TRN's
Making the Future report
The State of an Emerging Technology and a Look at What Lies Ahead

Report Number 12 March, 2004

Pattern Recognition Technologies: Getting the Picture

Executive Summary

Because our ability to recognize patterns is so ingrained, it seems easy. But attempts to program a computer to pick out individual words from the continuous flow of noise we call speech, to recognize the difference between a green apple and a tennis ball, to tell the difference between a female and male face, or to paraphrase a sentence have shown that pattern recognition is a sophisticated task.

Combine the ability to recognize objects and data patterns with a computer's fast computational abilities and round-the-clock hours, however, and you have the means to automatically organize mountains of video and audio, greatly improve the human-computer interface, and enable high-level security.

Improving automatic pattern recognition promises to speed data analysis tasks from analyzing DNA and proteins to classifying and retrieving text, audio and video. These abilities are poised to dramatically improve the processes of searching and mining data.

And not a moment too soon. Genetic material is copious. Human DNA contains 3 billion base pairs. At the same time, humanities' collection of facts and figures is growing at a rapid rate. From 1999 to 2002 the amount of stored data worldwide nearly doubled, according to a study by the University of California at Berkeley.

Pattern recognition is also intrinsic to computer vision, network intruder detection, forgery detection, and biometrics. Giving computers the ability to recognize human patterns of speech, gesture and handwriting promises to enable more sophisticated computer interfaces. Researchers are also working to find ways to automatically interpret language to give computers the abilities to paraphrase, translate, and understand language.

What to Look For

Data analysis:

- High-speed genome comparisons
- On-the-fly data stream mining
- Earthquake forecasting

Multimedia:

- Automatic classification of large music collections
- Video classification by soundtrack analysis
- Internet content-based image searching

Computer vision:

- Video hyperlinks
- Efficient quantum image-processing algorithm
- Robotic vision comparable to biological

Security:

- Automatic classification of Web attacks
- Automatic signature verification
- Practical network anomaly detection
- Hidden message deciphering

Interfaces:

- Natural gesture recognition
- Prosody recognition in speech interfaces
- Affective interfaces

Language processing:

- Document-scale automatic paraphrasing
- Multilingual speech translation
- Domain-independent speech translation

Neural networks:

- Massively parallel neural networks
- Quantum neural networks

Understanding the world

Humans have evolved to be great pattern recognizers. We have to be. Our senses gather an immense amount of information that cannot possibly all be analyzed carefully. We cope by naturally filtering out noise from patterns that are likely to be important, and turning most of our attention to these patterns.

This process is what allows you to tune out the chatter around you when you're reading a good book, but look up when someone calls your name.

Because the ability to recognize patterns is so ingrained, it seems easy. But attempts to program a computer to pick out individual words from the continuous flow of noise we call speech, to recognize the difference between a green apple and a tennis ball, to tell the difference between a female and male face, or to paraphrase a sentence have proven otherwise.

Seeing things our way

At the same time, computers that are able to recognize patterns are potentially very useful. Combine the ability to recognize objects and data patterns with a computer's fast computational abilities and round-the-clock hours and you have the means to automatically organize mountains of video and audio, greatly improve the human-computer interface, and enable high-level security.

In general, researchers are looking for ways to categorize a wide range of patterns, from words, music, images, and video, to genes, people's interests, handwriting, spoken language, and Web attacks.

Pattern recognition is poised to prove especially useful in five areas:

- Data analysis
- Computer vision
- Security
- Computer interfaces
- Language processing

Improving automatic pattern recognition promises to speed a wide range of data analysis tasks, from analyzing DNA and proteins to classifying and retrieving text, audio and video. These abilities are poised to dramatically improve data searching and mining.

Pattern recognition is intrinsic to interpreting data from computer vision devices that allow robots to see and that automate factory tasks and video monitoring. Pattern recognition also has many potential security uses, including network intruder detection, forgery detection, and biometrics, which involves recognizing human patterns like faces, irises, fingerprints and gaits.

And giving computers the ability to recognize human patterns of speech, gesture and handwriting promises to enable more sophisticated computer interfaces.

Language is also full of patterns that have meaning. Researchers are working to find ways to automatically interpret language to give computers the abilities to paraphrase, translate, and understand language.

Where things stand

Researchers from IBM, Johns Hopkins University, the University of Maryland and the Shoah Foundation are working on a project that illustrates both the utility and difficulty of automating the process of pattern-finding.

How It Works

Anatomy of a recognizer

Pattern recognition tasks involve three steps: data acquisition, feature extraction, and classification.

In the data acquisition step, an object — an image, an utterance or a DNA sequence — is fed into the pattern recognition system as an input pattern. Feature extraction involves describing the input pattern as a set of particular features — shapes, phonemes or base pairs. Classification involves determining whether the input pattern belongs in a certain category, or how closely it matches a template or previous input patterns.

To supervise or not to supervise

These tasks can be carried out in one of two modes: supervised or unsupervised.

In supervised pattern recognition, the categories are known ahead of time, and the goal is to determine the probability that a given input pattern belongs to a given category. Unsupervised pattern recognition uses no defined categories; it groups input patterns according to their relationships to each other.

The technologies

Pattern recognition technologies fall into four classes: template matching, structural, statistical and neural network.

Matching

Template matching is the simplest pattern recognition method. It involves directly comparing an input pattern to a template pattern, like the pixels of one image to the pixels of another.

Grammars

Structural pattern recognition assumes that patterns are derived from structures or hierarchies of patterns. This is analogous to the way grammar structures languages. The structures can be defined beforehand or inferred from known patterns used to train the system. Structural pattern recognition systems break a pattern into the basic components, or primitives, and classify the pattern based on the way the primitives are arranged.

Statistics

Statistical pattern recognition describes an input pattern as a set of n features and plots the pattern as a point in an n-dimensional imaginary space. Vehicles, for example, could be plotted using three features: size, passenger capacity and payload.

They are working to find ways to automatically index 116,000 hours of interviews in 32 languages from Holocaust survivors and rescuers. The Shoah foundation, established by Steven Spielberg, is manually indexing the material according to a thesaurus of keywords, but the process is time-consuming. It would take 40 years of eight-hour days to simply watch that many hours of tape.

The five-year automatic indexing project, which began in 2001, is exploring technologies that could speed things up, including speech recognition, multilingual speech recognition, and cross-lingual information retrieval. Although today's commercial speech recognition programs work tolerably well for a single trained user in a relatively quiet environment, it is a much bigger challenge to build a system that recognizes words contained in taped, emotional testimony from many users in many languages. (See "Speech Recognition to Sort Holocaust Tapes", page 19.)

Approaches to pattern recognition

Four basic technologies are used in pattern recognition:

- Template matching
- Structural pattern recognition
- Statistical pattern recognition, which includes the Nearest Neighbor, Support Vector Machine and Dynamic Bayesian Network methods
- Neural pattern recognition

Template matching is simply determining how closely an item matches an example or ideal. The approach is easy to implement but less effective than others because it is relatively rigid. It is not good at matching objects that are in different orientations, for example. Template matching is used in many image processing applications because its requirements are simple: just compare pixels.

Structural patterns are built from subpatterns, which are in turn built from primitives, or basic components. Patterns can be classified by their primitives and the rules, or grammar, that organize them. A wide range of complex patterns can be derived from a small number of primitives and simple rules. This approach works well when patterns have well-defined rules. Signal processing tasks like electrocardiogram interpretation fall into this category. The drawback to structural pattern recognition is the need to configure specific rules for each task.

The dominant approach in pattern recognition technology is statistical pattern recognition, which involves measuring features of an item, using these statistics to plot the pattern as a point in a graph, and comparing the point to a predefined range. Statistical methods can be used to handle virtually every type of pattern recognition task.

The nearest neighbor rule is a simple method of categorizing patterns according to their similarities. Support vector machines use more complicated mathematics to draw accurate boundaries between groups of patterns. Dynamic Bayesian networks, including the hidden Markov model widely used in speech

The three-dimensional space could have size as its x axis, passenger capacity as its y axis, and payload as its z axis. Where the three axes join is point 0, 0, 0. If the axes have 10 increments each, a dump truck might be represented as a point at 8, 2, 10, a bus as a point at 10, 10, 4, and a car at 3, 4, 2. Vans and pickup trucks would fall between these points.

Once a series of input patterns are plotted, the points are grouped in order to classify the patterns. This makes it possible to determine, for example, whether a new input pattern belongs in the truck group.

There are many classification techniques that use statistical pattern recognition. The simplest is the nearest neighbor rule, which says that a point belongs in the same category as the point nearest to it. A variation of the nearest neighbor rule classifies a point as belonging in the category of the majority of a certain number of nearest neighbors.

Another type of statistical classifier that has gained popularity in recent years is the the support vector machine, which maps input patterns into a higher-dimensional imaginary space and then determines the optimal boundary between groups of points. Finding the best boundary makes it more likely that a subsequent input pattern will be placed in the correct group.

A more complicated and widely used statistical pattern recognition technique is dynamic Bayesian networks, which is software that make a series of measurements to recognize patterns that involve changes over time. The method analyzes the transitions that occur as a system progresses through a series of observable states. The transition information makes it possible to both recognize the system and to predict future states. In speech recognition technologies, for example, a system is a spoken word and the phonemes, or sounds that make up a word, are the observable states. Two examples of dynamic Bayesian networks are the Kalman filter, which is often used in video-based pattern recognition tasks like object tracking, and the Hidden Markov Model, which is widely used in speech recognition technology.

Mimicking biology

Neural networks work according to the same principle as biological brains and similarly have the ability to adapt. Artificial neural networks are made up of many simple interconnected processing units, which are analogous to biological neurons. The output signal of one unit can be the input signal for others. Each unit has a threshold for firing that is based on strength or number of input signals. The connections between units are weighted, meaning some are stronger than others, and the weight of a connection increases the more the connection is used. Neural networks can be implemented in recognition technologies, use sequences of measurements to recognize patterns that involve changes over time.

Neural pattern recognition adds learning to the mix. Neural networks learn to recognize patterns by assigning weights to connections between neurons and changing the weights based on use. The technique, modeled after biological nervous systems, is popular with researchers because a system that learns requires less configuration than other pattern recognition technologies. Neural networks can be used for most types of pattern recognition tasks, but are especially appropriate for applications that involve sensory data. (See How It Works, page 2)

Keen eye for the numbers

Data analysis is all about categorizing, spotting trends, and spotting anomalies. Techniques for sifting through data, including statistical analysis, have been in use since the early part of the 20th century. The rise of computing has accelerated the science of data analysis through a feedback cycle of better number crunching techniques begetting greater amounts of data in need of crunching. Statistical analysis in particular is the foundation for much of pattern recognition technology.

Today's computer processing and storage capabilities, combined with the Internet, have caused humanities' collection of facts and figures to balloon to ridiculous proportions. From 1999 to 2002 the amount of stored data worldwide nearly doubled, according to a study by the University of California at Berkeley. The world stored about 5 million terabytes of information in 2002, not counting copies. Ninety-two percent of this increase is stored on magnetic media, principally computer hard disks. Five million terabytes is 500,000 times the size of the print collection of the Library of Congress.

The desire of corporations and governments to tap this vast resource has fueled efforts to apply the latest pattern recognition technologies to the number crunching problem, which has given rise to the field of data mining. Corporations, for instance, are using the technology to analyze the increasing amounts of personal information collected in their databases in order to target marketing efforts to ever smaller segments of the population.

Words

Though statistical data is a matter of numbers, our communications and the recordings of commerce, culture and education are largely in word form. There's a long history of using computers to attempt to recognize patterns in words; language processing was one of the first branches to emerge in artificial intelligence research. Word pattern recognition is used for document analysis and data mining.

Cornell researcher John Kleinberg has devised a method of organizing data based on sudden changes in content. He used the method to search American presidential State of the Union addresses to produce a list of words that summarized the

software on an ordinary computer or in specialized hardware, including reconfigurable electronic circuits and optical circuits.

A basic type of neural network commonly used for pattern recognition is the feed-forward network, which consists of a layer of input units connected to a layer of output units. Input patterns produce output patterns. Initially the relationship between input and output is random, but after a period of training with example patterns, certain connections become more heavily weighted. In this way the network learns to associate a range of input patterns with a specific output pattern.

Self organizing maps are another type of neural network used in pattern recognition. The maps are feed-forward networks that are not trained. Only one unit at a time receives input, and initially the process is random. But once a unit has received input from a particular pattern, it learns the pattern and is more likely to receive subsequent input from the same pattern than other units, which reinforces the learning. In addition, nearby units also learn and begin to receive input from similar patterns. In this way the network automatically forms a map that groups patterns.

Who to Watch

Data analysis

Eytan Domany, Weizmann Institute of Science
Rehovot, Israel
www.weizmann.ac.il/~fedomany

Chandrika Kamath, Lawrence Livermore National Laboratory
Livermore, California
www.llnl.gov/CASC/people/kamath

Vipin Kumar, University of Minnesota
Minneapolis, Minnesota
www.cs.umn.edu/~kumar

Xindong Wu, University of Vermont
Burlington, Vermont
www.cs.uvm.edu/~xwu/home.html

Philip S. Yu, IBM Research
Yorktown, New York
www.research.ibm.com/people/p/psyu

Biometrics and security

Anil Jain, Michigan State University
East Lansing, Michigan
www.cse.msu.edu/~jain

Sargur N. Srihari, State University of New York at Buffalo
Buffalo, New York
www.cedar.buffalo.edu/~srihari

Dit-Yan Yeung, Hong Kong University of Science and Technology
Hong Kong, China
www.cs.ust.hk/~dyyeung

important topics in politics at the time of the speech. (See "Word Bursts Reveal Hot Topics", page 12.)

University of Rome researchers have found that is possible to use differences in the way text files are compressed to compare how close a string of unknown text is to a block of text written in a specific language or by a specific author. (See "File Compressor IDs Authors", page x.)

Researchers from the University of Chicago have shown that it is possible to identify groups of users who have common interests by analyzing requests for data. The researchers are looking to use the patterns to design more efficient services for resource-sharing environments like Grid computing. (See "Net Scan Finds Like-minded Users", page 9.) Grid computing puts together disparate resources across networks like the Internet to form virtual computers that can give users the computing power to tackle large problems like processing the data from physics experiments.

Researchers from the Center for Mathematics and Computer Science (CWI) and the University of Amsterdam have devised a universal similarity metric that measures the information distance between two files by measuring how easily one file can be compressed using the information contained in the second file. They have used the metric to compare music files to determine how similar they are, to construct evolutionary trees using mitochondrial genes of different animal species, to put together language trees of 52 Eurasian tongues, and to detect plagiarism in students' programming assignments. (See "Software Sorts Tunes", page 10.)

These data analysis techniques can be used in practical applications now.

Genes — an ocean of data

Comparing genes is a particularly tedious pattern recognition task. Genetic material is copious. DNA consists of strings of base pairs connected to a sugar phosphate backbone. Human DNA contains about 3 billion base pairs. Some plant DNA can be even more voluminous: wheat has 16 billion base pairs. There are between 25,000 and 75,000 genes in human DNA, each made up of about 1,000 codons, which in turn consist of three base pairs each. Human genes encode the information for producing the 100,000-odd proteins that carry out the biochemical functions of human life.

Researchers from Cornell are using a programming algorithm similar to those used for analyzing language to cut down the possibilities when comparing DNA segments in order to determine how close one species is to another. (See "Software Speeds Gene Comparison", page 16.)

Weizmann Institute researchers have devised a clustering algorithm that identifies two types of subsets in data, then uses one subset to cluster the other. The cross-clustering method clusters based on similar traits even when the traits are not identified before hand. They used the method to sort gene data. (See "Software Cross-Sorts Gene Data", page 17.)

Computer vision and image processing

Jezekiel Ben-Arie, University of Illinois at Chicago
Chicago, Illinois
www.eecs.uic.edu/eecspeople/benarie.htm

Joachim M. Buhmann, University of Bonn,
Bonn, Germany
www-dbv.informatik.uni-bonn.de/~jb

Hany Farid, Dartmouth University
Hanover, New Hampshire
www.cs.dartmouth.edu/~farid

Ahmed Ghali, Nottingham University
Nottingham, England
www.cs.nott.ac.uk/~ahg

Bahram Javidi, University of Connecticut
Storrs, Connecticut
www.engr.uconn.edu/ece/labs/ospclab

Baback Moghaddam
Mitsubishi Electric Research Laboratory
Cambridge, Massachusetts
www.merl.com/people/moghaddam/moghaddam.html

Shree Nayar, Columbia University.
New York, New York
www1.cs.columbia.edu/~nayar

Tomaso A. Poggio, Massachusetts Institute of Technology
Cambridge, MA
www.ai.mit.edu/projects/cbcl

Jianbo Shi, University of Pennsylvania
Philadelphia, Pennsylvania
www.cis.upenn.edu/~jshi

James Wang, Pennsylvania State University
University Park, Pennsylvania
wang.ist.psu.edu

Harry Wechsler, George Mason University
Fairfax, Virginia
cs.gmu.edu/~wechsler

Interfaces and language processing

Nelson Morgan, University of California at Berkeley
Berkeley, California
www.icsi.berkeley.edu/~morgan

Rajeev Sharma, Pennsylvania State University
University Park, Pennsylvania
www.cse.psu.edu/~rsharma

Alex Waibel, Carnegie Mellon University
Pittsburgh, Pennsylvania
www.is.cs.cmu.edu/mie

Jie Yang, Carnegie Mellon University
Pittsburgh, Pennsylvania
www-2.cs.cmu.edu/~yang

Steve Young, University of Cambridge
Cambridge, England
mi.eng.cam.ac.uk/~sjy

Victor Zue, MIT
Cambridge, Massachusetts
www.sls.csail.mit.edu/sls/people/zue.shtml

These genetic analysis tools can be used in practical applications now.

Neural networks

Gail A. Carpenter, Boston University
Boston, Massachusetts
cns-web.bu.edu/~gail

David Horn, Tel Aviv University
Tel Aviv, Israel
neuron.tau.ac.il/~horn

David J.C. MacKay, University of Cambridge
Cambridge, England
www.inference.phy.cam.ac.uk/mackay

Paul Sajda, Columbia University
New York, New York
newton.bme.columbia.edu/~psajda

Audiovisual

An increasing amount of data is stored in the forms of sounds and images, and an increasing portion of the data is digitized. Given the difficulty of manually indexing this information, pattern recognition technologies are at the center of efforts to automate the classification of multimedia data in order to increase the accessibility of audio, image and video files.

Scientists from Microsoft Research have written an algorithm that allows computers to differentiate speech, music, environmental sounds, and silence in video soundtracks by mapping and comparing the characteristics of each type of sound. (See “Software Sorts Video Soundtracks”, page 18.)

Researchers from Pennsylvania State University have written image retrieval software that maps images’ key features and assigns images to several broad categories, then retrieves images by matching the features and categories of a query image to images stored in the database. (See “Image Search Sorts by Content”, page 20.)

Multimedia pattern recognition techniques could be used in practical applications in two to ten years.

Seeing things

In the area of computer vision and image processing, pattern recognition is the only game in town. Biological senses inherently involve pattern recognition and have as much to do with how the mind interprets data as how the data is gathered. Given that biology is the model for machine vision, neural networks are an active area of research.

Researchers from the State University of New York at Buffalo and Stanford University have built a silicon retina that uses a timing signal to mimic a form of data compression performed by biological eyes. The electronic retina processes the data that makes up an image in a way that allows it to extract information about the edges of images, which is usually sufficient for detecting and tracking objects. (See “Vision Chip Shines”, page 22.)

Researchers from the University of Amsterdam have come up with an object tracking technique that captures more than just the shape of an object. The software tracks changes in appearance of every part of an object, making it easier for a computer to recognize objects despite angle and lighting changes. (See “Computer Follows Video Action”, page 23.)

Researchers from the University of Pennsylvania have built a system that determines the gender of humans by examining faces and voices. (See “Computers Sort Gender in a Binary World”, page 24.)

A University of British Columbia physicist has come up with an algorithm that proves that quantum computers would be faster at finding patterns than today’s classical computers. (See “Quantum Software Gets the Picture”, page 25.) Quantum computers use the attributes of particles like atoms and electrons to store the 1s and 0s of computer information.

Many of these computer vision technologies can be used in practical applications now, but more sophisticated types of artificial eyes are at least a decade away. Practical quantum computers are one or two decades away.

Digital watchdogs

Pattern recognition can also be used in security tasks ranging from detecting irregular activity in computer networks to spotting forged signatures. Security boils down to being able to detect patterns that fall outside of a range that defines normal activity.

Researchers from the University of California at Davis have adapted text classification clustering techniques to categorize computer users into two groups — authorized users and intruders. The software treats system calls as words and sequences of system calls as documents, and classifies each document as one generated during normal or intrusive activity. (See “Text Software Spots Intruders”, page 26.)

Researchers from the University of South Carolina are using nuclear experimentation methods to map network traffic and extract patterns of typical network behavior that can then be used to tease out observations that point to known network intruders. (See “Physics Methods May Spot Intruders”, page 27.)

Researchers from the Spanish Institute of Applied Physics have come up with a way of sorting inconsistent sets of data into groups that share some similarity. The method compares the distance — the minimum number of elementary edit operations like deletions, insertions and substitutions — needed to transform one piece of data into another. The researchers used the method to group Web attacks by severity. (See "Degree of Different Sorts Data", page 28.)

A Dartmouth college researcher has devised a way to detect hidden messages in digital images by compressing a photo, then comparing it with a statistical profile of compressed data files that make up undisturbed images. (See "Statistics Sniff out Secrets ", page 28.)

And researchers from the University of Buffalo have written a handwriting recognition program that recognizes and analyzes features like character shape, line separation, and stroke slant and thickness. (See "Software Spots Forged Signatures", page 29.)

These security technologies could be used in practical applications in one to three years.

Catch my meaning

One of the most challenging tasks for pattern recognition technologies is human-computer interaction. Researchers are tackling the problem in pieces.

Researchers from Pennsylvania State University and Advanced Interface Technologies, Inc. are working to untangle the relationships between attributes of speech and attributes of gestures in order to improve the way computers recognize gestures. Although it's easy for humans to interpret gestures, these patterns are complicated; there's no one-to-one mapping of form to meaning. (See "Interface Gets the Point", page 30.)

A related group of Pennsylvania State University researchers are working on a human-computer interface that allows a computer to see where a human is pointing and uses that information to interpret mixed speech and gestural directions. (See "Interface Lets You Point and Speak", page 32.)

And researchers from the Chinese Academy of Sciences and the Chinese Institute of Modern Optics have come up with a way to use skin color to improve gesture recognition systems. Colors are elusive because the change depending on illumination. The researchers' system uses an artificial neural network to detect skin color, then processes the information to determine which skin-color objects should be connected together. (See "Neural Net Tracks Skin Color", page 33.)

Gesture recognition technologies are two to five years away from practical application.

More than words

Perhaps the most heavily researched aspect of human-computer interaction is speech recognition. Decades of research are beginning to yield usable technologies, and the average person is likely to have encountered speech recognition through corporate call centers. Nevertheless, there is still a large gap between the speech recognition capabilities of computers and those of the average 5-year-old.

Researchers from SRI International and Bilkent University are working to use patterns of prosody — information gleaned from the timing, pitch and loudness of speech — to improve speech recognition software. (See "Hearing Between the Lines", page 33.)

Speech recognition is also enabling technologies that extend beyond the average person's capabilities, including real-time speech translation.

Researchers from Carnegie Mellon University, Cepstral, LLC, Multimodal Technologies Inc. and Mobile Technologies Inc. have built a two-way speech-to-speech system that translates medical information from Arabic to English and English to Arabic. The system works by extracting key meaning from an input sentence and translating it to an interlingual representation. The process is akin to constructing a medical-context template that fits key information, then filling in the template. (See "PDA Translates Speech", page 34.)

The speech technologies are two to five years away from practical application.

Pattern recognition technologies also play a key role in language processing research. Language processing is a decades-old discipline that goes beyond word recognition to give computers a shot at working with meaning.

Researchers at Cornell University have employed two large databases of news stories and an algorithm borrowed from computational biology to construct a system that automatically paraphrases whole sentences. The system uses word clustering and gene comparison techniques to identify word patterns that describe similar events within each news source, then taps the differences between the patterns of the two different sources to construct paraphrases. (See "Software Paraphrases Sentences", page 36.)

The paraphrasing software could be used in practical applications now.

Learning from the brain

The master of pattern recognition is the biological brain, which is formed from large numbers of neurons that each have large numbers of connections to other neurons. Neural networks learn by assigning weights to the connections between neurons and changing the weights based on use. This allows a specific set of inputs to be associated with a pattern of weighted neural connections.

The key to making useful neural networks is making them large enough. The human brain contains billions of neurons and trillions of connections. This is an enormous number. A single human brain contains thousands of times more connections than there are stars in the Milky Way Galaxy.

Researchers from Carlos III de Madrid University and the Massachusetts Institute of Technology have devised a neural network architecture that uses a novel mix of optics and electronics to accommodate large numbers of neurons. (See "Design Enables Large Neural Nets", page 37.)

Researchers from Wichita State University are aiming to combine neural networks with quantum computing in order to use neural networks' proven capacity for learning to help realize quantum computing's potential to solve astronomically large problems. (See "Self-Learning Eases Quantum Computing", page 37.)

Large neural networks could be developed in two to five years. Quantum neural networks are a decade away.

Making sense

Developing pattern recognition technologies that make machines work more like humans is a matter of necessity rather than vanity. Taming the data overload monster that we have created will require adding a semblance of our superior pattern recognition abilities to computers' raw number-crunching power. And giving robots the means to interpret their surroundings will allow them to venture further into the everyday world. Powerful pattern recognition capabilities are also the keys to more efficient automation, better human-computer interaction and, for better or worse, widespread surveillance.

Recent Key Developments

Advances in data analysis:

- A method of identifying communities of interest from patterns of data requests (Net scan finds like-minded users, page 9)
- A system that groups music according to similarity based on abstract symbolic data (Software sorts tunes, page 10)
- A method of determining topics of interest for a given time based on word frequency (Word bursts reveal hot topics, page 12)
- A method of determining authorship based on differences in compressed text files (File compressor ID's authors, page 12)
- Two software tools that reduce the scope of vast amounts of real-world data to aid pattern recognition (Tools cut data down to size, page 14)
- An efficient method of comparing gene data (Software speeds gene comparison, page 16)
- A method of analyzing gene data based on clusters (Software cross-sorts gene data, page 17)

Advances in multimedia:

- A method of classifying the sounds on video soundtracks (Software sorts video soundtracks, page 18)
- A project to use speech recognition software to index videotape interviews of Holocaust survivors (Speech recognition to sort Holocaust tapes, page 19)
- A method of searching image databases by matching the features of a query image to those of target images (Image search sorts by content, page 20)

Advances in computer vision:

- An artificial retina with high-speed, optical output (Vision chip shines, page 22)

- A model of how the brain recognizes the faces of people it's seen before, Ohio State University, April 2003
- A system that tracks objects in video by monitoring changes to each pixel in the object (Computer follows video action, page 23)
- A method of determining gender based on face and voice analysis (Computers sort gender in a binary world, page 24)
- A theoretical analysis that shows that quantum computers could be more efficient than ordinary computers at visual pattern recognition (Quantum software gets the picture, page 25)

Advances in security:

- A method of detecting intrusions based on text classification techniques (Text software spots intruders, page 26)
- A system for identifying people by their gaits, Georgia Institute of Technology, October 2002
- A method of detecting intrusions based on particle physics analysis (Physics methods may spots intruders, page 27)
- A method of grouping dissimilar data based on degree of difference (Degree of difference sorts data, page 28)
- A method of detecting data hidden in images based on statistical analysis (Statistics sniff out secrets, page 28)
- A method of identifying forged signatures based on a neural network that learns the differences between examples from the same writer (Software spots forged signatures up, page 29)

Advances in interfaces:

- A computer interface that correlates gesture with the loudness, pitch and timing of speech (Interface gets the point, page 30)
- A computer interface that interprets gestures and speech (Interface lets you point and speak, page 32)
- A neural network that learns to track skin color to improve gesture recognition and gaze tracking software (Neural net tracks skin color, page 33)

Advances in language processing:

- A method of improving speech recognition software by tracking the loudness, pitch and timing of speech (Hearing between the lines, page 33)
- A system that translates spoken language about a given topic in real-time (PDA translates speech, page 34)
- A system that paraphrases whole sentences by analyzing multiple news sources using gene comparison techniques (Software paraphrases sentences, page 36)

Advances in neural networks:

- A design for a large neural network consisting of both electronic and optical circuits (Design enables large neural nets, page 37)

A simulation that shows that a quantum neural network could allow quantum computers to program themselves (Self-learning eases quantum programming, page 37)

Data

Net Scan Finds Like-Minded Users

By Kimberly Patch, Technology Research News
May 7/14, 2003

When you search for information on the Web, chances are you aren't alone — there are like-minded groups of users across the Web searching for the same sorts of things.

Researchers from the University of Chicago have shown that is possible to identify these groups by analyzing browsing patterns, even in networks as far-flung as the Web.

The researchers' method of graphing information across data distribution systems like the Internet shows that, given a large enough sample, computer users can be grouped according to their common interests based only on their requests for data. "One of the first questions we asked was is the group-based collaboration of scientists mirrored somehow in their usage of data," said Adriana Iamnitchi, a researcher at the University of Chicago.

The answer turned out to be yes, across all types of group-based interests. "Communities as heterogeneous as the Web seem to show this pattern of having users naturally group in interest-based groups," she said.

The information-request graphing method can be used to design scalable, adaptive methods for locating and delivering data, said Iamnitchi. The method could theoretically be used by anyone, including ecommerce vendors, to target communities of interest.

The researchers are working on using the patterns to design more efficient services for resource-sharing environments like Grid computing, Iamnitchi said.

Grid software coordinates a few or even hundreds of computers across networks like the Internet to piece together compute power and resources like databases into powerful virtual computers; the combined resources can speed up scientific and engineering applications like time-consuming equations and three-dimensional simulations.

The researchers found the data-sharing relationship pattern while looking for a way to leverage characteristics of the Grid computing community to make that type of computing more efficient, according to Iamnitchi. "Our idea was to... design mechanisms [that are] able to cope efficiently with large and dynamic numbers of resources — data files, computers, and storage space for results," she said.

One typical characteristic of the community that uses Grid computing is they tend to collaborate, said Iamnitchi. When the researchers analyzed traces of scientific computations from a high-energy physics collaboration that spanned 18 countries and involved 70-odd institutions and thousands of physicists, they found that the patterns of collaboration were mirrored in scientists' data requests.

The researchers looked at the relationships that formed among users based on the data they were interested in. "We captured and quantified these relationships by modeling the system as a data-sharing ... graph whose nodes are the data consumers in that system," said Iamnitchi. Nodes, or people, who requested a given number of the same files within a given time were connected.

In an analysis of six months worth of scientists' requests for data, the researchers found that group-based collaboration is visible in the way information isa requested, said Iamnitchi. "Scientists form groups of interest based on the data they used," she said. The researchers found the same pattern in a larger analysis of general Web requests.

The pattern of similar requests shared the small-world characteristic common in many networks, including the way data is arranged in networks like the Internet.

In small-world networks, it is possible to get from one node to any other node by traversing relatively few links. Social networks, with people as nodes and relationships as links, and the Web, with pages as nodes, and links between pages as links, are also small-world networks.

Looking at small-world topologies is not a novel idea, but the method of extracting a graph from an arbitrary data-sharing relationship and using it to study these structures is, said Filippo Menczer, an assistant professor of management services at the University of Iowa.

Data request patterns have been analyzed previously, but in different ways — to examine the popularity distribution of Web requests or to study the most efficient way to cache Internet traffic. In contrast, the Chicago researchers' analysis uncovered relationships between users based on their common interests in data.

The method is potentially useful, especially because a graph can be made from any Web usage log, said Menczer. "Any Webmaster can do this."

The method may be useful for discovering clusters of users who have interest in a certain type of data, Menczer said. "Ecommerce vendors are currently using collaborative filtering techniques, which are related to this," to do so, he said. The method can also be used for distributed caching and broadcasting, similar to the services offered by Akamai Technologies Inc., he said.

The researchers are now making the method more efficient for resource-sharing environments like Grid computing, said Iamnitchi. "We are currently looking... to design mechanisms to locate resources," she said. "The ultimate goal is to provide scalable, adaptive mechanisms [that are] able to deal with variations in resource participation."

The resource location mechanisms could be ready to use within two years, Iamnitchi said.

Iamnitchi's research colleagues were Matei Ripeanu from the University of Chicago and Ian Foster from Argonne National Laboratory. The research was funded by the National Science Foundation (NSF).

Timeline: 2 years
Funding: Government
TRN Categories: Internet; Distributed Computing
Story Type: News
Related Elements: Technical paper, "Data-Sharing Relationships in the Web," posted on the arXiv physics archive at iv.org/abs/cs.NI/0302016

TRN

Software Sorts Tunes

By Kimberly Patch, Technology Research News
April 23/30, 2003

Exactly what is it about The Fine Young Cannibals that is similar to Roxy Music?

Many of today's online retailers use the opinions of lots of people to suggest selections that are similar to music a customer has purchased. The challenge of automating the process comes down to teaching a computer how to tell that a piece of music written by Mozart sounds like music written by Mozart.

Researchers from the Center for Mathematics and Computer Science (CWI) and the University of Amsterdam in the Netherlands have tapped computational biology and information theory to find a way to automatically compare music files to determine how similar they are.

The work is the latest application of a universal similarity metric that the researchers have used to construct evolutionary trees using mitochondrial genes of different animal species, put together language trees of 52 Eurasian tongues, and detect plagiarism in student programming assignments.

The researchers' latest application shows that it could be used to categorize music, and to determine the true origin of music authorship. "Our method works well on abstract symbolic data, that is, data without a direct physical interpretation," said Rudi Cilibrasi, a researcher at CWI.

The method uses a universal similarity metric that measures the information distance, or differences between two files, as a number between zero and one. "Zero means they're the same file, and one means they have no relation whatsoever. Most file pairs fall somewhere in between," said Cilibrasi.

The universal metric measures how easily one file can be compressed using the information contained in the second file. Compression algorithms are regularly used to remove redundant information from files to make the files smaller. "Two files are closer if one can be compressed [more fully] given the information in the other," said Cilibrasi.

This method takes into account all similarities between a pair of sequences, according to Paul Vitanyi, a fellow at CWI and professor of computer science at the University of Amsterdam. "It is universal in that it discovers all effective similarities," he said.

A human expert will categorize pieces of music by searching for specific similarities related to pitch, rhythm or harmony, according to Vitanyi. Existing systems that automatically classify music do the same, looking, for instance, for similar rhythms.

The CWI researchers' universal metric approach does not look for particular features, according to Vitany. This means the clustering takes into account features that are not apparent, said Vitanyi. "Existing systems tend to use... very particular features to do the clustering," he said. "Our theory tells us we — ideally — account... for all features [including] the many arbitrary ones we don't know and even cannot know," he said.

The method could be used to classify music in commercial applications, said Cilibrasi. It could "enable better cataloging and indexing based on people's preferences... without the help of pre-existing purchase records or other human input," he said.

The researchers' method has two steps. The researchers first compared a group of MIDI files and calculated the information distance between every pair of files.

Then they used a hill-climbing algorithm to organize the differences among files into a branching tree, with closer files situated closer on the tree, said Cilibrasi. "If two files have a short distance in the [distance] matrix, then they should also be close together in the tree," he said.

A hill-climbing algorithm starts at a random place and aims to improve the solution at each step. A possible solution to a problem can be thought of as a hill, and many possible, unconnected solutions a group of hills. In order to keep the hill-climbing algorithm from getting stuck on a small hill, the researchers added a randomized escape sequence.

To speed the process, they modified the hill-climbing algorithm so that it examined many files in parallel. This enabled them to categorize larger groups of music than was previously possible, said Cilibrasi. The method was able to handle 60 items, but works best with sets of 20 to 30 items Vitanyi said.

The researchers tested the method by having it categorize increasingly similar files. The program correctly classified a group of 12 very different types of files: gene sequences, excerpts from a novel, classical MIDI files, jazz MIDI files, and linux and java computer code.

The algorithm was 85.8 percent accurate in distinguishing the genres of 36 classical, jazz and rock pieces. The program placed some of the rock music in a separate group near the main rock music group, and classified a pair of Bach preludes and a Chopin piece in the other genres.

The apparent misclassifications could be errors of the program, or could reveal less apparent likenesses among specific pieces in different genres, according to Cilibrasi.

The method was more accurate at sorting music by composer within genres, especially with smaller groups of music. It was 95.8 percent accurate in sorting 12 Bach, Chopin and Debussy pieces; 89.5 percent accurate in sorting 32 pieces from Bach, Chopin, Debussy and Haydn;

84.4 percent accurate when 28 pieces from Beethoven, Buxtehude and Mozart pieces were added to the mix; and 86 percent accurate in sorting 34 symphonies from Haydn, Mozart, Beethoven, Schubert and Saint-Saens.

Applying methods from computational biology and information theory to music is not new, said Elaine Chew, an assistant professor of industrial and systems engineering at the University of Southern California. The challenge, however, is finding a practical way to implement the theory, she said.

The particular metrics the researchers chose have had previous success in building phylogeny trees to help in classifying mitochondrial genes and Eurasian languages, Chew added. "What is new is the application of these methods to music classification," she said.

Further work needs to be done in order to better understand what it means for two pieces to be close according to the researchers' metrics, said Chew. In addition, it appears that the method doesn't work well with larger data sets, she said. This "would limit its applicability to the classification of today's digital music libraries," she said.

This type of research, however, does have potential practical applications, said Chew. "Similarity metrics are a critical part of query-by-humming and... query-by-example systems," she said. These types of metrics also form the core of music search and recommendations systems for personalized music applications, she said.

The basic software works to detect patterns that appear meaningful, said Cilibrasi. "We still lack... a clear idea of the best way to practically use these techniques," he said. "We think [they] need to be applied to a larger variety of subject areas."

The researchers are currently working on making the algorithm more consistent, and developing better methods of visualizing the information that the algorithm brings to light, said Cilibrasi. "A binary tree is not the only way to visualize this information. We may consider using arbitrary trees or other pictures," he said.

The basic method could also be applied to disciplines like art history and forensics, said Cilibrasi.

Cilibrasi and Vitanyi's research colleague was Ronald de Wolf at CWI. The research was funded by the Netherlands Organization for Scientific Research (NWO).

Timeline: Unknown
Funding: Government
TRN Categories: Databases and Information Retrieval
Story Type: News
Related Elements: Technical paper, "Algorithmic Clustering of Music," posted on the Computer Research Repository (CoRR) at arxiv.org/abs/cs.SD/0303025; "The Similarity Metric," presented at the 14th Association for Computing Machinery-Society for Industrial and Applied Mathematics (ACM-SIAM) Symposium on Discrete Algorithms, January 12-14, 2003 and posted at www.cwi.nl/~paulv/selection.html

Word Bursts Reveal Hot Topics

Technology Research News, February 26/March 5, 2003

Cornell University researcher Jon Kleinberg has shown that you can learn a lot about what topics are becoming important at any given time by analyzing large collections of documents for sudden changes in content.

Kleinberg tested the theory by searching his email using an algorithm that examined the frequency of words over time. These provided clues about the hot topics of the moment. For example, the searches made it apparent when exam time was near — the word "prelim" suddenly became more frequent.

Searching American presidents' state of the union addresses produced a list of words that neatly summarized the important topics in american politics at any given time: gentleman was popular in the late 1700s, militia burst onto the scene in 1801, bank showed up in 1833, California in 1848, slaves in 1859, rebellion in 1861, emancipation in 1862, paper in 1867, coinage in 1877, forest in 1901, interstate in 1907, marketing in 1919, atomic in 1947, inflation in 1971, oil in 1974, crime in 1991, businesses in 1990, health in 1992, and teachers in 1996.

Kleinberg presented his work at the American Association for the Advancement of Science (AAAS) annual meeting in Denver on February 18.

File Compressor ID's Authors

By Kimberly Patch, Technology Research News
April 17/20, 2002

Using computers to find language patterns like the forms that make up the grammar of a particular language or the turns of phrase a certain author tends to use is both useful and tricky.

The main challenge is a version of the classic chicken-or-egg conundrum: if the patterns are unknown, how you know what to tell a program to look for?

Researchers from the University of Rome in Italy are using differences in the way text files are compressed as points of comparison. Using these patterns, the researchers can compare how close a string of unknown text is to a block of text written in a specific language, or even by a specific author. Data compression techniques are used to shrink files so they take up less storage space and travel more quickly over networks.

The researchers tested the method using gzip, a well-known compression program, and 90 texts written by 11 authors. They included a short string of unidentified text at the end of each of a group of longer, known texts, and used the compression process to compare the short string of text to the group of known texts.

This type of research could make it easier to automatically identify languages, identify authorship of a document, or adapt computer tools like spell checkers to different languages. It could also be used to automatically classify other types of data like DNA sequences and seismic information.

Compression algorithms, or zippers, save space by rooting out redundancy. They replace an original file with a group of building blocks plus instructions for putting the building blocks together to reconstruct the original file. "A zipper is... explicitly conceived to take a file and try to transform it into the shortest possible file," said Vittorio Loreto, a physics researcher at the University of Rome.

The researchers measured how similar the short, unknown string of text was to the file it was compressed with by noting how concisely the zippers for each known file compressed the unknown string.

"If we take a long English text and... apend to it an Italian text, then zip [the resulting file], the zipper begins reading the file starting from the English text," said Loreto. "So after a while it is able to encode optimally the English file. [But] when the Italian part begins, the zipper starts encoding it in a way which is optimal for the English," he said.

As long as the Italian portion of the file is short enough, the zipper will compress the Italian using the English rules, Loreto said. This makes it possible to use the compression rate as a measure of how close the two languages are, he said. "If the length of the Italian file is small enough, the difference between the length in bits of the zipped Italian plus English text and the length in bits of the English text zipped alone will give a measure of the distance between the two texts," he said.

This distance, or entropy between two languages is the difficulty a native speaker of one language encounters in attempting to read text written in another, said Loreto.

This distance between known and unknown texts is a tool that computer algorithms can use to recognize the context of a given sequence — things like its language, subject and author, Loretto said.

For automatic language recognition, for example, a piece of unknown text can be compared with many texts in many different languages. The procedure will show the language that is the closest to the unknown file. If the language collection includes the language in question, the closest file will be the same language, said Loreto.

In a similar way, unknown files can be automatically sorted by subject and author as well, he said. "If you are able to guess automatically the subject a given text is about without the need to read it, you can exploit this feature for automatic classification," he said.

The distance data can be used to automatically build tree-type representations of relationships among texts, said Loreto. "This way one also has a graphical representation of the corpus structure."

A key problem plaguing computer linguists is finding patterns even when you don't know exactly what you're looking for, said Patrick Juola, an assistant professor of computer science at Duquesne University. The University of Rome researchers have tackled the problem by adapting the "best-known pattern-finding machines — compression algorithms — to the task of learning patterns useful for categorizing language," he said.

There are several groups working on similar methods, Juola added. "Language has interesting patterns that can be found automatically. The question that remains is what's the best way to find and exploit these patterns, and what kinds of problems can we solve with them," he said.

This type of research is useful for things like Web searches and finding authorship, said Juola. "The major implication... is that this kind of technique makes it much easier to analyze linguistic documents without having a detailed knowledge of the language in question," he said.

The method could help adapt existing language tools like Web search engines, and spelling and grammar checkers to new languages that don't have a long scientific and analytical tradition, said Juola. "If I were the University of Nepal, I'd be very interested in seeing how this could help with the development of a modern information processing system for Nepali documents."

One language recognition challenge is gaining information by using just a few characters. The ability to identify the language a document is written in in as few as 20 characters may allow the researchers' method to identify the language of picture captions and table headings, which is hard to do right now, Juola said. One possible drawback to the method is that because it needs a lot of text in various languages to make a comparison, it may take a long time to do so, Juola added.

The researchers have tested the method by classifying large law and literature collections, said Loreto. The method can also automatically classify strings of characters that represent non-text coding, [like] time sequences, genetic sequences... stock market data, and medical monitoring data, he said.

The next step in the work is applying it in one of these areas, said Loreto. The researchers are looking at "the analysis of biological data — DNA or protein sequences — and the investigation of phenomena described in terms of time-series: earthquake sequences, stock-market data, or other kinds of signals," he said.

In both cases, "we expect to be able to implement procedures for automatic recognition and classification [to identify] specific significant patterns like the sequences codifying the genes in DNA or some premonitory pattern in geological or medical data," he said.

Loreto's research colleagues were Dario Benedetto of the University of Rome and Emanuele Chaglioti of the University of Rome and the Center for Statistical Candidates and Complexity (INFM) in Italy. They published the research in the January 28, 2002 issue of Physical Review Letters. The research was funded by the University.

Timeline: Now
Funding: University
TRN Categories: Applied Technology; Pattern Recognition
Story Type: News
Related Elements: Technical Paper, "Language Trees and Zipping," Physical Review Letters, January 28, 2002

Tools Cut Data down to Size

By Kimberly Patch, Technology Research News
March 14, 2001

Humans quickly recognize objects despite the many combinations of angle and lighting changes that make nothing look exactly the same from one moment to the next.

The key to this recognition is our instinctive ability to quickly determine which of the objects' many traits, or data points to pay attention to. What it really comes down to is a lifelong version of the old Sesame Street game One of These Things Is Not like the Other.

Getting a computer to glean the traits needed to determine that two views of the same face are just that is actually a difficult problem involving a lot of higher math. Its solution also has many potential uses.

Gaining this ability would help computers interact with their environments by better understanding spoken speech, deciphering handwriting and recognizing faces. It would also make it easier to extract information from large databases and sort noise from data in audio and video applications.

Two research teams are pushing the field forward with algorithms that reduce the dimensionality of nonlinear data.

Linear data can be represented algebraically in steps that come one after another. Classic data algorithms like principal complement analysis (PCA) and multidimensional scaling (MDS) have been used for years to reduce the dimensionality of linear data. As it turns out, however, relationships in real-world data are usually nonlinear, involving things like loops.

Dimensionality refers to the number of measurements available for each item in a data set, according to Sam Roweis, a senior research fellow at University College London's Gatsby Computational Neuroscience Unit. A set of data that has just two dimensions can be plotted on a flat graph, while data with three dimensions can be plotted on a graph with x,y and z axes. The dimensionality of real world items, however, is inevitably much higher than can be pictured using the three dimensions we can see.

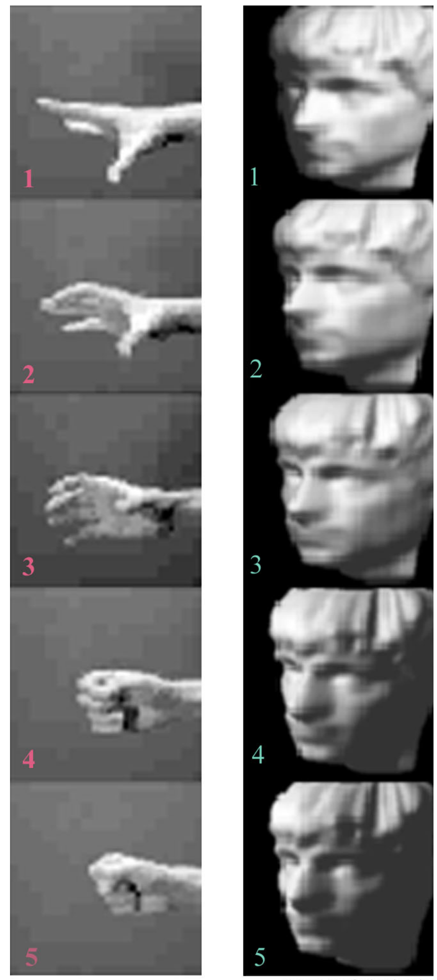
"For example, if you have a bunch of 600 by 600 images, their dimensionality is 360,000 since that is how many pixels each one has," Roweis said. Similarly, the dimensionality of data measuring the frequency of a group of words across a number of documents equals the number of documents.

The key to analyzing data is comparing these measurements to find relationships among this plethora of data points.

In nonlinear data these measurements are highly redundant, and relationships among data points predictable, said Roweis. "If I tell you the value of a pixel in an image it is easy to predict the values of nearby pixels since they tend to be similar. Or if I tell you that the word 'corporation' occurs often in articles about economics, employment and environment but not very often in articles about art, poetry and classical music then it is easy to predict that it will not occur very often in articles about love," he said.

Although there are lots of measurements per item, there are far fewer that are likely to vary. Using a data set that only includes the items likely to vary allows humans to quickly and easily recognize changes in high dimensionality data.

These measurements that are likely to vary are referred to as 'degrees of freedom.' Noticing how things like lighting, pose and expression change a person's face across all 360,000 pixels of a picture is both tedious and unnecessary when the face really only has 30 degrees of freedom. "The difference between 30 and [360,000] is what makes data like pictures of faces amenable to dimensionality reduction," said Roweis.



Source: Stanford University

Using the Isomap algorithm, a computer determined feasible paths from initial face and hand positions to end positions, choosing from a very large set of possibilities.

The research teams have each written algorithms that winnow out relationships between data points that, for instance, mean the difference between a face looking up and a face looking down.

The trick is not only finding the important relationships, but preserving those relationships in a lower dimensionality space that is easier to picture and work with. "What's meaningful isn't the property of any one pixel, it is somehow a merged property of the whole object," said

Josh Tenenbaum, an assistant professor of psychology and computer science at Stanford University.

Astronomy data, for example, includes the amount and wavelengths of light stars emit. This constitutes a huge body of data because there many stars, wavelengths and amounts. “You detect maybe 1,000 different wavelengths of light just in a regular optical telescope,” said Tenenbaum. The underlying parameters that astronomers are interested in is a much smaller data set that can be deduced from this data, like an individual star’s gravity, temperature and the amount of metal it contains.

The trouble is, gravity, temperature and metal content may not be related in a simple way to the amount of light reflected or coming out of any one wavelength. “You want to find a few meaningful dimensions that are somehow hidden in this high dimensional mess,” said Tenenbaum.

Tenenbaum’s Isomap algorithm extracts meaningful dimensions by measuring the distance between data points in the geometric shapes formed by items in a nonlinear data set. “Our technique works by [measuring] the intrinsic geometry of the surface. That’s captured... by the distances between points along the surface,” he said. This would be fairly easy using linear data, which can be represented mathematically as a group of planes in a space. Nonlinear data, however is represented by more complicated shapes. The class of shapes Tenenbaum’s algorithm works best for are shapes that can be flattened out, like cylinders or Swiss rolls.

For data that maps mathematically to a roll shape, for instance, Tenenbaum’s algorithm measures the distance between any two points on the shape, then uses these geodesic distances in combination with the classic multidimensional scaling algorithm in order to make a low dimensional representation of that data.

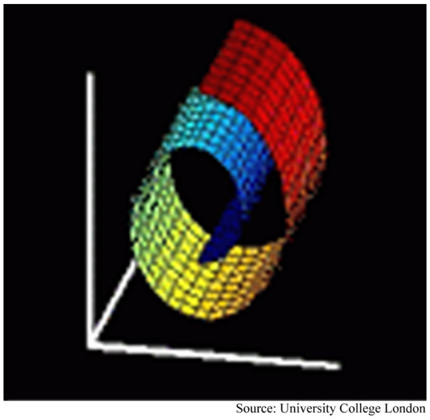
Roweis’ Local Linear Embedding (LLE) algorithm works a little differently. It figures out how each item in a data set is related to the few items nearest to it. It then preserves these neighborhood relationships when the data is converted from its high dimensional form into a low dimensional form.

“Making a low dimensional form of data whose mathematical shape in high dimensional space is a sphere, for instance, is like making a flat wall map of the earth,” said Roweis. “We need to preserve the local relationships between places, and this involves a bit of distortion since we have to flatten... the globe. But if we do the flattening over small areas it is not so bad since the earth is not curved too much over short distances.”

The LLE algorithm chops the sphere into small patches that each do not involve a lot of curve. “Once we flattened each small patch we can reassemble patches in a lower dimensional space, but still trying to preserve the relative positions of things within each patch,” Roweis said.

Neither technique is perfect—they both introduce some distortions of the data, especially for more complicated shapes that include curves. The different approaches may prove to be better or worse for different types of data, said Tenenbaum. “Our approach, based on estimating and preserving global geometry, may distort the local structure of the data. Their technique, based only on local geometry, may distort the global structure,” he said.

The techniques are “an important, useful advance in pattern recognition,” said Javier Pena, an assistant professor of operations research at Carnegie Mellon University. “It is commonplace to have huge databases of information in a variety of contexts. The described techniques attempt to... filter the essence of what these huge databases contain,” he said.



Nonlinear dimensionality reduction algorithms can map the relationships among points in a mathematical surface, like the above Swiss roll, in a lower dimensional space than the initial data set, making it easier to work with.

Roweis is working on taking the LLE algorithm a step further by mapping the difference between the original data space and the low dimensionality set. This mapping could be used to find the low dimensionality data for new data items, to interpolate, or morph new data items between existing items and to extrapolate new data beyond the range of the original data set, according to Roweis.

The Isomap researchers are looking to extend their algorithm to data sets that have intrinsic curvature, to make it work with data sets that have a lot of noise and to make it run faster with data sets that have one million points or more, said Tenenbaum. They are also working with scientists who are using the algorithm on real-world data to see what types of data the algorithm is useful for, Tenenbaum said.

This includes a collaboration with mathematicians who deal with the topology of shapes. “There’s a whole bunch of different mathematical properties of surfaces... topology like holes and connectedness properties. These are all things mathematicians have studied on abstract surfaces for... at least a century. What we’re trying to do here is to study the [mathematical] surfaces that live in real-world data—there’s a lot of open questions like what really are the important geometric and topological features of the surfaces,” said Tenenbaum.

Roweis' research colleague was Lawrence K. Saul of AT&T Research. The research was funded by AT&T, the Gatsby Charitable Trust, the National Science Foundation (NSF) and the Canadian Natural Sciences and Engineering Research Council (NSERC).

Tenenbaum's research colleagues were Vin de Silva of Stanford University and John C. Langford of Carnegie Mellon University. The research was funded by Mitsubishi, the Schlumberger Foundation the National Science Foundation (NSF) and the Defense Advanced Research Projects Association (DARPA).

Both groups published the results in the December 22, 2000 issue of *Science*.

Timeline: < 2 years
Funding: Corporate, Government, Private
TRN Categories: Pattern Recognition
Story Type: News
Related Elements: Technical papers, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science, December 22, 2000. Download of the Isomap algorithm: isomap.stanford.edu

TRN

Software Speeds Gene Comparison

By Kimberly Patch, Technology Research News
February 14, 2001

When biologists want to compare a group of individuals from one species to a group from another species in order to determine how closely the two species are related, they make comparative gene maps, a tedious process done by hand that can take weeks or months.

It takes a long time because of the large number of comparisons that need to be made. For example, to find out how far back corn and rice diverged from a common ancestor, a biologist would compare about 100 markers, or DNA segments, on each of the 10 chromosomes that make up the corn genome to 100 markers on each of the 12 chromosomes of rice. The larger the number of matches, the more recently the plants became separate species.

As a computer problem, it is difficult because simply going through those comparisons by rote, comparing everything to everything else, would result in a number of comparisons so large it would take half a page of space to write the number out: a 12 followed by 1,000 zeroes. Although the hand comparisons are tedious, humans can cut down on the work by intuitively skipping over blocks of comparisons that are not likely to produce results.

A group of Cornell researchers has addressed the problem, however, by using a dynamic programming algorithm similar to those used for analyzing human and programming languages. The algorithm considers all the possibilities without actually going through them by determining what to compare next based on the results it has amassed so far. "What it has looked at so far... will limit what it considers next," said Cornell applied mathematics graduate student Debra Goldberg.

The algorithm produces the comparison results in a few minutes of computer time, Goldberg said.

The quick results may prove especially useful because the markers used to compare chromosomes are from many individuals, and the more samples, the better the comparison. But because going through a full comparison is difficult, new data from more individuals is not incorporated into the maps very quickly. "We felt we [could] do something that would speed up the process and then revise it as more data came in," said Goldberg.

The problem was also a difficult one because the practice of comparing chromosomes is not standardized.

This is because gene mapping is of necessity a subjective process. The genes the process is trying to compare have been shuffled and reshuffled over millions of years. Finding evidence of common markers in different species involves a good bit of detective work.

Chromosomes are very long strings of DNA, the molecules that contain sequences of the four bases that make up the genetic material of life. Genes are specific regions, or sequences of bases, that a cell uses to make certain proteins.

The order of genes on chromosomes can change in several ways. During the course of reproduction, the progeny gets a mix of its parent's chromosomes. During the course of cell division, portions of chromosomes can break off and recombine in the wrong order. Common forces like radiation can damage DNA, causing further changes.

All these variables can make comparisons difficult. For example, if a gene is common to two species, but has also been copied within a single species, "you'll have trouble figuring out how to match them up," said Dannie Durand, an associate professor biological sciences at Carnegie-Mellon University.

Mice and humans both use hemoglobin proteins to carry oxygen in the blood, and these proteins are coded by a certain gene. Humans, however, have at least eight copies of the gene, four of which have mutated into genes that are only turned on while a fetus is in the womb's low oxygen environment. "You can see how if you didn't have a complete dataset you might think you found two analogous genes when you haven't," said Durand.

For their algorithm, the Cornell researchers decided on a set of standards. This is likely both a plus plus and a minus. One of the reasons gene maps are not standardized is biologists studying different organisms care about different aspects of the data. A given approach "may work very well for some types application problems but not others," simply because the scientists are asking different questions about very different species, said Durand.

The Cornell researchers program has just one variable, which allows biologists to choose how closely they would like groups of bases to match in order to be called a match. "We came up with some simple rules that are general enough that they're not going to violate any biologists sense of what is correct," said Goldberg. The idea, she said was to balance accuracy with parsimony.

Parsimony requires not changing things constantly. "We don't want to be flip-flopping back and forth for every gene in our labeling," said Goldberg.

It is tricky to know exactly what means what in gene comparisons. For example, a single instance of a certain portion of maize chromosome six matching a certain portion of rice chromosome three may be misleading, Goldberg said. "Perhaps this is evidence of an ancient linkage group, but just a single marker is not enough evidence that we would want to commit to such a thing. There're many other ways that any one gene might have a match in another genome, so we don't want to assign too much weight to any one piece of data," she said.

At the same time, if there are several instances in comparisons among individuals of different species where six markers in a row match, a seventh marker does not match, and then the next four markers match, it may be enough evidence to call the region as a whole a match. This is because events like mutations could have independently changed a single marker on one of the plants after they diverged from a common ancestor. Decisions like these "can be thought of as a smoothing function," said Goldberg.

In some ways the algorithm makes for simpler comparisons than the more complicated hand produced comparison maps.

On the other hand, because they're consistent, comparison maps produced by the algorithm can be quickly, easily and consistently compared to each other. "A formalized set of rules [means research] groups can compare different comparative maps," said Goldberg.

The algorithm may also allow comparisons in areas where it is too costly to do them otherwise. The data comparing humans and mice, for instance, has been looked at thoroughly because of its importance in human medicine. "But in many other species we don't have the resources, the money or biologists' time," Goldberg said.

The researchers are working on providing a better interface to the program that will allow researchers to more easily extract data from the comparison by asking questions, said Goldberg.

The algorithm will be available later this year for researchers to use, she said.

Goldberg's research colleagues were Jon Kleinberg and Susan McCouch of Cornell. They presented their research at the Plant and Animal Genome Conference in San Diego, January 13-17, 2001. The research was funded by the National Science Foundation, the Packard Foundation, the U.S. Department of Agriculture (USDA), The Cooperative State Research Education and Extension Service, the Alfred P. Sloan foundation, and the Office of Naval Research (ONR).

Timeline: Now
Funding: Government, Private
TRN Categories: Applied Computing; Data Structures and Algorithms
Story Type: News
Related Elements: Technical paper, "Automated Comparative Mapping," presented at the Plant and Animal Genome Conference in San Diego, January 13-17, 2001; Technical paper, "Algorithms for Constructing Comparative Maps," presented at the Gene Order Dynamics, Comparative Maps and Multigene Families (DCAF) conference in September, 2000 in Sainte-Adèle, Québec. The paper is posted at www.cam.cornell.edu/~debra/research.html

Software Cross-Sorts Gene Data

By Kimberly Patch, Technology Research News
November 22, 2000

As anyone who uses the Internet knows, there's a vast difference between simply gathering information, and organizing it so that you can actually find anything. Analyzing large amounts of genetics or financial data presents and even larger problem: teasing patterns from reams upon reams of data.

One popular method that allows for quicker analysis of large groups of data is clustering, which organizes data into groups with similar traits.

Researchers from the Weizmann Institute have taken the process a step further with coupled two-way clustering (CTWC), an approach that identifies two types of subsets in the data, then uses one to cluster the other.

CTWC works with any clustering algorithm. For their demonstrations, the researchers used it in conjunction with their Super Paramagnetic Clustering algorithm, which clusters data based on the way grains of magnetic materials organize themselves into magnetic clusters as they cool.

The CTWC method allows the clustering algorithm to zero in on relevant data and spend less time chunking away on data that has less in common. "We were looking for ways to reduce the problem and break it up into its constituent parts," said Eytan Domany, a physics professor at the Weizmann Institute.

In one demonstration that involved genetic data, for example, "we were able to zero in on small groups of genes whose expression levels are able to distinguish between particular subgroups of tissues," Domany said.

The method also allows for clusters based on similar traits, even when the traits were not identified beforehand. "We want to find also new partitions — say of two different kinds of tumor — of whose existance we were unaware before the experiment," said Domany.

For example, in a demonstration that analyzed gene participation in colon cancer, the method was used to analyze expression levels of thousands of genes for 40 tumor samples and 22 normal tissue samples.

This type of data can be clustered in two basic ways, said Domany. The first way is according to gene expression profiles across the genes, like tumor versus healthy, or different kinds of tumor. It can also be organized into groups of genes whose expression is strongly correlated. The second type of cluster is important because such groups of genes could belong to the same biological mechanism that may have caused the disease.

To identify new clusters the researchers "first use all [tissue] samples to divide the genes into clusters, and all genes to partition the samples into groups. Now we take [the clusters] of genes, and use only its members... to partition every group of samples into subgroups," said Domany.

"The procedure gives us numerous partitions of the genes and of the samples. We look at them and check whether they are... statistically significant and... biologically meaningful," he said.

In one experiment, the researchers first clustered the genes using both tumor and normal tissue samples, then using the tumor tissues only. The two resulting clusterings each had two similar gene clusters, but only in the tumor tissue sample were the expression levels of the two clusters strongly correlated.

This showed that colon cancer was more likely when people had both types of genes, which is known to be the case.

Using the same data, the researchers found a cluster of genes which parsed the 62 tissue samples into two groups, each containing both tumor and normal tissues. The two groups were distinct in a different way, however. "We discovered that at a certain date the experimental protocol had been changed," said

Domany. "Nearly all the tissues in one of our groups have been measured before this date, and those of the other group after this date. Hence we discovered a set of genes whose expression levels are sensitive to the changing measurements induced by the change of protocol," he said.

The algorithm is still being tuned, but the researchers plan to have a version available to download from their web site within two months, according to Domany.

The next step is to apply the algorithm to two more types of data, said Domany. "We are applying the algorithm for document classification [and also] plan to look at financial data," he said. In addition, the researchers are planning to use the existing algorithm on more gene expression data. They are also taking steps towards commercializing the algorithm, Domany said.

The current emphasis in clustering algorithm research is improving their performance for very large data sets, said Johannes Gehrke, an assistant professor in the computer science department at Cornell University.

"Algorithms for very large data sets have to be scalable, which means that their running time has to increase about linearly with the number of records in the input data set. Many existing algorithms... just take too much time to run on today's data sets, even with the speed of computers increasing according to Moore's Law," he said.

Researchers are also beginning to turn their attention to making clustering algorithms work on the fly.

"One challenge is the mining of high-speed data streams," said Gehrke. "As an example, Yahoo had about 450 million page views per day in December 1999 and about 680 million page views per day last month. This amounts to 6 gigabytes of clickstream data per hour. Given this enormous flood of data, we need to develop stream data mining algorithms that can digest these rivers of data just by looking at each record exactly once, in the order the records arrive," he said.

Domany's research colleagues were doctoral students Gad Getz and Erel Levine. They published their research in the Oct. 17 issue of the Proceedings of the National academy of Sciences (PNAS).

The research was funded by the U.S.-Israel Binational National Science Foundation, the Germany-Israel Science Foundation, the Ministry of Science and the Minerva Foundation.

Timeline: < 2 months
Funding: Government
TRN Categories: Data Structures and Algorithms
Story Type: News
Related Elements: Technical paper, "Coupled Two-way Clustering Analysis of Gene MicroArray Data," Proceedings of the National Academy of Sciences (PNAS), Oct. 17, 2000; Site where algorithm will be available in early 2001: www.weizmann.ac.il/physics/complex/compphys/ctwc

TRN

Multimedia

Software Sorts Video Soundtracks

By Chhavi Sachdev , Technology Research News
November 28, 2001

The theme music for the nightly news and the newscaster's voice sound inherently different to us, but distinguishing between the two is not an easy trick to teach a computer.

When it comes to cataloguing and indexing, however, computers are much faster than humans. To use this indexing speed for audio in a way that's practical, computers must be able to distinguish a human voice from a saxophone, and the tympani in the fourth movement of Beethoven's sixth symphony from real thunder.

Scientists at Microsoft Research have come up with an algorithm that allows computers to differentiate among speech, music, environmental sounds, and silence in video soundtracks by mapping and comparing the characteristics of each type of sound.

Classifying sound as noise, speech, or music is an important key to coding audio data, said Hong-Jian Zhang, a senior researcher and assistant managing director at Microsoft Research in China. "Audio segmentation and classification is [the] first step in audio content analysis and content-based [searching] in large audio or music databases," he said. It can also help group together segments of speech by a particular person.

To identify sound, the algorithm goes through a two-step process, Zhang said. It analyzes each audio clip in one-second sections, then further divides this window into forty, 25-millisecond snippets. Using pattern classification, it characterizes each 25-millisecond frame as either speech or non-speech, then further classifies non-speech sounds into silence, music and environment sounds.

The researchers set thresholds to allow the algorithm to distinguish the sound of a slamming door from a drum roll.

Below a certain level the sound is classified as environmental noise and above it, as music.

The most difficult task the computer faces is sorting mixed sounds, said Zhang. Speech with a noisy background, for instance, is frequently classified into music, and music clips with drum sounds are falsely pegged as speech, he said.

Despite the overlaps, the algorithm has a 96-percent accuracy rate in classifying audio data, according to Zhang. When the algorithm does not go the second step of breaking sounds into 25-millisecond components, it has a 91.47 percent accuracy rate in classifying 2- to 8-second clips, he said.

This whole area of research is very exciting, said Dan Ellis, an assistant professor of electrical engineering at Columbia University. "Automatic classification and segmentation of [sound] could make searching for a particular example or instance orders of magnitude easier," he said.

Segmentation algorithms like Microsoft's "could be used as a basis for more complex classifications, for instance to distinguish between news and sitcoms by the patterns of alternation between speech and music, or to detect commercials based on their particular soundtrack signatures," Ellis said.

There are generally several things going on at once in soundtracks, said Ellis. The sound in most video "doesn't fit very comfortably into the black-and-white classification paradigm assumed by this and similar work. Ultimately, we will need computer soundtrack analysis systems that can describe segments with more nuance, for instance as 'male speech with orchestral music in the background'," he said.

The system can be used now for simple classification, but the researchers plan to expand the classification categories to make it more accurate, said Zhang. It should be in practical use within three years, he said.

Zhang's research colleagues were Lie Lu and Hao Jiang at Microsoft Research in Beijing. They presented the research at the International Conference on Multimedia held between September 30 and October 5 in Ottawa, Canada. The research was funded by Microsoft.

Timeline: 3 years
Funding: Corporate
TRN Categories: Multimedia
Story Type: News
Related Elements: Technical paper, "A Robust Audio Classification and Segmentation Method," at the International Conference on Multimedia of the Association for Computing Machinery (ACM) in Ottawa, Canada, September 30 - October 5, 2001

TRN

Speech Recognition to Sort Holocaust Tapes

By Kimberly Patch, Technology Research News
October 31, 2001

When Steven Spielberg established the Shoah Foundation to record eyewitness accounts of Holocaust survivors and rescuers seven years ago, speech recognition software that took dictation was barely usable.

Now, after videotaping 52,000 eyewitness accounts in 57 countries and 32 languages, the foundation is looking to speech recognition software — which has also come a long way in the past seven years — to help with the arduous task of indexing the 116,000 hours of interviews.

The foundation is currently indexing the material manually according to a thesaurus of keywords. "Annotators mark down... codes from the thesaurus as they watch the interviews," said Bill Byrne, an associate research professor of electrical and computer engineering at Johns Hopkins University.

The process is very time-consuming: it would take 40 years of 8-hour days to simply watch the entire collection. "It's also difficult to determine beforehand how to annotate the data so that subsequent searchers can find exactly what they're looking for," he said.

Teams of researchers from IBM, Johns Hopkins University, the University of Maryland, and the Shoah Foundation will take several approaches over the next five years in an attempt to automate the process and make the material more accessible to historians and teachers, said Byrne.

"We hope to be able to use speech recognition and a cross-lingual information retrieval technique to both speed up the annotation so it will be easier for the skilled translators to annotate and also to, at some point, make it possible for people to be able to search these data collections directly without the need of human annotation at all," said Byrne.

Current speech recognition software, which works fairly well for a single trained user, is still not up to the task of transcribing from tape emotional testimony from many users in many languages. The nature of this job makes an excellent research project, however, said Byrne.

Speech recognition systems work well when people are speaking specifically to be understood, like dictating directly to a computer or professionally announcing the news, said Byrne. This is why the real-time speech translators used in loud bars to subtitle news or sports broadcasts work fairly well.

In contrast, in the Shoah foundation material, "people are speaking to an interviewer... and their speech is highly emotional and about topics that are something out of the general realm of experience. They're heavily accented in the English collections. And the speakers are also elderly. Children and elderly people [have] a lot more variability in their speech, [which] makes it hard to recognize as well," he said.

Another challenge is the acoustics. In contrast to newscasts, the videotaping "was not done in a sound booth... there's just a microphone in the camera several feet away from the speaker," said Byrne.

It's a difficult project, said Alex Waibel, a professor of computer science at Carnegie Mellon University. "The biggest challenges are that the recorded speech is conversational, not read, and therefore presents greater variability, leading to higher error rates, [it is in] multiple languages, [and it involves the] expression of emotion, which makes recognition harder."

Usually speech recognition systems address the multiple language problem by individually training recognizers for each language, said Waibel. The project is an obvious fit for an alternative approach that has already shown some promise — multilingual speech recognition models, he said.

Multilingual models proposed five years ago by Carnegie Mellon and University of Karlsruhe researcher Tanja Schultz showed that a multilingual translator can do as well as the approach that uses multiple translators for individual languages, said Waibel.

The fictional Star Trek universal translator presages this approach. It uses a speech model that "can essentially be used for any new language with little adaptation data," Waibel said.

The researchers are taking several different tacks, said Byrne.

The IBM researchers are adapting an English translation module using 100 hours of tape from the collection that will be transcribed by people, essentially giving the module the answers to the first 100 hours of words. It is not possible to do this much work with each of the 32 languages, however, so the researchers will next use about 20 hours of translated Czech to adapt the Czech module, said Byrne. "We're going to see if we can develop techniques that allow us to train systems with much less data," he said.

Speech recognition is just part of the project, he added. Its goal is finding information, and the speech recognizers will be embedded in a much larger search and retrieval system, he said. The idea is to "make this data usable by historians and educators and teachers... they're going to want to search through the material to find discussion of certain events or themes... related to their research or the classroom material," he said.

The advantage of this goal is "the speech recognition systems don't need to work perfectly to be useful for searching archives. Returning good answers to the user's query [is] what we're really after," he said. The researchers will also concentrate on retrieval lexicons, which are lists of words used by search engines. "We will try to make sure that we do a very good job on these words, because these are the words [search engines are] looking for," he said.

The general plan is the Maryland researchers will work on information retrieval and interaction with users; the Johns Hopkins researchers will work on speech recognition and the problems of working with multiple languages; the IBM researchers will focus on the project of transcribing English; and the Shoah foundation researchers will concentrate on cataloging and adapting the approaches to their specific needs, said Burns. "All these efforts fit together tightly.

The project is scheduled to last five years. Improved access to the Shoah foundation archives is likely to be available sooner, said Byrne. "We could start seeing initial results from the effect of our work within a year or so," he said.

Byrne's research colleagues are Frederick Jelinek, Sanjeev Khudanpur and David Yarowsky from Johns Hopkins University; Douglas Oard, Bruce Dearstyne, David Doermann, Bonnie Dorr, Philip Resnik and Dagobert Soergel of the University of Maryland; Bhuvana Ramabhadran and Michael Picheny from IBM T. J. Watson Research; and Sam Gustman, Douglas Greenberg and Ella Thompson of the Survivors of the Shoah Visual History Foundation. The research is funded by the National Science Foundation (NSF).

Timeline: 5 years
Funding: Government
TRN Categories: Human-Computer Interaction
Story Type: News
Related Elements: None

Image Search Sorts by Content

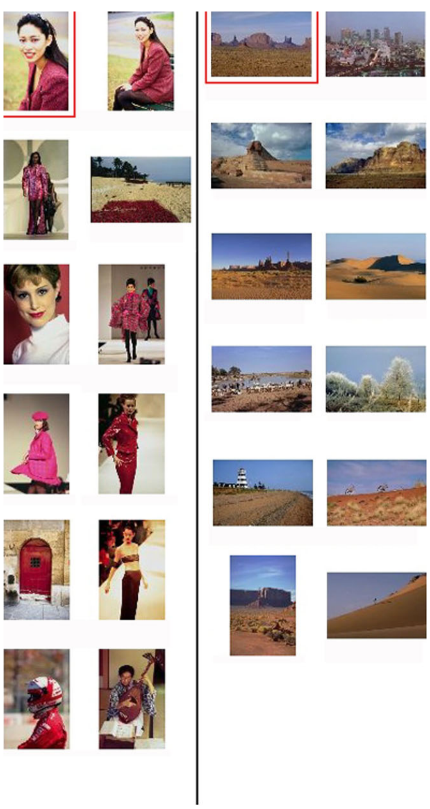
By Chhavi Sachdev, Technology Research News
September 26, 2001

If you've ever tried to search for a picture on the web, you know it's a hit-and-miss affair. You type in some words, but only retrieve those images that contain the search terms in their meta tags or file names.

Some digital image libraries, on the other hand, let users search on the content of the images themselves. One problem with these libraries is keeping searches manageable across thousands of images.

A group of researchers at Pennsylvania State University has come up with software that maps images' key features and assigns the images to several broad categories. The Semantics-sensitive Integrated Matching for Picture Libraries (SIMPLIcity) system retrieves images by matching the features and categories of a query image to those of images stored in the database.

The system could cut the time and expense involved in sifting through large databases of images in biological research, said James Z. Wang, an assistant professor of Information Sciences and Technology and Computer Science and Engineering at Pennsylvania State University. It could also be used to index and recover images in vast museum and newspaper archives, he said.



Source: Pennsylvania State University

An image-based search returned these sets of photographs. The image in the upper-left corner of each column is the query image. The other images were found by comparing their features and subjects to the query image.

To use the program, a user gives it a query image or image URL. The program indexes the image by converting it to a common format and extracting signature features, such as color, texture and shape of certain segments. These features are stored in a features database while the whole image is relegated to a system database.

Images are also classified semantically - as graph or photograph, textured or non-textured, indoor or outdoor, and objectionable or benign, he said. "We first extract features. Then we use semantics classification to classify images into categories. Then within each category, we retrieve images based on their features," Wang said.

After feature extraction and classification, the program "can tell that a picture is of a certain semantic class, such as photograph, clip art, indoor, [or] outdoor," he said. Although it is still impossible for it to tell that the picture is about a horse, given a picture of a horse, the program can find other images with related appearances, he said.

It does this by matching the most similar areas and features of an image and comparing the remaining areas in the query image. In this way, all the areas of a query image are considered and the similarity of the query image to the database images is based on the entire image.

"An image with 3 objects is like a set of 3 points, each with a significance, in the feature space. Now the question is how to match two sets of points," said Wang. A region-matching algorithm uses assigned significance of features in each region to do this, he said. The program's matches remain consistent even when query images are rescaled or rotated, he said.

It takes a couple of seconds to put all the images in a database of 200,000 images in order, Wang said. "Based on the interface, the system can provide a collection of best matches," he said. According to Wang, the image retrieval system is more accurate and substantially faster than others available today.

The image retrieval system is one of several to have emerged in recent years, said Tomaso Poggio, a professor of Brain Sciences at the Massachusetts Institute of Technology. Although the system may be an improvement over previous systems, "I do not see any major breakthrough," he said. "The intermediate semantic level makes sense but the list of semantic categories is arbitrary [from] what I can judge. It may be interesting to try to ground it in … studies of human subjects… to evaluate whether people do the task based on semantic level categories," he added.

The program is currently used in several universities, Wang said. "Most of them are using SIMPLIcity to search for stock photos, pathology images, and video frames," he said. The system could be in wide use within 10 years, he said. It is not likely to ever replace human skills completely, especially with much bigger image collections, he said. "I would not know how long it will take to develop a system suitable to search hundreds of millions of images" such as on the Internet, he said.

The researchers are continuing their search for more efficient image retrieval techniques, said Wang. "At the same time, we plan to apply our systems to domains such as biology and medicine."

Wang's research colleagues were Jia Li at Penn State University and Gio Wiederhold at Stanford University. They published the research in the IEEE Transactions On Pattern

Analysis And Machine Intelligence, September 2001 issue. The research was funded primarily by the National Science Foundation (NSF).

Timeline: >10 years
Funding: Government
TRN Categories: Computer Vision and Image Processing; Databases and Information Retrieval; Pattern Recognition
Story Type: News
Related Elements: Technical paper, "SIMPLIcity: Semantically-Sensitive Integrated Matching for Picture Libraries" in IEEE Transactions On Pattern Analysis And Machine Intelligence, September 1; "Scalable Integrated Region-based Image Retrieval using IRM and Statistical Clustering," presented at the Joint Conference on Digital Libraries (JCDL '01), Roanoke, VA, June 24-28, 2001

TRN

Computer vision

Vision Chip Shines

By Eric Smalley, Technology Research News
September 10/17, 2003

The video cameras and complicated image processing software that are used to give machines the ability to see are relatively bulky and expensive. Many research teams are working toward a better solution — eyes-on-a-chip.

Researchers from the State University of New York at Buffalo and Stanford University have built a silicon retina that uses a timing signal to mimic a form of data compression performed by biological eyes, and transmits high-speed optical rather than electrical output.

The silicon retina could be used to give small robots a better understanding of their visual environment, according to Albert Titus, an assistant professor of electrical engineering at the State University of New York at Buffalo. The electronic retina could also be used in smart sensors and remote monitoring cameras, where its ability to sort out important information would allow reduced amounts of data to be analyzed, transmitted and stored.

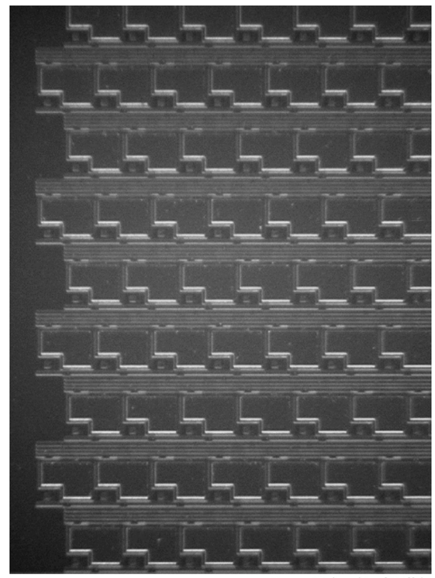
Like its biological forerunners, the electronic retina processes the larger amount of data that makes up an image in order to transmit a smaller amount of key information. The silicon retina provides information about the edges of images rather than a whole picture. Edge information is usually sufficient for detecting and tracking objects.

The device's pixels are an array of light detectors made from metal oxide semiconductor. The array takes in an image, processes the information, and passes the compressed output to a liquid crystal spatial light modulator on the chip. Spatial light modulators pattern light, in this case allowing it through in positions corresponding to pixels that generate an electrical "on" signal.

The spatial light modulator enables output from each receptor, or pixel, to be transmitted optically, which allows the process to take place nearly in real-time. "Optical output... allows for maximum parallelism of the data output, and requires no wires to send the data," said Titus.

Otherwise, with electrical output, an array of 4,096 pixels would require 4,096 output wires from the chip, or it would have to slow down the process by sending more than one output per wire, said Titus. "These are both unrealistic approaches," he said.

Biological retinas use two types of light receptor cells — rods and cones — to convert optical energy into electrochemical responses that can be processed by nerve cells. Cones are sensitive to color, and work best in bright light. Rods allow for vision in dim light.



Source: University of Buffalo

This chip is an artificial retina. The larger rectangles are mounts for tiny liquid crystal light modulators that control the chip's optical output. The smaller rectangles are light detectors.

Three other types of cells — amacrine, bipolar and horizontal — work together to share signals between receptors, and to transmit the signals to nerve cells. The result is the light pattern the retina picks up gets transformed, or filtered, into a more concise set of information for the ganglion cells that make up the optic nerve, said Titus. "There are anywhere from 10 to 1,000 [times] fewer ganglion cells than receptors, so there is a significant amount of data compression that occurs between the light input and what is transmitted to the brain," he said.

One form of retinal data compression is "a response that corresponds to the edges of objects," said Titus. "If you break an input image into objects represented by just their edges — changes in intensity — then you remove a lot of information from the image, but you still have quite a bit of information about the scene," he said.

The researchers' design models the function of the receptors and the bipolar, amacrine and horizontal cells, said Titus. "Our silicon retina produces information about the edges and performs edge enhancement based on motion," he said.

Edge detection is a common capability of artificial retinas. The researchers' design is unique because it uses a clock signal to synchronize the pixels, which allows the chip to work efficiently, according to Titus. A pixel in the artificial retina is about 10 times faster than a photoreceptor in a biological retina, so it can perform several operations for every photoreceptor operation, said Titus. This helps the artificial retina perform edge detection using a relatively small number of pixels, he said.

The chip also draws very little power. Each cell requires less than one ten thousandth of a Watt to turn on and off at speeds of a few kilohertz, or thousand times a second.

The researchers have built a prototype that contains 256 pixels, and are working to make a more complete silicon-based system that can be used in autonomous robots and smart sensors, said Titus. They're also aiming to use the silicon retina in cameras for remote monitoring for safety, identification and biometrics purposes, he said.

The researchers are also working on artificial retinas that do more than just edge detection, said Titus.

The silicon retina could be used in practical applications within one to five years, according to Titus. Applications using optical output will have to wait 10 years or so until optical interconnects are available for interchip communications, he said.

Titus's research colleague was Timothy J. Drabik. The work appeared in the August, 2003 issue of *Optical Engineering*. The research was funded by Displaytech, Inc.

Timeline: 1-5 years, 10 years
Funding: Corporate
TRN Categories: Computer Vision and Image Processing; Optical Computing, Optoelectronics and Photonics
Story Type: News
Related Elements: Technical paper, "Optical Output Silicon Retina Chip," *Optical Engineering*, August, 2003

TRN

Computer Follows Video Action

By Eric Smalley, Technology Research News
December 5, 2001

Teaching a computer to watch video is a lot harder than it might seem. What is obviously a face or a car or a sunset to us is often a confusing blur of shadows, colors and shapes to the computer.

Although computers are good at recognizing light, dark and color, and they do a reasonable job of picking out objects in still images, they get lost when objects move around. This is because the appearance of a moving object is constantly changing.

The key to giving computers a decent shot at following the action in a video is getting them to track the objects.

A team of researchers at the University of Amsterdam in the Netherlands has come up with an object-tracking technique that works by capturing the essence of a video object rather than just relying on its shape.

The underlying assumption of most video object tracking research, which focuses on the position and pose of an object, is that the object's appearance does not change much, said Marcel Worring, an assistant professor of computer science at the University of Amsterdam.

In contrast, the researchers' approach makes a full assessment of an object by monitoring every point, or pixel, associated with it. This makes it easier for the computer to recognize that a face turned upward in the sun is the same a moment later when it is angled down and in the shade.

"Our contribution is to... track the changes in appearance of every part of the object," said Worring. "In practice this means that we track changes in pixel values that belong to the object." Those pixel changes can be due to changes in lighting and the orientation of the object, and whether the object appears to have moved closer or farther away, he said.

By keeping track of the state of each pixel that makes up an object, the system is also better able to keep tabs on the object when it momentarily disappears, said Worring.

Like other object-tracking techniques, the researchers' system uses a cycle of measuring an object, predicting where the object will be next, using the prediction to narrow the area to be measured, and using the measurements to improve the prediction. When the measurements are completely off from the prediction, it usually means the object has disappeared from view.

"The system detects that the object has been lost by observing that the prediction becomes totally useless," said Worring. "It then... keeps doing a search in the neighborhood of the old position till the object reappears. At that point the tracking continues." Tracking the whole of an object rather than just its shape makes the researchers' system more efficient at recognizing when an object has disappeared, which gives it a better chance of picking up the object when it reappears, he said.

The system could be used for creating video hyperlinks, tracking faces to select the best view for face recognition, and tracking cars, said Worring. "The current system has to be tested thoroughly, but we expect it to be ready for practical use within a year," he said.

The algorithm is useful, said Aggelos K. Katsaggelos a professor of information technology at Northwestern University. "It's a nice combination of template matching and temporal prediction."

Worring's research colleagues were Hieu T. Nguyen and Rein van den Boomgaard of the University of Amsterdam. They presented the research at the International Conference on Computer Vision in Vancouver, July 9 to 12, 2001. The research was funded by the Dutch Organization for Scientific Research.

Timeline: < 1 year
Funding: Government

TRN Categories: Computer Vision and Image Processing
Story Type: News
Related Elements: Technical paper, "Occlusion Robust Adaptive Template Tracking," International Conference on Computer Vision, Vancouver, July 9 to 12, 2001

Computers Sort Gender in a Binary World

By Ted Smalley Bowen, Technology Research News
January 30, 2002

In some respects, computers have helped mitigate the significance of gender in society. The Internet, for instance, gives people control over whether to reveal their gender.

At the same time, however, computers can sort people by gender by comparing faces or voices to a database of features or voice samples. But today's practical applications have weaknesses. Off-center images of faces can be hard to interpret, as can crowd shots. Ambient noise can make it difficult to decipher voice samples.

A group of researchers at The Pennsylvania State University are using a type of pattern recognition software to determined the gender of both faces and voices, then merging the data to produce more accurate results.

Support vector machines (SVMs) are a type of computer learning system that can be trained to screen for certain data in order to make a given classification. They analyze data by comparing the information to a pair of previously defined choices, such as the sexes.

SVMs have been used to identify gender using images of faces, but not voice clips. Today's methods for identifying the gender of voices are generally less sophisticated than those for facial ID.

To test their scheme, the Penn State researchers trained their system to screen thumbnail images of faces and voice samples for gender characteristics. They then presented it with a separate set of pictures and voices, which the SVMs designated male or female. Finally, they merged the image and voice results.

The twice-sifted results had a 95-percent accuracy rate, according to Rajeev Sharma, associate professor of computer science and engineering at Penn State.

The researchers' multi-modal, multi-stage learning scheme is generic enough to apply to other decision fusion scenarios as well, said Sharma. "It involves first building classfiers for each of the two modalities separately, followed by a separate learning stage in which the fusion of decisions is learnt. This creates a robust decision fusion from two disparate sources," he said.

The researchers have tested the method with head-on, static images and sound clips that are free of background noise.

The method is fairly easy to implement, according to Sharma. It calls for basic audio visual equipment and a computer to analyze the data. The system can handle images that are rotated as much as about 20 degrees, Sharma said.

To train the face-screening SVM, the researchers used 1,056 facial images from 600 20- by 20-pixel thumbnail pictures and their mirror images. The researchers culled the images from several databases.

The group then trained a speech-classifier SVM with 300 voice samples derived from a spoken alphabet database dictated by 150 male and female subjects.

The researchers boiled the training material down to 147 image and 147 voice samples. They grouped the results into two-dimensional matrices, and used 47 of them to train the fusion SVM and 100 for testing it.

A Penn State spin-off venture plans to commercialize the technology for market research in about six months, according to Sharma.

Eventual uses could include applications that tailor digital content based on gender in a variety of settings, including information kiosks, according to Sharma.

The work could potentially have widespread applications, according to Jeffrey Cohn, associate professor of psychology and psychiatry at the University of Pittsburgh. "Men's and women's faces differ in both local features and in shape. The Penn State algorithms appear to capture and represent [the appropriate] features. They also include vocal parameters when available. By combining these types of multi-modal data in a classifier, they potentially can achieve robust discrimination," he said.

At the same time real-world conditions could derail the applications, Cohn added. "The question is under what range of parameters can the system perform satisfactorily. Technical challenges include pose, image resolution, occlusion, number of individuals in the image, and image complexity. Sun glasses, for instance, foil face recognition algorithms, and may do the same for gender recognition."

According to Sharma, the system should function well with voice recognition systems, which require relatively unadulterated sound. The group has yet to test the system's tolerance of extraneous, ambient noise in real world situations, he added.

Sharma's research colleagues were Leena Walavalkar and Mohammed Yeasin. The research was funded by the National Science Foundation and Penn State.

Timeline: 6 months
Funding: Government, University
TRN Categories: Pattern Recognition; Computers and Society
Story Type: News
Related Elements: None

Quantum Software Gets the Picture

By Eric Smalley, Technology Research News
September 4/11, 2002, 2002

When you look at a tile floor, you may think about how well the pattern goes with the rest of the room, but you won't wonder whether there is a pattern there in the first place.

A computer, on the other hand, would have a hard time simply figuring out that a black tile followed by a white tile followed by a black tile followed by a white tile constitutes a pattern.

It is clear that quantum computers, which use the quirks of quantum physics to compute, will be orders of magnitude more efficient at many tasks than ordinary, classical computers, if and when sufficiently large quantum computers can be built.

A physicist at the University of British Columbia has come up with an algorithm that proves that quantum computers would be faster at finding patterns, too. "Finding and recognizing [a linear] pattern can be accomplished much faster on a quantum computer than on a classical one," said Ralf Schützhold, a researcher at the University of British Columbia.

The algorithm would allow quantum computers to detect an 8-by-8 grid of alternating black and white squares set in an array of 640 otherwise randomly distributed squares.

This seemingly simple task takes a classical computer about 6,000 steps because it would have to compare each square to every other square, one at a time.

A quantum computer, however, can examine all of the possible solutions to a problem at the same time, in this case comparing all the squares to each other at once. The algorithm proves that this particular task can be represented mathematically in a way that a quantum computer can carry it out.

Quantum computers can check all solutions at once because they use atoms or subatomic particles to make quantum bits, or qubits. The particles have two opposite orientations that can represent the 1s and 0s of computer information.

The power of a quantum computer comes from the quirky physics of these tiny particles. When a particle is isolated from its environment it is in the weird quantum state of superposition, meaning it is in both orientations at once, and so can represent a mix of 1 and 0. This allows a string of particles in superposition to represent every combination of 1s and 0s at the same time, and a quantum computer to process all the numbers that represent possible solutions to a problem with one set of operations.

The pattern-finding algorithm is an addition to a growing set of quantum algorithms based on the quantum Fourier transform, a mathematical formula for finding order.

Other researchers have demonstrated that quantum computers would be exponentially faster than classical computers for pattern-matching tasks like finding a mug shot in a database that matches an image from a security camera. Schützhold's pattern-finding algorithm performs the first task of a pattern recognition application: finding patterns in raw data.

Pattern finding is a key component of speech, face, and handwriting recognition programs, and of software that sorts seismographs and other large sets of scientific data, said Schützhold. The exponential speed-up promised by quantum computers might enable us to attack problems that would take classical computers "longer than the age of the universe" to solve, he said.

This particular algorithm is not likely to be used in practical applications, however. "The problem I discussed is very simple and probably not extremely important or relevant for practical applications," said Schützhold. "My main point is to demonstrate the possible exponential speed-up," he said.

The pattern-finding algorithm is also not a particularly efficient quantum algorithm, said David Meyer, a mathematics professor at the University of California at San Diego. But it is important for demonstrating that quantum computers could be used to speed up image processing tasks, he said. "There are probably other image processing problems for which quantum algorithms will be more successful," he added.

Researchers generally agree that it is likely to take at least two decades to develop practical quantum computers. Quantum computing research is now at a stage comparable to when electrical engineers began to build and combine small numbers of transistors half a century ago, said Schützhold.

Transistors are electrical switches that combine to form the basic logic circuits of computers. Today's PCs have about one billion transistors. Useful quantum computers will require at least one million qubits, the quantum equivalent of transistors. The largest prototype quantum computer built so far had seven qubits.

The research was funded by the Alexander von Humboldt Foundation in Germany and the Natural Sciences and Engineering Research Council of Canada (NSERC).

Timeline: > 20 years
Funding: Private, Government
TRN Categories: Data Structures and Algorithms; Quantum Computing and Communications
Story Type: News
Related Elements: Technical paper, "Pattern Recognition on a Quantum Computer," posted on the arXive physics archive at arXiv.org/abs/quant-ph/0208063

TRN

Security

Text Software Spots Intruders

By Kimberly Patch, Technology Research News
October 30/November 6, 2002

The computer anti-virus programs in common use today use signature detection schemes that can only protect a machine from viruses that have been previously identified and entered into the programs' virus databases.

Anomaly detection systems, however, sense when normal patterns of communications change in order to stop new viruses — or any other system intruders like worms or unauthorized users — in their tracks.

The trouble is, existing anomaly detection schemes all generate high error rates — they cry wolf so often that they are impractical. In order to identify the real intrusions, system managers must spend time checking out every possibility.

Researchers from the University of California at Davis have taken an unusual tack in anomaly detection by adapting text classification techniques to intrusion detection. Their initial results suggest that the technique could produce an anomaly detection system with a reasonable error rate.

The idea to apply text classification to intrusion detection began with a conversation about categorizing Web pages into clusters that share a given property, said V. Rao Vemuri, a professor of applied science and computer science at the University of California at Davis, and a scientist at Lawrence Livermore National Laboratories.

Instead of categorizing Web pages, however, the researchers used the classification system to categorize computer users into just two groups — authorized users and intruders. "The problem is to decide what 'text'" to use for the problem, Vemuri said. We wanted some objective way of characterizing a user that the user... cannot consciously influence" in order to prevent an intruder from fooling the system, he said.

They turned to system calls to characterize a user. System calls are the internal requests various pieces of software make to each other in the course of carrying out a user's instructions. "The system calls are generated by the computer, and the user cannot really influence the sequence in which they are generated," said Vemuri. The scheme treats each system call as a word and each sequence of system calls as a document, and classifies each document as one generated during normal activity or intrusive activity, he said.

The nearest-neighbor text categorization technique the researchers used categorizes Web pages based on how they are linked. The nearest neighbors in terms of links also tend to be closer in terms of content.

The researchers' detection scheme characterizes an authorized user by building a profile of activities. "For example, in the course of my normal life, I use email, browse Web pages, use Word, PowerPoint [and] printers," said Vemuri. "Let's suppose that I rarely, if ever, use Java or C++. I rarely use root privileges. If someone logging onto my machine uses these, that departure from normal usage should signal... abnormal, [possibly] intrusive activity," he said.

The problem turned out to be easier than categorizing Web pages, said Vemuri. "Usually we use many categories. In our example, we have only two categories — authorized or intruder, and in the worst-case three" if the system has to resort to classifying activity as unknown.

In addition, Web pages can be very long and the size of the English vocabulary is around 50,000 words, which makes categorizing Web pages a computer-intensive task. "In our case, the vocabulary — distinct system calls — rarely exceeds 100, and the size of the 'pages', [or groups of calls], is also very small," he said.

Short sequences of system calls have been used before to characterize a person's normal behavior, but this requires building a database of normal sequences of system calls for each program a person uses. The text categorization technique, however, calculates the similarities between program activities, which involves fewer calculations.

This allows the system to detect an intruder as the intruder is affecting the system, said Vemuri. "The computational burden in our case is much smaller, to the extent we started to dream about the possibility of detecting an intruder in real-time," like the way contestants called out titles as songs played on the TV show "Name That Tune", he said.

The researchers' current implementation is almost real-time, said Vemuri. "We have to wait until [a] process, terminates, or halts" before completing the classification, he said. Intrusive attacks, however, are usually conducted within one or more sessions, and every session contains several processes, said Vemuri. Because the classifier method monitors the execution of each process, it's likely that an attack can be detected while it is happening, he said. The researchers are also working on allowing the system to make a classification before a process terminates, he added.

The researchers tested their scheme with 24 attacks within a two-week period. The method detected 22 of 24 attacks, and had a relatively low false-positive rate of 31 false alarms out of 5,285 events, or 0.59 percent, according to Vemuri.

The method shows promise, said Bennet Yee, an assistant professor of computer science and engineering at the University of California at San Diego. "The novelty is noticing that text classification techniques can be adapted to intrusion detection, and doing the experiments that validate it," he said.

If it proves practical and is widely deployed, the technique could help prevent malicious software like the Internet worms Code Red and Klutz, Yee said. "It should be able to recognize new attacks as anomalous behavior and raise alarms earlier [than] signature detection schemes where a database of bad behavior must be compiled first," he said.

There is still work to do to determine if the method can be improved to a low enough false positive rate, however, said

Yee. A practical anomaly detection system must have a very low false positive rate in order to be commercially useful because if system administrators spend too much time chasing down false alarms, they "will not want to use the system and will turn the intrusion detector off," he said.

Even a false positive rate of 0.44 percent could mean 23 false alarms per day if there are 5,285 events per day, Yee said. "Most people will not want to handle a false alarm per hour per machine," he said. The researchers' method is an improvement over earlier anomaly detector designs, but "further improvements are still necessary for broader use," he added.

It is theoretically possible to use the method today, said Vemuri. The researchers are working on proving that the method can be used without raising too many false alarms, he said.

To cut down on false alarms, the researchers are looking to make a redundant system "where we use different methods on different data sets, combine the results of both those methods, or use a best of three voting system," he said. One method could use system call data, for instance, while another could analyze instructions used, he said.

The researchers hope to have their anomaly detection system worked out and supported with performance data within a few years, said Vemuri.

Vemuri's research colleague was Yihua Liao. They published the research in the *Proceedings of the 11th Usenix Security Symposium*, which was held in San Francisco August 5 through 9, 2002. The research was funded by the Air Force Office of Scientific Research.

Timeline: 2-3 years
Funding: Government
TRN Categories: Cryptography and Security; Computer Science; Internet; Networking
Story Type: News
Related Elements: Technical paper, "Using Text Categorization Techniques for Intrusion Detection," *Proceedings of the 11th Usenix Security Symposium*, San Francisco August 5-9, 2002

TRN

Physics Methods May Spot Intruders

By Kimberly Patch, Technology Research News
December 5, 2001

The key to detecting uninvited visitors is recognizing them.

This gets difficult in crowded situations, like large networks, because there is a lot of normal traffic, or noise, that can cover an intruder's comparatively quieter signal. What's even more difficult, however, is detecting a new type of intrusion the first time it happens. Essentially what's needed is a way to detect what you don't know you're looking for.

Researchers from the University of South Carolina have tapped the methods of nuclear experiments to map network traffic and extract patterns of typical network behavior. When scientists looking into the makeup of matter cause nuclear particles to collide, hundreds of detectors monitor every facet of the complicated reaction to capture any slight derivation that may point to an unknown phenomenon.

Analyzing network traffic data this way makes it easier to tease out derivations that point to known network intruders, said Vladimir Gudkov, a physics research professor at the University of South Carolina. "If... almost complete monitoring and data collection [of nuclear events] is possible in physics, why not try to find a way to do similar things in network monitoring?" he said.

The research could also eventually be adapted to the really difficult problem of detecting new methods of intrusion as they are happening, said Gudkov. "We have an opportunity to detect even unknown intrusions in the reconnaissance stage of an attack," he said.

When a file is transmitted over a network it is first broken up into many small packets, which traverse the network using whatever route is available and are reassembled when they arrive at their destination.

To closely monitor a network, the researchers track all the properties of these packets, including how they change over time. Routers, the specialized computers the control traffic around the Internet, put time stamps and other marks on the packets. The advantage of using this time-dependent information is it provides a complete description of the process. "This is exactly what we need for reliable numerical analysis," Gudkov said.

The researchers translate this information into mathematical functions in order to use the complex systems theory that physicists use to extract information from large, changing sets of data, said Gudkov.

The method captures raw data from a network node, then on a separate system plots the mathematical functions in two or three-dimensional imaginary space, and uses pattern recognition to find deviant signals. The result is an "ability to optimize signal-to-noise ratio and to analyze signals in real-time," Gudkov said.

This makes the faint tracks of an intruder more apparent. "The basic idea is to define the normal network behavior using the complete network monitoring. The deviation from the normal traffic behavior will give an alert for possible... intrusions," he said.

In plotting the signals the researchers also found something surprising: some of the ways information flows in these imaginary spaces are independent of how a network is laid out and what system software the computers are running. "This looks natural [to] me now, but some months ago we did not even suspect that... characteristics like the dimension of information flow in the parameter space are... not sensitive to network topology [or] operating systems," Gudkov said.

The researchers are working on a test model of a system that will detect known intrusions as they are happening, said Gudkov. If the research goes as expected, a model for detecting unfamiliar types of intrusions could be available within a year, and a practical working system a couple years after that, Gudkov said.

The researchers are also working on finding a way to detect unfamiliar intrusions by analyzing all the data rather than just looking for known intrusion patterns. The challenge is finding a method of pattern recognition that will work in real-time data plotted in imaginary spaces that have more than three dimensions, according to Gudkov. “The next step for this is the study of multidimensional pattern recognition methods based on wavelet analysis,” he said. Wavelets are a form of compressed data.

The researchers’ idea of modeling network traffic characteristics as functions is an interesting one, but “the question of whether such a view is meaningful, or if it would lead to useful results,” cannot be answered without testing the method on real networks, said R. Sekar, an assistant professor of computer science at the State University of New York at Stony Brook.

It is also difficult to predict whether it will be possible to find unfamiliar intrusions this way, according to Anita Jones, a professor of engineering and applied science at the University of Virginia. “Any mathematical approach depends upon detecting some properties that distinguish the intrusive traffic from normal traffic. Just as in real life, what is harmful can often be masked to appear benign. Such traffic can sometimes be very hard to distinguish from normal traffic,” she said.

Gudkov’s research colleague is Joseph E. Johnson of the University of South Carolina. The research was funded by the Defense advanced research projects agency (DARPA) and the Air Force Research Laboratory.

Timeline: 3 years
Funding: Government
TRN Categories: Networking; Internet
Story Type: News
Related Elements: Technical paper, “New Approach for Network Monitoring and Intrusion Detection,” posted on arXiv physics archive at xxx.lanl.gov/abs/cs.CR/0110019

TRN

Degree of Difference Sorts Data

Technology Research News, April 23/30, 2003

Researchers from the Institute of Applied Physics in Spain have found an efficient way to sort inconsistent sets of data into groups that share some similarity.

Grouping objects that are similar in some way is the first step in many types of data searching and analysis.

Being able to group sets of data that do not necessarily share the same characteristics, such as disparate Internet attacks or incomplete telephone surveys, is more difficult than grouping data whose pieces match up.

Key to the researchers’ method was comparing the edit distance, or the minimum number of elementary edit operations — like deletions, insertions, substitutions — needed to transform one piece of data into another. The method allowed the researchers to sort the data without using a hierarchical structure.

They used the method to classify different types of Web attacks, grouping them by severity. This more fine-grained way of classifying Web attacks should help programmers build more secure software, including better intrusion detection systems and firewalls, according to the researchers.

The method is ready to be used in practical applications, according to the researchers.

TRN

Statistics Sniff out Secrets

By Kimberly Patch, Technology Research News
September 26, 2001

As digitized pictures, audio and text proliferate, people are exploring ways to exploit these media by hiding messages within the information, which leads others to try to detect these hidden messages.

Although steganography — the practice of hiding a secret message in written or audio information — is hardly new, computers and the Web have added a new twist simply because the volume of information that makes up digitized media is so large. This provides for historically large haystacks that easily obscure needles.

A Dartmouth College researcher has found a method that makes it easier to detect hidden messages in digital images, which can contain a megabyte — a string of one million ones and zeros — of information, or more.

Digital images are made up of pixels, or dots of color. Especially with high-resolution digital images that have one million or more different shades of color, it’s easy to hide a message by slightly altering these colors in ways that are imperceptible to the human eye.

In an image that has not been tampered with, however, the information that makes up the image is not simply random. The key to the Dartmouth detection method is creating a statistical profile of the compressed data files that make up natural, or undisturbed images, then checking a given image against the profile, according to Hany Farid, an assistant professor of computer science at Dartmouth. “In order to detect hidden messages in an image we need to start by characterizing the statistics of natural images. The hope, then, is that when a message is hidden in an image, these statistics are disturbed," he said.

When images are compressed so they can be stored as smaller files, the digital information that indicates the color of each pixel is changed into wavelet information. Wavelet mathematics includes functions like spatial position, orientation and scale. Wavelets allow for compression because all the information that makes up a wavelet can be reconstructed from only a portion of that information. An image is compressed by storing only the portion that is needed to reconstruct the whole.

Farid collected two types of wavelet statistics: variations like mean, variance, skewness and kurtosis in the coefficients, or numbers that make up the wavelets, and information about the rate of errors that occur when reconstructing full wavelets from compressed information.

Variance shows how spread out the data is from the mean, or average; skewness shows how evenly distributed the data is on either side of the mean, and kurtosis shows how peaked the distribution of data is around the mean, said Farid.

He then combined the variation and error rate statistics into a vector — a mathematical construction that is like a virtual sculpture with 70 to 100 dimensions rather than the usual three.

By comparing the statistical vector information with the same information in an individual image, Farid was able to tell if the image had been disturbed with a hidden message, he said.

Source: Dartmouth College

These pictures show different aspects of the wavelet mathematicsused to compress images. The fainter images show diagonal information only; they're more ghostly because this particular image has many strong horizontal and vertical lines.

The practice of information hiding, or steganography is related to, but different from cryptography. In cryptography a message is encrypted and then transmitted. If you saw the transmission you wouldn't be able to decipher the message, but you would know the sender and receiver might be trading secrets. The goal of information hiding is to go a step further by camouflaging the transmission entirely, said Farid.

The statistical vector method only detects hidden messages, and cannot read or remove them, but may eventually be adapted to do so, said Farid. "This work cannot obviously be adapted to remove or decipher the hidden message. I do believe, however, that it is possible to do so," he said.

The technique is an extension of previous steganography detection schemes, said Neil F. Johnson, associate director of the Center for Secure Information Systems. "It is potentially useful if the techniques for detection are repeatable," he said.

In addition to determining if there's information embedded in a message, it is also useful for a detection method to identify the steganography technique used to hide the information, said Johnson. Another goal is to be able to extract the embedded information, he said. Steganography tools exist that can do this in at least some cases, he added.

Steganography has many applications, both good and bad, said Farid. "It can be used to protect copyrights in digital media, for unobtrusive military and intelligence communication, covert criminal communication, trafficking of illegal pornography, and for the protection of civilian speech against repressive governments."

An unfortunate side effect of research that reveals hidden messages, is "repressive governments could use this research to limit civilian speech," Farid said. Because of the possible unpleasant applications, "some will be very critical of this research, possibly with good reason," said Farid. "Nevertheless, I believe that the development of these techniques are inevitable and... will lead to better techniques for hiding information, which in turn will lead to better detection schemes and so on. My larger research vision is in authenticating digital media so that [neither] the 'good-guys' [nor] the 'bad-guys' will... be able to manipulate digital sound, image or video to suit their needs," he said.

The method can also eventually be applied to analyzing works of art to detect forgeries or to determine if more than one artist painted a single painting, Farid said.

The method could be used practically in less than two years, said Farid.

Farid's research was funded by the National Science Foundation (NSF) and the Department of Justice (DOJ).

Timeline: < 2 years
Funding: Government
TRN Categories: Cryptography and Security; Pattern Recognition
Story Type: News
Related Elements: Technical paper, "Detecting Steganographic Messages in Digital Images," posted at www.cs.dartmouth.edu/farid/publications/tr01.html

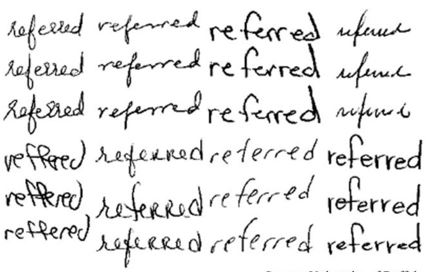
Software Spots Forged Signatures

By Kimberly Patch, Technology Research News
March 21, 2001

Researchers from the University of Buffalo are working on a handwriting analysis system aimed at determining who penned a ransom note or forged a check.

The move was motivated by several high court rulings that required expert testimony to be substantiated by scientific evidence, according to Sargur Srihari, a professor of computer science and engineering at the University of Buffalo and director of the school's Center for Document Analysis and Recognition. "Since individual handwriting had not been subjected to such study, we undertook this work," he said.

The researchers' handwriting recognition program extracts handwriting features like character shape, line and word separation, and stroke slant and thickness. To compare handwriting from different documents, the program computes the differences in these features between the two samples measured, then determines if the differences fall within the limits of an individual writer's variability, according to Srihari.



Source: University of Buffalo

This shows a variety of handwriting styles. Handwriting analysis software compares features like loops and arches in handwriting samples to determine if the same writer wrote both samples.

In tests to determine whether or not two documents were written by the same person, the results were 95 percent accurate, said Srihari. In a second type of handwriting task, where the program determines which member of a known group of writers wrote a certain document, the performance varies from 98 percent for two writers to 89 percent for 1,000 writers, he said.

The program is based on a pair of well-known pattern recognition algorithms: Artificial Neural Network and Nearest Neighbor. It differs from conventional handwriting recognition in a key way, however. "Our work is about determining the variability between writers and within writers, [while] in handwriting recognition, the goal is to identify the message by averaging out the differences between writers," said Srihari.

The researchers' next steps are to capture finer features of handwriting in order to increase the accuracy. "For example, we currently are not measuring very many word-level features such as ascenders or descenders, or the presence of garlands or arcades," Srihari said. Ascenders and descenders are the parts of lowercase letters that extend above or below most other lowercase letters. Garlands resemble circles and arcades, arches.

Capturing features like these should make it possible to boost accuracy past 99 percent, he said.

The research is scientifically thorough, and is certainly useful, but is aiming for a difficult goal, said Nasser Sherkat, an associate professor of real-time machine vision at Nottingham Trent University in England.

"Variability within a writer's [handwriting] is very high, especially when time elements and conditions of writing are taken into account. The ultimate question is whether we can tell if the writer is the same when the constraints [used in the study] have been removed," Sherkat said.

The program can eventually be used as the basis for human experts testimony in court, and eventually as a purely objective handwriting evaluation tool that doesn't require human intervention, said Srihari. This would be useful, said Sherkat. "Checking for forgeries manually is expensive and we could do with automating at least part of the process," he said.

The software could be available in less than a year, said Srihari.

Srihari's research colleague is Sung-Hyuk Cha of the University of Buffalo. The research was funded by the National Institute of Justice.

Timeline: < 1 year
Funding: Government
TRN Categories: Applied Computing
Story Type: News
Related Elements: Technical paper, "Handwriting Identification: Research to Study Validity of Individuality of Handwriting and Develop Computer-assisted Procedures for Comparing Handwriting," downloadable from www.cedar.buffalo.edu/NIJ/publications

Interfaces

Interface Gets the Point

By Kimberly Patch, Technology Research News
January 1/8, 2003

Tone of voice can mean a lot. Your colleague can be giving you a complement or an insult depending on how she inflects the phrase "great work." Gestures can be just as expressive.

Communicating with computers is much more basic. Try to insult an uncooperative speech recognition system by telling it where to go, and, assuming your diction is clear, it will simply show the words on-screen without gleaning anything about your dark mood. Adding an appropriate gesture would make things very clear to even a tone-deaf human, but computers are generally gesture-blind as well.

Researchers from Pennsylvania State University and Advanced Interface Technologies are trying to change that. They are working to untangle the relationships between prosody — the loudness, pitch, and timing of speech — and gestures in an attempt to improve the way computers recognize human gestures.

The research could eventually be applied to many different situations where humans try to get information across to computers, including computer games, surgical applications, crisis management software, and security systems, according to Rajeev Sharma, an associate professor of computer science and engineering at Penn State University and president of Advanced Interface Technologies, Inc.

Although it's child's play for humans, getting a computer to recognize gestures is difficult, said Sharma. Gestures "do not exhibit one-to-one mapping of form to meaning," he said. "The same gesture... can exhibit different meanings when associated with a different spoken context; at the same time, a number of gesture forms can be used to express the same meaning."

In previous work, the researchers analyzed hours of tape of meteorologists giving weather forecasts in order to link prosody to gestures.

The researchers increased their understanding of the phenomenon by plugging speech pitch and hand velocity into the Hidden Markov Model, which breaks information into very small pieces and makes predictions about a given piece of information based on what comes before and after it. The model is commonly used to predict words in commercial speech recognition systems.

The researchers used the system to help detect speech segments that commonly occur along with a particular class of gesture. "We [combined] visual and speech signals for continuous gesture recognition," said Sharma. "The basic idea... is to detect emphasized parts of speech and align them with the velocity of the moving hand."

For instance, a pointing gesture commonly precedes these emphasized segments of speech, a contour-type gesture is more likely to occur at the same time as an emphasized speech segment, and auxiliary gestures, which include preparation and retraction movements, tend not to include emphasized speech segments at all, according to Sharma.

The researchers are using the method in a geographical information system prototype that uses a large screen display, microphones attached to the ceiling and cameras that track users gestures.

The state-of-the-art in continuous gesture recognition is still far from meeting the naturalness criteria of a true multimodal human-computer interface, said Sharma. Computers have achieved accuracies of up to 95 percent in interpreting isolated gestures, but recognizing significant gestures from a full range of movements is much harder, he said.

Taking into consideration prosody when trying to interpret gestures, however, increased the accuracy of gesture recognition from about 72 percent to about 84 percent, Sharma said.

One of the challenges of putting together the system was to define when the visual and audio signals corresponded, said Sharma. "Although speech and gesture... complement each other, the production of gesture and speech involve different psychological and neural systems," he said.

Further complicating things, speech contains both phonological information, which are the basic sounds that make up words, and intonational characteristics, which include some words louder than others and raising the pitch at the end of a question. The system had to accurately pick up changes in intonation amidst the phonological variation in the speech signal, Sharma said.

Modeling and understanding prosody in systems that combine speech and gesture is important in the long run to help transition from a low-level, or syntax-based, to a high-level, or semantics-based understanding of communication, said Matthew Turk, an associate professor of computer science at the University of California at Santa Barbara.

The field has applications in "just about every human-computer interaction scenario, and in many computer-mediated human-to-human communication scenarios [like] remote meetings," Turk said.

The researchers are currently working on incorporating the prosody-based framework into a system to manipulate large displays. The researchers' next step is to run a series of laboratory environment studies to investigate how it works with real people, according to Sharma.

The researchers are ultimately aiming for an environment where a user can interact with the gestures he is accustomed to in everyday life rather than artificially-designed gestural signs, said Sharma.

The system could eventually enable more natural human-computer interfaces in applications like crisis management, surgery and video games, Sharma said.

Another possibility is using the method in reverse for biometric authentication, said Sharma. "This research [could] enable a novel way to identify a person from [a] video sequence... since a multimodal dynamic signal would be very hard to fake," he said.

Understanding how humans and computers can interact using several different types of communication will become increasingly important "as we deal with the need to interact with computing devices... embedded in our environment," said Sharma.

The first products that incorporate the prosody-based system could be ready within two years, said Sharma.

Sharma's research colleagues were Sanshzar Kettebekov and Muhammad Yeasin. The research was funded by the National Science Foundation (NSF) and Advanced Interface Technologies, Inc.

Timeline: > 2 years
Funding: Corporate, Government
TRN Categories: Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "Prosody Based Co-Analysis for Continuous Recognition of Co-Verbal tures," posted at the computing research repository at arXiv.org/archive/cs/intro.html

TRN

Interface Lets You Point and Speak

By Kimberly Patch, Technology Research News
July 25, 2001

One of the reasons speech recognition software remains inferior to human speech recognition is computers can't read hands.

Humans convey a surprising amount of information through the gestural cues that accompany speech. We point things out, convey concepts like "big" or "small", get across metaphorical ideas, and provide a sort of beat that directs conversational flow.

No matter how often or how vigorously you shake your fist at at your computer screen, however, it won't help the computer tune in to your mood.

Researchers from Pennsylvania State University are working on a human-computer interface that goes a step toward allowing a computer to glean contextual information from our hands. The software allows a computer to see where a human is pointing and uses that information to interpret the mixed speech and gestural directions that are a familiar part of human-to-human communications.

These pointing, or deictic gestures are commonly mixed with speech when talking about things like directions, for example, saying "from here to here," while pointing at a map.

The researchers used Weather Channel video to glean a database of deictic gestures, which include directly pointing to something, circling an area, or tracing a contour. "Looking at the weather map we were able to classify pieces of gestures, then say which pieces we can interpret, and what kind of gestures would be useful. We came up with algorithms [that] extract those gestures from just the video," said researcher Sanshzar Kettebekov, a Pennsylvania State University computer science and engineering graduate student.

The researchers used this database to create a pair of applications designed for large screens that allow the computer to interpret what people mean when they use a mix of speech and pointing gestures.

One application, dubbed IMAP, is a campus map that responds to pointing and spoken queries. "It brings the computer into the loop with the human," said Kettebekov. For example, if a person asks the map for a good restaurant in an area she is circling with her hand, the computer will reply based on the spoken request for a restaurant and the gestural request for a location, according to Kettebekov.

The second application is a battlefield planning or city crisis management simulation that allows a person standing in front of a large screen to direct vehicles around a battlefield or city. "A person has limited resources [and there are] alarms going off all over the city. The person is using... a 50-inch display... to direct the resources to where the alarm is going [off]," said Kettebekov.

Even though it seems easy to us, giving a computer the ability to sense and make sense of gestures in a verbal context is a complicated problem that involves several steps, according to Kettebekov. The computer must be able to track the user's hands, recognize meaningful gestures, and interpret those gestures.

The first problem is tracking. "We have a vision algorithm that tracks a person and tries to follow a person's hand," Kettebekov said. The second stage is picking out the pointing gestures. "You're trying to delimit gestures from a continuous stream of frames where the hands are just moving — saying 'from here to here was this gesture'," he said. "The third stage is interpretation when you really associate [the gesture you have isolated] with parts of speech and try to extract meaning," he said.

Multimodal human computer interaction is an active research topic with a long history, said Jie Yang, a research scientist at Carnegie Mellon University.

"Coordination of speech and gestures is an old but still open problem," he said, noting that there was a paper published 20 years ago on a computer system that integrated speech and gesture, and there have been many studies on the advantages of using speech and gesture. "Yet, we cannot naturally interact with a computer using speech and gesture without constraints today."

When all the difficult computer problems have been worked out, however, systems that recognize speech and gesture will allow a person to "efficiently manipulate multimedia information regardless of whether the person is communicating with a computer or with another human," he said.

The Penn State researchers are working on improving their gesture recognition algorithms by adding an understanding of the prosodic information that lends speech its subtle shades of meaning, said Kettebekov.

"We're working on using prosodic information in speech: tone of voice, stresses, pauses... to improve gesture recognition and interpretation," he said.

The toughest of the three gesture problems is improving gesture recognition, said Kettebekov. Currently the system identifies keywords and tries to correlate them with gestures. Adding prosodic information would help the system to both recognize gestures and interpret them, he said.

For example, when a TV meteorologist wants to emphasize a keyword, he raises the tone of his voice, said Kettebekov. "If I want you to pay attention I not only point, but my voice would change so that I would attract more attention to that concrete point," he said. "You can extract those most prominent parts of speech, and those parts of speech nicely relate with the gestures — in this case it was pointing," he said.

The researchers may eventually turn their sights to iconic, metaphoric and beat gestural information, but there is a lot of work to be done in the deictic area first, said Kettebekov.

In addition, understanding what these subtler gestures mean from a linguistics point of view "is not there yet — so there's not enough theoretical basis," to use to give that understanding to computers, he said.

Kettebekov's research colleague was Rajeev Sharma of Pennsylvania State University. They presented the research at the Engineering for Human-Computer Interaction conference in Toronto in May, 2001. The research was funded by the Army Research Laboratory and the National Science Foundation (NSF).

Timeline: Now
Funding: Government
TRN Categories: Human-Computer Interaction; Computer Vision and Image Processing
Story Type: News
Related Elements: Technical paper, "Toward Natural Gesture/Speech Control of a Large Display," presented at the Engineering for Human-Computer Interaction conference in Toronto, May 11-14, 2001

TRN

Neural Net Tracks Skin Color

Technology Research News, September 10/17, 2003

Our seemingly easy ability to spot and distinguish one object from another is actually a complicated process that evolved over millions of years.

Researchers working to give computers and robots the ability to recognize gestures are up against several challenges. This system must recognize a face or a gesturing hand, and it must be able to continue to distinguish faces and hands as they move around among other objects.

Researchers from the Chinese Academy of Sciences and the Chinese Institute of Modern Optics have come up with a way to use skin color to detect faces and hands. This is trickier than it sounds, because colors are elusive, changing depending on illumination.

The researchers' system uses a camera connected to a processor that uses an artificial neural network to detect skin color, then processes the information further to determine which skin-color objects should be connected together.

The system could be used to enable gaze and gesture control of electronics like appliances and robots. Tests of the system show that it was able to segment gestures with 96.25 percent accuracy, according to the researchers.

The researchers' gesture recognition method could be used practically in five years, according to the researchers. The work appeared in the August, 2003 issue of *Optical Engineering*.

TRN

Language processing

Hearing Between the Lines

By Kimberly Patch, Technology Research News
July 19, 2000

When humans talk, we exchange a lot of audio information along with the words. Computers, however, don't hear between the lines, which is one reason speech recognition applications can seem so frustratingly stupid. Essentially, today's computers are socially inept, blind to the meanings of subtle pauses or even drastic changes in tone.

The technical reason for this is the Hidden Markov Model (HMM) most speech recognition programs rely on only looks at tiny, 10 millisecond slices of speech. The model works well for picking out words, but misses contextual cues that span words, phrases or sentences.

"When you pause at the end of the sentence or you lengthen or you drop your pitch, that [spans] a region that's at least 10 times larger than the HMM can capture and sometimes 100 times larger," said Elizabeth Shriberg, a senior research psycholinguist at SRI International.

Hoping to remedy the situation, Shriberg and other researchers have shown in a pair of experiments that computers can use speech attributes like prosody — information gleaned from the timing and melody of speech — to better understand human speech.

In one experiment, prosody significantly improved a computer's accuracy in adding punctuation and paragraphs to databases of speech from news broadcasts and phone conversations. Prosody proved even more helpful in sorting the broadcast feed into topics. (See chart)

Prosody includes the duration, pitch and energy of speech. Duration, or the way people stretch or speed certain parts of speech, is most important, said Andreas Stolcke, a senior research engineer at SRI International. "People use the duration of speech sounds in certain ways to emphasize things," he said.

The researchers found that pauses and pitch were most useful in segmenting news speech, while pauses, duration of syllables and word based cues proved significant in the more difficult task of segmenting natural conversation.

Prosodic information is slowly being recognized as an important source of information in speech understanding, said Julia Hirschberg, Technology Leader in the Human-Computer Interface Research Department at AT&T Labs. "The SRI work applies prosodic information to a very important task, topic segmentation, with considerable success. [It's] the first that I know of which improves topic segmentation performance," she said.

In another experiment, researchers used word choice and order as well as prosodic cues to improve the task of automatically categorizing telephone conversations into 42 types of phrases like statements, opinions, agreement, hedging, repeated phrases, apologies, and phrases that signal non-understanding.

Prosody's ability to mark emotional levels of speech may eventually help in certain types of searches, like news footage of politicians having an argument. A similar, real-time, application could be call center operators wanting to know "who the angry customers are right away because you don't want them to have to [continue listening] to a computer," said Shriberg. Prosody also allows computers to gauge attention levels, which may allow educational applications to automatically adjust the difficulty of a task. Prosodic information, because it differs among languages, may also prove useful in discerning what language is being spoken.

The researchers are also looking at using prosody to make speech recognition more accurate — "the holy grail right now," said Stolcke. "The general idea is simply to have a more comprehensive model of everything that can vary within speech. [You] can get significantly better speech recognition if you know the type of utterance," he said.

Better recognition based on prosody is also likely to create a feedback loop that will make talking to computers more natural, said Shriberg. "If the machine is using [pitch and emphasis, people] will put that in their speech because it's getting a response from the machine. They'll adapt to what the machine is able to do — that's a well-known principle."

Real world applications of prosody are at least two years away, said Shriberg.

Shriberg and Stolcke were joined in the prosody topic segmentation research by Dilek Hakkani-Tür and Gökhan Tür of Bilkent University in Ankara, Turkey. They were joined in the automatic tagging of conversational speech research by Noah Coccaro and Dan Jurafsky of the University of Colorado Boulder, Rebecca Bates of the University of Washington, Paul Taylor of the University of Edinburgh, Carol Van Ess-Dykema of the U.S. Department of Defense, Klaus Ries of Carnegie-Mellon University and the University of Karlsruhe in Germany, Rachel Martin of Johns Hopkins University and Marie Meteer of BBN Technologies.

The researchers' work on prosody for topic segmentation was funded by the National Science Foundation (NSF) and the Defense Advanced Research Projects Agency (DARPA). The work on automatic tagging of conversational speech was funded by the Department of Defense (DOD).

Timeline: > 2 years; > 5 years
Funding: Government
TRN Categories: Databases and Information Retrieval; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper "Prosody-Based Automatic Segmentation of Speech into Sentences and Topics" posted in the Computing Research Repository; Technical paper "Dialog Act Modeling for Automatic Tagging and Recognition of Conversational Speech," posted in the Computing Research Repository

TRN

PDA Translates Speech

By Kimberly Patch, Technology Research News
December 17/24, 2003

As speech recognition technology gets better, and as handheld computers get more powerful, audio translators are becoming a more practical proposition.

Researchers from Carnegie Mellon University, Cepstral, LLC, Multimodal Technologies Inc. and Mobile Technologies Inc. have put together a two-way speech-to-speech system that translates medical information from Arabic to English and English to Arabic and runs on an iPaq handheld computer.

The prototype falls short of Star Trek's fictional universal translator in several ways. The system is not transparent — it must be switched between Arabic-to-English and English-to-Arabic modes. It also works only when the speakers are talking about medical information, and it's only about 80 percent accurate in the lab.

The device shows that it's becoming possible, however, to provide automatic translation using a portable device. "It's good enough to make yourself understood," said Alex Waibel, a professor of computer science at Carnegie Mellon University and a founder of Mobile Technologies Inc.

The effort is one of a series of projects aimed at providing the armed forces with automatic translation for medical and force protection situations and making automatic translation in a wider set of subject areas available for tourists during the 2008 Olympics in Beijing, said Waibel.

The Speechalator prototype uses a built-in microphone and a language-selection button. "You push on the button on the iPaq and speak a sentence and then the translation comes out... in the other language," said Waibel. "You can switch it into the opposite mode when the other person answers and it translates back into your own language."

The software consists of three components: a speech recognizer, a translator, and a speech synthesis engine. "Each one of these components have slight twists to them... in order to work properly for speech translation," said Waibel.

The researchers modified the speech recognition engine to optimize it for handling spontaneous speech.

The translation system has the biggest twist. It extracts the key meaning from the input sentence and translates it to an interlingual, or intermediate representation, and the process depends on the speech being contained in a certain domain, or context, like medical information. "It's just certain nuggets in the phrase that... you need to extract," said Waibel.

The process is akin to constructing a medical-context template that fits the key information, then filling in the template, said Waibel. This process makes it possible for the system to handle spontaneous speech. "We go fishing for the nuggets," he said. But it is also a limitation — the system must know what domain a speaker is talking about.

The researchers are working on a system that can handle multiple contexts and automatically switch between them, said Waibel. "It can, for example, recognize 'now you're in the hotel reservation domain', or 'now you're in the conference registration mode', or 'now you're talking about medical problem'," he said.

To come up with templates that handle different domains, the researchers collect a lot of data from people talking in those domains, said Waibel. "The more data we collect the better coverage of all the possible ways you could be saying [these things] becomes," he said.

The difficult part was fitting the software required to do two-way translation in the 64 megabytes of memory contained in the handheld computer, said Waibel. "You need two recognizers, two synthesizers and two translators to make [it] happen in both directions," he said.

The prototype also has a camera attachment that translates text like that on street signs, said Waibel. Snap a picture of a sign with the camera and it automatically extracts the text region, puts the text through a character recognition program, then translates it, he said. "What you then see on the screen is the picture of the scene with a sign and then underneath an English subtitle," he said.

The Speechalator is a practical proof of concept, said Bernard Suhm, a senior scientist at BBN Technologies. "They have engineered the recognizers and other algorithms sufficiently to make them work in real-time on the very limited computational resources of a consumer PDA," he said.

The device carries the promise of being useful not only for medical translation, but also situations such as travel or business, said Suhm. "This work could facilitate the transition of speech-to-speech translation research from the technology side of research, which focuses on algorithms and engineering, to the human factors side of research, which focuses on how people interact with devices, and how useful devices are to tasks from real-life," he said.

The device hasn't yet been run through its paces in a field test, however, Suhm said. "Until then we don't know whether the additional challenges in the field, [like] high levels of noise... or usability issues make it unusable," he said.

The researchers' next steps are to increase the accuracy of the device so that it can deal with ambient noise, and expand the coverage by collecting more data about how people communicate in different domains, said Waibel. The researchers are also working on building learning algorithms that automatically sort out different ways to say the same things.

The researchers' next prototype is scheduled to be finished in the summer of 2004, and will initially have two domains: hotel reservations and medical situations. "Then it [it will] gradually expand towards other domains as are necessary for tourists," he said.

The device can eventually be used to provide translation services for soldiers and relief workers in foreign countries and for travelers, said Waibel.

It could also address a medical problem in the U.S., he said. "There are a number of people in the U.S. who don't speak English and then when going to doctors... feel embarrassed to explain their health problems in front of somebody else who translates," he said.

The researchers are also working on a multilingual speech recognizer that can recognize speech in any of a set of languages, said Waibel. "In that case you might not have to switch the system between the two languages — you just talk in any language and it will come out in any other language you choose," he said.

And they are aiming to develop a system that combines speech translation with human-to-machine translation, said Waibel. "There are certain situations as a traveler... where you want to communicate with a person in another language, but then there are certain other things which you could just as well do communicating with [a computer]," he said. You would want to talk to another person when ordering food, but communicate with a machine to get directions to a railway station, for example.

Longer-term the researchers are looking for ways to deal with spontaneous speech that is not limited to a certain domain, said Waibel.

Waibel's research colleagues were Ahmed Badran, Robert Frederking, Donna Gates, Alon Lavie, Lori Levin, Tanja Schultz and Dorcas Wallace from Carnegie Mellon University, Alan W. Black from Carnegie Mellon University and Cepstral, LLC, Kevin Lenzo from Cepstral, Monika Woszczyna from Multimodal Technologies Inc., and Jürgen Reichart and Jing Zhang from Mobile Technologies Inc. The researchers presented the results at Eurospeech 2003 in Geneva, Switzerland, September 1 to 4. The research was funded by the Defense Advanced Research Projects Agency (DARPA).

Timeline: Now, 4 years
Funding: Government
TRN Categories: Applied Technology; Human-Computer Interaction
Story Type: News
Related Elements: Technical paper, "Speechalator: Two-Way Speech-To-Speech Translation on a Consumer PDA," Eurospeech 2003, Geneva, September 1-4 posted at cmu.edu/~awb/papers/...speechalator.pdf

TRN

Software Paraphrases Sentences

By Kimberly Patch, Technology Research News
December 3/10, 2003

We paraphrase all the time, often without thinking about it. Try to give a computer the means to reword a sentence, however, and it becomes apparent that figuring out how to say it differently is complicated.

Researchers at Cornell University have tapped a pair of unlike sources — on-line journalism and computational biology — to make it possible to automatically paraphrase whole sentences. The researchers used gene comparison techniques to identify word patterns from different news sources that described the same event.

The method could eventually allow computers to more easily process natural language, produce paraphrases that could be used in machine translation, and help people who have trouble reading certain types of sentences.

Two ideas led to the system, said Regina Barzilay, one of the Cornell researchers who is now an assistant professor of computer science at the Massachusetts Institute of Technology.

First, there is a lot of duplication online, which is potentially useful fodder for systems trying to learn to paraphrase. When two reporters describe the same news event, for instance, they may use different details, but they tend to report about the same basic facts, said Barzilay. "This redundancy can help us to learn ways to paraphrase the same information," she said. "If we have a lot of [different sources] we can clean up the noise and identify the pieces of information where they say the same thing."

Even given similar writing styles, sentence-level paraphrasing is more than simply recognizing synonyms. The researchers' example of a pair of business journalism paraphrases makes this clear:

- After the latest Fed rate cut, stocks rose across the board.
- Winners strongly outpaced losers after Greenspan cut interest rates again.

Second, to sort out sentence similarities, the researchers borrowed techniques from computational biology that determine how closely related organisms are by finding similarities among genes. "In computational biology... you have genes which started from the same kind of seed, and then they change during evolution [but] there is some similarity," said Barzilay.

Key to the technique is comparing news sources that cover the same events but employ slightly different styles. Because they are writing about the same events they contain the same facts, or arguments, said Barzilay. "This gives us patterns which are kind of the same — and this is the core of the paraphrasing technique."

The researchers tested the system by comparing articles produced in English between September 2000 and August 2002 by Agence France-Presse (AFP) and Reuters news agencies.

The researchers' system performs two types of grouping: the first comparison is across articles of the same source, said Barzilay.

The researchers' system uses word-based clustering methods to identify sets of text that have a high degree of overlapping words, said Barzilay. Using this method, the researchers identified articles that described individual acts of violence in Israel and army raids on the Palestinian Territories.

They then employed computational biology techniques to identify sentence templates, or lattices. Lattices are made up of words or parallel sets of words that occur across several examples, and arguments, or slots, where names, dates or number of people hurt or killed occur.

The challenge is to identify which sentence differences are due to lexical variability and which are due to different subjects, said Barzilay.

The technique allowed the researchers to identify common templates that journalists use to describe similar events, said Barzilay. Journalists "use a similar style, but then change it — add one word, remove words. With this technique we can still identify this common pattern," she said.

One pattern, or lattice, read: Palestinian suicide bomber blew himself up in NAME on DATE killing NUMBER (other) people and injuring/wounding NUMBER. In addition to the injuring/wounding variable, there are several variables within the name argument: settlement of, coastal resort of, center of, southern city, or garden cafe.

The system found 43 AFP lattices and 32 Reuters lattices. Once these were identified, the researchers did a cross-comparison.

The researchers compared the lattices from the two sources by comparing the slot values of articles written on the same days. They used a statistical technique to identify patterns that tend to take the same arguments in both sources, said Barzilay.

Twenty-five lattices from each source matched up. Taking into account the variables contained within the lattices, there were 6,534 template pairs.

Given a sentence to paraphrase, the system finds the closest match among one set of lattices, then uses the matching lattice from the second source to fill in the argument values of the original sentence to create paraphrases. The sentence can be paraphrased in perhaps as many as 20 ways using different variables, according to Barzilay.

The researchers' ultimate goal is to use the system to allow computers to be able to paraphrase like humans, and to understand paraphrases, "but that's very far [off]", said Barzilay.

Their next step is to find ways to put paraphrased sentences together in order to paraphrase whole documents. Barzilay's previous work, which used a different technique to paraphrase at the level of words and phrases rather than sentences, is part of the Columbia News Blaster project, which summarizes news stories.

The sentence-based paraphrasing system could improve machine translation, according to Barzilay. "Today a majority of machine translation is statistical... you have large amounts of data in English and in French and then the system learns how to translate that," she said.

These systems work best when they have many different translations of a given sentence, however. "To create such a corpus where 10 people are to translate huge amounts of French text is very expensive," said Barzilay. The researchers' system has the potential to accomplish the same thing by taking one human translation and creating 10 paraphrases of it automatically, she said.

Sentence paraphrasing is also useful for people with certain disabilities, said Barzilay. The system could be used to produce paraphrases based on a specific model, for example, for aphasic readers, who find it difficult to read certain types of phrases, she said.

The system also produced a couple of insights into reporters' writing, said Barzilay. It showed that the writing was very formulaic, and it pointed out bias, she said.

For example, the system learned incorrectly that "Palestinian suicide bomber" and "suicide bomber" were the same, and that "killing 20 people" is the same as "killing 20 Israelis", said Barzilay. These mistakes made by the system are "due to how reporters are reporting," she said. "In some sense... the teacher here is what the reporter writes," she said.

Barzilay's research colleague was Lillian Lee. The researchers presented the work at the Human Language Technology Conference held in Edmonton, Canada, May 27 to June 1, 2003. The research was funded by the Sloan Foundation and the National Science Foundation (NSF).

Timeline: Now
Funding: Government, Private
TRN Categories: Databases and Information Retrieval; Natural Language Processing
Story Type: News
Related Elements: Technical paper, "Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment," posted on the Computer Research Repository (CoRR) at arxiv.org/abs/cs.CL/0304006 and presented at the Human Language Technology Conference, Edmonton, Canada, May 27-June 1, 2003

TRN

Neural networks

Design Enables Large Neural Nets

Technology Research News, October 8/15, 2003

Electricity and light each have strengths and weaknesses as communications media. This means, inevitably, that the two must be used together.

Researchers from Carlos III de Madrid University in Spain and the Massachusetts Institute of Technology have devised a neural network architecture that uses a different mix of optics and electronics than previous schemes in order to accommodate large numbers of neurons. The architecture leverages the computational strength of electronics and the fast communications abilities of light. It could be useful in systems that require optical input and neural net computation — like those used for robotic vision.

Neural networks, like the human brain, have many interconnected elements. Neural networks learn by assigning weights to the connections between neurons and changing the weights based on use. This allows a specific set of inputs to be associated with a pattern of weighted neural connections.

The researchers' system carries out the neural weights assignment electronically, which cuts down on the number of difficult-to-align optical connections compared to previous optoelectronic neural networks, according to the researchers.

The design makes it possible to scale the system to a very high number of elements, or neurons, and the system's optical interconnects allow for fast communications among neurons.

The device can be used in practical applications in two to five years, according to the researchers. The work appeared in the September, 2003 issue of *Optical Engineering*.

TRN

Self-Learning Eases Quantum Computing

By Eric Smalley, Technology Research News
July 10/17, 2002

Ordinary computers are rather simple devices, logically speaking. They cannot learn and they have to plow through large problems one step at a time.

Two largely experimental methods of computing — neural networks and quantum computing — go beyond these limitations by mimicking biological brains, and exploiting the quirks of quantum physics, respectively.

A team of researchers at Wichita State University is aiming to combine the two techniques in order to use neural networks' proven capacity for learning to help realize quantum computing's potential to solve astronomically large problems.

The researchers made a simulation of a quantum neural network and used it to show how the theoretical devices could calculate the quantum mechanical property of entanglement. Calculating entanglement was an unsolved problem in field of quantum computing, said Elizabeth Behrman, an associate professor of physics at Wichita State University.

Entanglement is one of the weirder traits of quantum physics. When a subatomic particle or atom is isolated from the environment and cannot be observed, it enters into the quantum mechanical state of superposition, meaning it is in some mixture of all possible states. For example, a particle can spin in one of two directions, up or down. In superposition, however, the particle spins in some mixture of both directions at the same time.

When two or more particles in superposition come into contact with each other, they can become entangled, meaning one or more of their properties, like spin or polarization, become linked, and move in lockstep. Two entangled photons could, for example, have linked polarizations. When one of the photons is knocked out of superposition to become vertically polarized, the other photon leaves superposition at the same instant and also becomes vertically polarized, regardless of the distance between them.

Entanglement allows quantum logic operations to work on many particles at once. A quantum computer can take advantage of entanglement to check every possible answer to a problem with one series of operations across a group of entangled particles rather than having to check each possible answer one at a time.

“Entanglement is the basis for the power of quantum computing,” said Behrman. It is the reason quantum computers are theoretically able to do computations that cannot be done by even the fastest possible classical computer, she said.

Quantum computers are extremely delicate, however, and only a few simple prototypes have been built. And researchers have only come up with a few algorithms to use them with. Bringing neural networks into the picture could solve both of these problems, according to Behrman.

Artificial neural networks consist of virtual nerve cells, or neurons, linked by virtual synapses. Neurons communicate with each other through the synapses, with the output from one neuron becoming the input to another. Like biological synapses, the virtual synapses grow stronger with use and weaken with disuse. Repeated input to a neural network wears a distinct path through the neurons. In other words, neural networks learn.

“A neural network, like the one between your ears, is different from the computer on your desk in several important ways,” said Behrman.

Because neural networks learn, they can handle incomplete data and scale up automatically to handle larger problems. “All three of these characteristics would be great boons to quantum computers,” said Behrman.

For example, it is difficult to construct the algorithms, or software a quantum computer needs to solve a problem, she said. “A quantum neural computer... essentially constructs its own algorithm,” she said.

The entanglement calculation demonstrates that quantum neural networks should be able to work for any of the many difficult problems that researchers are building quantum computers to solve, said Behrman. “We don’t need to construct an algorithm for each of them if we have a quantum neural network,” she said.

The ability of neural networks to handle incomplete data could be helpful because quantum computers are very sensitive to noise, said Behrman. “We’re working on showing that neural computers can help with this [problem], too, but we haven’t yet demonstrated it,” she said.

Quantum neural networks could also help make larger quantum computers. “The advantages of scale-up are obvious,” said Behrman. “At the moment, we have quantum computers that are only of the size of a few qubits.”

Quantifying entanglement is a very important question in quantum information theory, said Vlatko Vedral, a lecturer of physics at Imperial College and Oxford University in England. “Entanglement is at the root of quantum teleportation, quantum cryptography and some quantum algorithms,” he said.

Having a computer that learns to compute an entanglement algorithm is useful because the algorithm “involves an optimization procedure that is difficult to perform,” he said.

The idea of using a quantum neural network to compute entanglement needs to be more thoroughly explored to determine if it offers any advantages, however, said Vedral. “One thing that is not convincing in the [researchers’] paper is that their method is efficient,” he said.

The Wichita State University researchers are beginning a collaboration with other researchers to attempt to build a quantum neural network, said Behrman. Quantum neural networks could be used in practical applications within 10 years, she said.

Behrman’s research colleagues were Vishwas Chandreshekar, Zhonghua Wang, Chaitra Belar, James Steck and Steven Skinner of Wichita State University. The research was funded by the National Science Foundation (NSF).

Timeline: < 10 years
Funding: Government
TRN Categories: Quantum Computing and Communications; Neural Networks
Story Type: News
Related Elements: Technical paper, “A Quantum Neural Network Computes Entanglement,” posted on the arXiv physics archive at arXiv.org/abs/quant-ph/0202131

Index

Executive Summary 1

What to Look For 1

Main report:
- Understanding the world 1
- Seeing things our way 2
- Where things stand 2
- Approaches to pattern recognition 3
- Keen eye for the numbers 4
- Words 4
- Genes — an ocean of data 5
- Audiovisual 6
- Seeing things 6
- Digital watchdogs 6
- Catch my meaning 7
- More than words 7
- Learning from the brain 8
- Making sense 8

How It Works 2
- Anatomy of a recognizer 2
- To supervise or not to supervise 2
- The technologies 2
- Matching 2
- Grammars 2
- Statistics 2
- Mimicking biology 3

Who to Watch 4
- Data analysis 4
- Biometrics and security 4
- Computer vision and image processing 5
- Interfaces and language processing 5
- Neural networks 6

Recent Key Developments 8

Stories:

Data
- Net Scan Finds Like-Minded Users 9
- Software Sorts Tunes 10
- Word Bursts Reveal Hot Topics 12
- File Compressor ID's Authors 12
- Tools Cut Data down to Size 14
- Software Speeds Gene Comparison 16
- Software Cross-Sorts Gene Data 17

Multimedia
- Software Sorts Video Soundtracks 18
- Speech Recognition to Sort Holocaust Tapes 19
- Image Search Sorts by Content 20

Computer vision
- Vision Chip Shines 22
- Computer Follows Video Action 23
- Computers Sort Gender in a Binary World 24
- Quantum Software Gets the Picture 25

Security
- Text Software Spots Intruders 26
- Physics Methods May Spot Intruders 27
- Degree of Difference Sorts Data 28
- Statistics Sniff out Secrets 28
- Software Spots Forged Signatures 29

Interfaces
- Interface Gets the Point 30

Interface Lets You Point and Speak 32
Neural Net Tracks Skin Color 33
Language processing
Hearing Between the Lines 33
PDA Translates Speech 34
Software Paraphrases Sentences 36
Neural networks
Design Enables Large Neural Nets 37
Self-Learning Eases Quantum Computing 37

TRN's Making The Future Report is published 10 times a year by Technology Research News, LLC. Each 20- to 40-page package assesses the state of research in a field like biochips, data storage or human-computer interaction.

Single reports are $300 to $500. A one-year subscription is $1,600. To buy a report or yearly subscription, go to www.trnmag.com/email.html.

We welcome comments of any type at feedback@trnmag.com. For questions about subscriptions, email mtfsubs@trnmag.com or call (617) 325-4940.

Technology Research News is an independent publisher and news service dedicated to covering technology research developments in university, government and corporate laboratories.

Kimberly Patch
Editor
kpatch@trnmag.com

Eric Smalley
Editor
esmalley@trnmag.com

Ted Smalley Bowen
Contributing Editor
tbowen@trnmag.com

Chhavi Sachdev
Contributing Writer
csachdev@trnmag.com