

WHAT ARE YOU LOOKING AT?

By Peter J Bentley

You're waiting at the station for your train and you glance at the electronic poster next to you. It notices that you're looking at it, and from your gaze it works out what you would most like to see. The display changes to show you new brands of mobile phone, and then changes again to show handheld computers as it notices your gaze flicker. You glance at the clock and so it brings up a list of forthcoming trains, then zooms in on yours and shows you exactly how long it will take to arrive. It's as though it's reading your mind – but really it's reading your eyes.

Intelligent displays such as this may only be a few years away thanks to the fascinating research of computer scientists who specialise in eye tracking and machine learning software. Like computerised mind-readers, eyetracking technology follows the strange patterns of our gazes, and machine learning software is used to learn what it all means. But reading our gazes is not easy. It's so hard to do that in 2005, PASCAL (a European funded network of scientists who specialise on pattern analysis, statistical modelling and machine learning) sponsored a challenge: could a computer learn to tell whether we found something useful, just by watching our eyes? Could a computer look deep into our eyes and guess our thoughts? If it could, what kind of program would it need to run?

Our eyes really do give away many of our thoughts. Only a tiny part in the middle of our vision called the fovea is capable of seeing detailed images. Everything else is just a blur. To give us the illusion that we see everything around us in perfect clarity, our eyes dart about several times a second in saccades, sampling different parts of the scene around us, and our brains glue together the separate parts to make the complete view that we think we see. Even as you read this text right now, your eyes are not sliding smoothly along as a camera might. You are hopping from word to word, often focussing on the middle of a word, maybe focussing twice on a longer or unfamiliar word such as *saccade*, sometimes backtracking to resample previous words, and often skipping the smaller words entirely. If a line was drawn, following the path of your gaze as you read this document, it would resemble a messy child's scribble, not the smooth line from left to right that you might have imagined.



When we look at a more complex document such as a web page or poster, it's even worse. Our eyes are flitting about the screen or paper like demented grasshoppers, and even when we fixate on something for a moment, our eyes may drift slightly, tremor or even continue to dart about in tiny micro-saccades. Not only that, but our irises also change depending on our mental state. Their main function may be to dilate depending on the light, but they also fluctuate if we're thinking hard or having an emotional response such as anger, guilt or desire. If we're thinking particularly hard or remembering



Not a leaking pen, this is the pattern of eye fixations over a period of less than 5 seconds as a person searches up and down a Web page for the right link. Larger blobs mean the eye fixated on one spot for longer. Surprisingly few words are read. The red 'x' marks the link that was chosen. *Image produced using a Tobii X50 eye tracker, operated by Sven Laqua, Research Student, Human Centred Systems Group, UCL. Eyes belong to Peter Bentley.*

something, we may look away in a particular direction, let our eyes unfocus and ignore our vision altogether.

The good news is that all of the tiny, barely perceptible changes in our gazes can be measured. In 2005 a Finnish group of computer scientists did exactly that with a series of tests on 11 people. They used a Tobii eye tracker built into a computer monitor that beams near-infrared light at the pupils of the eyes to create patterns of reflections. These patterns were then used to track exactly where the people were looking on the screen, at 50 times a second. The test subjects were given a task rather like a multiple-choice questionnaire: they were presented with a question and 10 possible answers (5 wrong, 4 relevant, and 1 right) and asked to find the right answer. The patterns of the gaze of each person were then measured, as they read and reread the text on the screen. The data gathered was then used for the challenge: could a computer predict which text a person finds most relevant from only the shifting movement of his or her eyes?

“Gaze patterns contain both direct and subtle cues about users' attention and interests, but being very noisy they require sophisticated modeling and signal processing,” according to Samuel Kaski, one of the organisers of the competition. “We thought it would be good to give the machine learning community the chance to try out their methods in a new field of application.”

Two competitions were set for European scientists: in the first, the data was preprocessed into useful, time-independent categories such as length of saccade, length of fixations, pupil diameter; in the second just the raw time-series of measurements of the eyes were provided – a much harder task.

The entries to the competitions were published in a PASCAL-sponsored workshop on Machine Learning for Implicit Feedback and User Modeling. Some attempted to use software based on finite state machines to learn to predict the child's scribble eye gaze pattern. Others tried to assign probability distributions to the data label sequences in an approach known as conditional random fields. Fascinatingly, although each competition was won by a different group of scientists, the same kind of method came out top for both: machine learning software based on neural networks (see box).

Michael Pfeiffer and his colleagues at the Graz University of Technology, Austria, won the first challenge. They used the clever observation that, in a multiple choice exam, the answer that a person perceives as being correct is likely to be read more times, and is likely to be the last line read before the person gives their final answer. So their method ignored most of the tiny movements of the eye and

Neural Networks

Unexpectedly, both winners of the challenge used neural networks to enable their computers to learn this task. Real biological neurons, such as the ones in your head, send electrical pulses to each other and are linked together in super-complex networks. Computer models that approximate this behaviour give computers the ability to learn just as we do. One of the most common models is known as the multi-layer perceptron (MLP), and it was this model that won the first competition. The MLP is a simple network of very basic “neurons”, one for each input parameter to the problem, one for each output, and one or more “hidden layers” connecting the two input and output layers. Neurons send their signals forward through the network, emitting a value as a weighted function of the values on their inputs.

MLPs are popular because they have a good mathematical foundation and they are flexible models that are easy to use. They typically use a sigmoid or hyperbolic tangent function to transform the inputs into an output, and it has been shown that a linear combination of these nonlinear functions can approximate any continuous function of one or more variables. Essentially this means that even when you have no idea how your output may be related to your input, the MLP can approximate the function that produces the output from the input. This is perfect if you're trying to get a computer to learn something tricky, such as which series of eye movements mean a piece of text is useful, and which series mean that it is not.

The winner of the second competition used a more complex type of neural network model. Instead of using the very abstract model of neurons of MLPs, Finnish researcher Tuomas Lepola used a newer approach known as generic neuron microcircuits. This method uses more biologically realistic neurons that fire pulses at each other, and connects them in recurrent networks (the outputs feed back into the inputs) unlike the feedforward networks of MLPs. It has more connections between neurons that are closer to each other in a three-dimensional space, resulting in the formation of “circuits” that are used like a “fading memory” to represent time-series data as it is input to the network. Then the overall state of the network is read by readout functions, trained to extract the desired pattern of information. The whole idea resembles biological neural networks far more than traditional approaches, and its success at solving the challenge is perhaps fitting: a neural network that resembles our brains was best at understanding the movement of our eyes, caused by the real neural networks in our heads.

concentrated on the large and conscious movements. Although clever – attaining the best accuracy of 72.31%, this idea couldn't work for all applications, and indeed it did not win the second challenge where only the raw eye movement data was given. In this more difficult problem, Tuomas Lepola of the University of Helsinki was most successful, with an accuracy of 64.8% on unseen test data.

The future looks promising for this remarkable area of research. Since the competitions were run, PASCAL funded a “pump-priming” project to investigate the ideas further. In this recent feasibility study, researchers from the Helsinki University of Technology, the University of Southampton and UCL collaborated to try an even harder task: could a computer learn whether you found a whole section of text relevant to a single keyword or search topic? Would the pattern of words scanned by your eyes provide enough clues for the computer to figure out what you are looking for? David R. Hardoon and John Shawe-Taylor were responsible for the creation of the machine learning software that had to perform this task. One method they used was support vector machines – a method of statistical machine learning that happens to be a cousin of neural networks. It is a

Support Vector Machines

SVMs are cousins of neural networks, and in fact a certain kind of SVM is exactly the same as a multi layer perceptron. But SVMs originate from the world of mathematics rather than biology. They work by automatically dividing a set of values (vectors) into two classes – effectively figuring out the best straight line that can separate the values from each other. When several values are used in each vector, this line becomes a plane, or more commonly, a hyperplane (a plane in more than 3 dimensions). But this only provides a linear (straight-line) separation between the vectors, so the trick used by SVMs is to use a kernel function, which maps vectors onto a new *twisted* space where they can then be separated by a hyperplane. (Mapping the flat plane backwards to the old space would twist the plane until it was a lumpy and convoluted surface, able to separate the overlapping data points.) An SVM with a sigmoid kernel function is equivalent to a two layer perceptron neural network.

So SVMs work because they are able to use a simple hyperplane in combination with kernel functions to separate data. And if you can separate data into two classes then you can use the computer to learn. For example, one set of eye movements corresponds to you finding a passage of text relevant; another set of eye movements corresponds to you finding the text irrelevant. SVMs can distinguish between the two sets, and once it has learnt how, it can predict if future eye movements will correspond to you finding text relevant or not.

EyeTools

The idea of eye tracking is now big business. Eyetools is a company specialising in the area, providing their own analysis of eye movements when presented with adverts and web pages of a huge range of commercial and corporate clients. Rather than helping users to find relevant content, Eyetools helps their corporate clients to design eye catching websites by analysing where people look. If nobody ever looks at an advert, headline or contents list, then this is indicative of a serious design flaw. In this way Eyetools is able to help companies produce the most effective visual designs possible. Companies such as Eyetools work offline, analysing data in order to improve a document. Researchers of *PINView* want to use machine learning and analyse our gazes in real time.

powerful technique, but even so, this was a tremendously challenging task.

“When we first started we thought there was no way it could work... but actually for some topics it performed amazingly well,” says Hardoon. The main problems with accuracy were actually caused by the test subjects themselves. The system was learning to understand how we skim-read text in order to pick out just a few key words – this is how we determine if something is useful to us. But this meant that if any test subject decided to read the whole passage of text, the computer couldn't tell if the person found specific words in the text relevant or not. According to Hardoon, because computer scientist “geeks” were used in the tests, “they were too interested in subjects such as astronomy and so read the whole text, spoiling the experiment.” For subjects on sport, the tests were much more successful.

The feasibility study has now been expanded into a full-scale project (*PINView – Personal Information Navigator adapting through Viewing*), under negotiation for funding by the European Union. The ambitious study aims to link several novel forms of input, including speech recognition and the analysis of eye movements, to a search engine. If these scientists are successful, future Internet search engines may involve just speaking a word and then glancing at the results, with every movement of your eyes fine-tuning the search until you find exactly what you want.

For some, the mind-reading technology of eye tracking and machine learning may seem alarming. Mental images of the scene from the movie *Minority Report* may spring to mind, with electronic posters automatically detecting who we are and force-feeding us irresistible advertisements tailored to our every glance and mood. This may be highly desirable for companies wishing to sell their products to us, but

you would always be able to escape. If eye tracking ever did become as ubiquitous and intrusive as television or Internet advertisements, you could always block the systems with a pair of dark glasses.

The scientists involved with this work are aware of these issues. Kai Puolamaki, another of the organisers of the original competition acknowledges, "The privacy issues have to of course be taken seriously. In this sense the eye movements are no different than other personal data stored in hard drives and sent through the net." Luckily, your eye movement data without the machine learning software that interprets them are unlikely to be as easily exploited as information such as your emails or typing on a keyboard, so the movement of your eyes will always be more secure than your fingers.

Consequently, the researchers prefer to take an optimistic view. In the words of Samuel Kaski, "I think it would make a lot of sense to integrate eye tracking technology to computer systems in the future... gaze direction is special because it is tied very closely to our attention and intentions."

The goal of researchers like Kaski, Puolamaki, Hardoon and Shawe-Taylor is to help the public find

what they want with the minimum of difficulty. Ideally this technology will be the perfect way to enable us to navigate through the vast and ever-growing information that surrounds us today. Before long, the right information for you may be just a glance away.

© Dr Peter J. Bentley, 2007
Honorary Research Fellow
Department of Computer Science

Resources

PASCAL: <http://www.pascal-network.org/>
Eye Movements Challenge: <http://www.pascal-network.org/Challenges/IREM/>
Workshop: <http://www.cis.hut.fi/inips2005/>
Pump-priming project:
<http://www.cis.hut.fi/projects/mi/pump06>
Eyetools: <http://www.eyetools.com/>