

IEEE Information Theory Society Honors UC San Diego Engineering Researchers

Researchers at UC San Diego have history of winning the award

October 12, 2006

Maureen C. Curran

Narayana Prasad Santhanam, Junan Zhang and Alon Orlitsky at the UCSD Jacobs School of Engineering Convocation

For the second time in three years, researchers at the University of California, San Diego have been awarded the IEEE Information Theory Society (ITS) Paper Award.

Professor Alon Orlitsky and two of his doctoral students, Narayana Prasad Santhanam and Junan Zhang, authored the paper selected for the 2006 award, "Universal Compression of Memoryless Sources Over Unknown Alphabets." It originally appeared in the July 2004 issue of *IEEE Transactions on Information Theory*. Selected by the ITS Board of Governors, the prestigious annual award recognizes exceptional publications in the field of information theory over the previous two-year period.

Orlitsky is a professor at the UCSD Jacobs School of Engineering with joint appointments in the Electrical and Computer Engineering and the Computer Science and Engineering departments. He is also the director of the Information Theory and Applications (ITA) Center established in 2006 by the California Institute for Telecommunications and Information Technology (Calit2) to study the fundamentals of information theory and their cross-disciplinary applications.

Santhanam and Zhang were graduate students at UCSD when the work was performed. Both graduated recently. Santhanam is currently a postdoctoral fellow at UCSD and will soon assume a similar position at UC Berkeley. Zhang is a postdoctoral research associate at Duke University, North Carolina.

IEEE Information Theory Society

IEEE Information Theory Society Paper Awards

Information Theory and Applications Center

Alon Orlitsky

Center for Wireless Communications

UCSD Jacobs School of Engineering

Computer Science and Engineering Department

Electrical and Computer Engineering Department

The paper considers modeling and compression of sequences over large and even unknown alphabets. Information theory typically considers binary sequences, consisting of 0's and 1's. Over the last four decades, algorithms for processing such sequences have been well studied and understood. However, many practical applications, for example those involving text, audio, or video, typically utilize very large alphabets. It was previously shown that existing techniques would perform poorly when applied to such sequences.

The paper addresses sequences over large alphabets by decomposing them into two parts: a dictionary indicating which values appear in the sequence and a pattern showing how the data changes with time. Since the pattern can be shown to contain essentially all the information about the sequence, the paper concentrates on the pattern, proving that under some general conditions it can be optimally modeled and compressed.

"Good modeling requires understanding the distribution generating the data, hence we first had to derive good estimates for distributions over large alphabets and for low-probability events" said Orlitsky referring to "Always Good Turing: Asymptotically Optimal Probability Estimation" a paper with Santhanam and Zhang that appeared in the October 2003 issue of *Science* magazine. "Probability estimation is an intriguing topic studied by many famous scientists such as Laplace, Ramanujan, Turing, and Fisher," added Santhanam, referring to works on estimating the probability of the sun not rising, the number of butterfly species in Malaysia and breaking the Enigma Code during World War II, "so it was fascinating to see how they addressed the same problem."

The work has a variety of applications and can be used whenever analyzing random phenomena that have many possible outcomes or where some events have very low probability. Examples include intrusion detection, estimating the likelihood and number of gene mutations, etc. The application that motivated the UCSD group was language compression, which plays an important role in many fields, such as speech recognition, data mining and text processing. Natural languages such as English have extremely large vocabularies and many words are used very infrequently, making them excellent candidates for treatment with the techniques addressed in the paper. "I believe that in terms of both the algorithms and applications, we have only scratched the surface," said Santhanam.

About the award itself, Orlitsky said "I am very pleased that the work is appreciated." He added that he is "particularly happy that Junan and Prasad won the award at such an early stage of their careers. It is rare to receive this award for work you have done while a student, but Junan and Prasad are extremely talented and have worked very hard to deserve it." Responded Santhanam: "I love the work we do, the award is icing on the cake."

UCSD researchers are not new to the Information Theory Paper Award. In 2004 the award went to Alexander Vardy, an ECE professor in the Jacobs School (and Ralf Koetter of the University of Illinois at Urbana-Champaign), in 1992 it was given to Paul Siegel (and Razmik Karabed), in 1974 to Jack Wolf (and David Slepian) and in 1968 to Andrew Viterbi. In addition to faculty members, frequent visiting scholars have also won the award. Jacob Ziv for example has won the award twice -- in 1977 with Aaron Wyner and in 1979 with Abe Lempel. Ziv is a professor of Electrical Engineering and a Distinguished Professor at the Technion-Israel Institute of Technology.

Asked about the large number of awards that have gone to UCSD faculty and affiliates, Ramesh Rao, director of Calit2's UCSD Division, and himself a former member of the governing board of the IEEE Information Theory Society said: "UCSD has a world-class group of faculty with extensive experience in information theory."

Media Contact: Maureen C. Curran, (858) 822-4084

