

LAMBERT MATHIAS

100 West University Pkwy, Apt 6-D, Baltimore, MD 21210
+1-484-680-1694 • lambert@jhu.edu • <http://www.clsp.jhu.edu/~lambert>

RESEARCH INTEREST KEYWORDS

Machine Translation
Natural Language Processing
Speech Recognition
Information Extraction
Machine Learning

RESEARCH OVERVIEW

My principal area of interest is statistical modeling for natural language processing applications. One of the key problems in most statistical natural language processing systems is modeling uncertainty. This problem is compounded further in applications which rely on further downstream processing. One of my long-term research goals is to contribute to the development of robust statistical models and learning techniques, capable of handling uncertainty, both in the general statistical modeling field and in specific speech and language processing domains.

Towards this, my research has focused on two core areas - statistical machine translation (SMT) and automatic speech recognition (ASR). In machine translation, I have contributed significantly to the development of a weighted finite state translation system for translation of speech word lattices. I have also worked on discriminative procedures for optimizing various SMT objective functions based on growth transformations. In ASR, my work has primarily focused on lightly supervised approaches for discriminative training, where I helped develop novel frame-filtering based training strategies to discriminatively train the acoustic models using non-literal transcripts.

I am also interested in exploring machine learning based approaches in the natural language processing domain. In this direction, I have conducted some research on applying boosting techniques to minimum bayes risk discriminative training of acoustic model parameters. In the long-term, I am interested in a diverse range of problems such as adapting existing models to new domains, developing complex yet scalable models capable of incorporating diverse features, semi-supervised approaches to discriminative parameter estimation. The convergence of machine learning and natural language processing can lead to some exciting new methodologies in language technologies.

EDUCATION

- 2002-2007 **PhD Candidate, Center for Language and Speech Processing**
Johns Hopkins University, Baltimore, MD, USA
Research Advisor: Dr. William J. Byrne
Dissertation: Statistical Speech Translation and Automatic Speech Recognition under uncertainty
- 2000-2002 **M.S. in Electrical and Computer Engineering**
Michigan State University, Baltimore, MD, USA
Advisor: Dr. Jack Deller, Jr.
Thesis: Feasibility Study of Voice Access to Computers for People with Limited Speech
- 1996-2000 **Bachelor of Engineering in Electronics**
University of Mumbai, India

RESEARCH EXPERIENCE

- May 2003-Present **Research Assistant, Center for Language and Speech Processing
Johns Hopkins University, Baltimore, MD, USA**
- Designed and implemented optimization algorithms for Statistical Machine Translation
 - Made extensions to the Translation Template Model, a weighted finite state based translation system to handle translation of speech lattices.
 - Worked on applying boosting techniques to minimum bayes risk discriminative training in automatic speech recognition.
- June 2004 - August 2004 **Summer Research Intern, Multimodal Technologies, Inc., Pittsburgh, PA, USA**
- Made significant contributions to the medical transcription system
 - Implemented discriminative training using non-literal transcripts for the medical domain
- August 2001-July 2002 **Research Assistant, Michigan State Vincent Voice Digital Library
Michigan State University, East Lansing, MI, USA**
- Worked on the digital libraries initiative (<http://www.ngsw.org>)
 - Improved the user interface for the word-spotting tool used to search oral archives
- August 2000-July 2002 **Research Assistant, Speech Processing Laboratory
Michigan State University, East Lansing, MI, USA**
- Demonstrated feasibility of a phone-based voice recognition system for people with dysarthric speech
 - Worked in collaboration with Invotek, Inc. (<http://www.invotek.org>) on voice enabled assistive device technologies

TEACHING EXPERIENCE

- Spring 2007 **Information Extraction from Text and Speech**
Graduate level course taught by Dr. Frederick Jelinek
Assisted instructor in course assignments and teaching
- Spring 2003 **Digital Systems Fundamentals**
Undergraduate level course taught by Dr. Gerard Meyer
Graded class work and held problem solving sessions for students
- Fall 2002 **Circuits, Johns Hopkins University, Baltimore, MD, USA**
Undergraduate level course taught by Dr. Weinert
Designed and planned recitation sessions, and administered all grades

PUBLICATIONS

• Conferences

1. Lambert Mathias and William Byrne, "Growth Transformations for Statistical Machine Translation", Manuscript (In preparation)
2. Lambert Mathias and William Byrne, "Statistical Phrase-based Speech Translation", IEEE Conference on Acoustics, Speech and Signal Processing, 2006.
3. Lambert Mathias, Girija Yegnanarayanan, and Juergen Fritsch, "Discriminative Training of Acoustic Models Applied to Domains with Unreliable Transcripts", IEEE Conference on Acoustics, Speech and Signal Processing, 2005

• Workshops

1. Lambert Mathias and William Byrne, "JHU Spanish-English Translation System TCSTAR 2007", Presented at TCSTAR evaluation meeting, Aachen, Germany, 2007
2. Participated in the OpenLab workshop for Speech Translation, Trento, Italy, 2006

• Unpublished Work

1. Lambert Mathias, Peng Xu, and Bill Byrne, "Boosting Minimum Bayes Risk Discriminative Training", 2003 (report available at <http://www.clsp.jhu.edu/~lambert/docs/boostingMBR.pdf>)

SKILLS

- Proficiency in software development in Matlab, Java, Python, C, C++, Perl
- Familiarity with UNIX, GNU/Linux, Win32, MacOSX.

HONORS AND AWARDS

- Mathias, L, Yegnanarayanan, G, and Fritsch, J, “Discriminative training of document transcription system ”, filed for US Patent in April 2006 (Publication No. 20060074656)