

VICTOR SOTO MARTINEZ

CEPSR 701 • Columbia University • New York, NY 10027
<http://www.cs.columbia.edu/~vsoto> • vsoto@cs.columbia.edu

Research Interests

Natural Language Processing, Machine Learning, Speech Processing

Education

Columbia University in the City of New York

Ph.D. Computer Science Expected 2018
M.S. Computer Science February 2013

Universidad Autonoma de Madrid

M.S. Computer and Telecommunication Engineering June 2011
B.S. Computer Science & B.S. Mathematics June 2009

Research Experience

Columbia University, Speech Lab 2011-2018

Ph.D. advisor: Julia Hirschberg

- Web scraping for code-switched resources.
- Neural part-of-speech tagging and language modeling for code-switched sentences.
- Crowdsourcing for language ID and part-of-speech tag annotations.
- Rescoring of word confusion networks for improved keyword search.
- Cross-language modeling of intonational phrase boundaries.

Google Inc., Speech Group

- Mixed-language sentence discovery on the web.
Mentor: Françoise Beaufays Summer 2016
- Selection and combination of hypotheses for dialectal speech recognition.
Mentor: Pedro Moreno Summer 2015
- Maximum entropy features extracted from DNNs for large scale acoustic modeling.
Mentor: Pedro Moreno Summer 2014

Telefonica Research, User Modeling and Data Mining Group 2009 - 2011

Supervisor: Enrique Frias-Martinez

- Land use classification using call detail records.
- Prediction of cell-phone users' socioeconomic level.

Universidad Autonoma de Madrid, Computer Science Department 2009 - 2011

Research advisors: Gonzalo Martinez-Muñoz and Alberto Suarez.

- Double pruning strategies for ensemble classifiers.
- Incorporating prior classification knowledge to dynamic pruning strategies.

Professional Activities

Third Workshop on Computational Approaches to Linguistic Code-Switching (CALCS).

Collocated with ACL, Melbourne, Australia, July 2018.

Co-Organizer. G. Aguilar, F. AlGhamdi, **V. Soto**, T. Solorio, M. Diab and J. Hirschberg.

Teaching

Seminar on Spoken Dialog Systems (COMS E6998), 16 students. Spring 2015
Teaching Assistant. Instructor: Svetlana Stoyanchev.

Natural Language Processing (COMS W4705), 92 students. Fall 2013
Teaching Assistant. Instructor: Michael Collins.

Awards and Honors

Interspeech Student Travel Award, ISCA.	2014
'La Caixa' Foundation Fellowship for Graduate Studies in the USA.	2010
Graduate Student Research Fellowship, Universidad Autonoma de Madrid	2009
Award for Outstanding Undergraduate Academic Performance, Comunidad de Madrid	2008

Publications

Refereed conference papers

- V. Soto**, N. Cestero, J. Hirschberg. *The Role of Cognate Words, POS Tags, and Entrainment in Code-Switching*. In Interspeech, Hyderabad, India, September 2018.
- V. Soto**, J. Hirschberg. *Joint Part-of-Speech and Language ID Tagging for Code-Switched Data*. In Third Workshop on Computational Approaches to Linguistic Code-Switching, at ACL 2018, Melbourne, Australia, July 2018.
- G. Mendels*, **V. Soto***, A. Jaech, J. Hirschberg. *Collecting Code-Switched Data from Social Media*. In LREC, Miyazaki, Japan, May 2018.
- V. Soto**, J. Hirschberg. *Crowdsourcing Universal Part-Of-Speech Tags for Code-Switching*. In Interspeech, Stockholm, Sweden, August 2017.
- V. Soto**, G. Martinez-Muñoz, A. Suarez. *An Urn Model for Majority Voting in Classification Ensembles*. In NIPS, Barcelona, Spain, December 2016.
- F. AlGhamdi, G. Molina, M. Diab, T. Solorio, A. Hawwari, **V. Soto** and J. Hirschberg. *Part of Speech Tagging for Code Switched Data*. In EMNLP, Austin, TX, October 2016.
- V. Soto**, O. Siohan, M. Elfeky and P. Moreno. *Selection and Combination of Hypotheses for Dialectal Speech Recognition*. In ICASSP, Shanghai, China, March 2016.
- M. Elfeky, P. Moreno and **V. Soto**. *Multi-Dialectal Languages Effect on Speech Recognition*. In ICNSLP, Algiers, Algeria, October 2015.
- G. Mendels, E. Cooper, **V. Soto**, J. Hirschberg, M. Gales, K. Knill, A. Ragni and H. Wang. *Improving Speech Recognition and Keyword Search for Low Resource Languages Using Web Data*. In InterSpeech, Dresden, Germany, September 2015.
- M. Ma, J. Richards, **V. Soto**, J. Hirschberg and A. Rosenberg. *Strategies for Rescoring Keyword Search Results Using Word-Burst and Acoustic Features*. In InterSpeech, Singapore, September 2014.
- V. Soto**, L. Mangu, A. Rosenberg and J. Hirschberg. *A Comparison of Multiple Methods for Rescoring Keyword Search Lists for Low Resource Languages*. In InterSpeech, Singapore, September 2014.
- V. Soto**, E. Cooper, L. Mangu, A. Rosenberg and J. Hirschberg. *Rescoring Confusion Networks for Keyword Search*. In ICASSP, Florence, Italy, May 2014.
- V. Frias-Martinez, **V. Soto**, J. Virseda and E. Frias-Martinez. *Can Cell-Phone Traces Measure Social Development?* In NetMob, Cambridge, MA, May 2013.
- V. Soto**, E. Cooper, A. Rosenberg and J. Hirschberg. *Cross-language Phrase Boundary Detection*. In ICASSP, Vancouver, BC, Canada, May 2013.
- V. Frias-Martinez, **V. Soto**, H. Hohwald and E. Frias-Martinez. *Characterizing Urban Landscapes using Geolocated Tweets*. In SocialCom, Amsterdam, The Netherlands, September 2012.
- V. Soto**, V. Frias-Martinez, J. Virseda and E. Frias-Martinez. *Prediction of Socioeconomic Levels using Cell Phone Records*. In UMAP, Girona, Spain, July 2011.

*Both authors are first authors.

Journal papers

- V. Soto**, S. Garcia-Moratilla, G. Martinez-Muñoz, D. Hernandez-Lobato and A. Suarez. *A Double Pruning Scheme for Boosting Ensembles*. In IEEE Transactions on Cybernetics, Volume 44, Issue 12, pp. 2682-2695, December 2014.
- V. Frias-Martinez, **V. Soto**, A. Sanchez and E. Frias-Martinez. *Consensus Clustering for Urban Land Use Analysis using Cell-Phone Network Data*. In International Journal of Ad-Hoc and Ubiquitous Computing, Volume 17, Issue 1, pp. 39-58, January 2014.

Workshop papers

- M. Ulinski, **V. Soto** and J. Hirschberg. *Finding Emotion in Image Descriptions*. In WISDOM at KDD, Beijing, China, August 2012.
- V. Frias-Martinez, **V. Soto**, J. Virseda and E. Frias-Martinez. *Computing Cost-Effective Census Maps From Cell Phone Traces*. In PURBA, Newcastle, UK, June 2012.
- V. Soto** and E. Frias-Martinez. *Automated Land Use Identification using Cell-Phone Records*. In HotPlanet, Washington, DC, June 2011.
- V. Soto** and E. Frias-Martinez. *Robust Land Use Characterization of Urban Landscapes using Cell Phone Data*. In PURBA, San Francisco, CA, June 2011.
- V. Soto**, G. Martinez-Muñoz, D. Hernandez-Lobato and A. Suarez. *A Double Pruning Algorithm for Classification Ensembles*. In Multiple Classifiers Systems, Cairo, Egypt, April 2010.