Emma Strubell

http://cs.umass.edu/~strubell/
strubell@cs.umass.edu | 207-712-4019

SUMMARY

Natural language processing researcher and software engineer with over 5 years of research and programming experience.

Passionate about developing new machine learning techniques to facilitate fast and accurate natural language processing.

My goal is to leverage large-scale data to solve real-world problems.

EDUCATION

UMASS AMHERST
PhD Computer Science
May 2017 (expected)
MS Computer Science
Dec 2014 • 3.8 GPA

UNIVERSITY OF MAINE
BS Computer Science
Minor: Mathematics/Statistics
May 2012 • 3.9 GPA

SKILLS

LANGUAGES
Scala • Java • Python • C

TOOLS
Unix • SQL • LaTeX • git • Grid
Maven • IntelliJ • Eclipse • vim

LINKS

GitHub:// **strubell**
LinkedIn:// **Emma Strubell**
Quora:// **Emma Strubell**

EXPERIENCE

INFORMATION EXTRACTION AND SYNTHESIS LAB Researcher
Sep 2013 – Present | UMass Amherst, Amherst, MA

- Developed a novel learning algorithm facilitating 5-10x test-time speedups in linear models used for part-of-speech tagging, syntactic parsing, named entity recognition.
- Core member of team responsible for building a knowledge base of entities and their relationships from over 2 million raw web and newswire documents for TAC KBP. Ranked 2nd out of 7 in Cold-Start task.
- Frequent contributor to FACTORIE open-source machine learning and natural language processing toolkit (Scala).

IBM Software Engineering Intern
May 2013 – Aug 2013 | Littleton, MA

- Worked with an agile development team to design and implement the backend for the next release of InfoSphere DataStage (Java/REST), a tool for integrating and transforming data from distributed sources at a very large scale.

BINDS LAB Researcher
Sep 2012 – May 2013 | UMass Amherst, Amherst, MA

- Developed an efficient library (C/C++) for implementing recurrent, arbitrary-precision neural networks to verify, predict functionality of novel hardware analog recurrent neural networks (ARNNs).
- Collaborated with a colleague in neuroscience to implement an ARNN model for identifying salient areas of an image inspired by our understanding of the neurological mechanisms for human attention in aural and visual modalities.

SPEED LAB Researcher
May 2011 – Aug 2012 | University of Maine, Orono, ME

- Modeled the spread of biologically-inspired Internet worms using the household epidemiological model
- Implemented parallelized simulations of worm epidemics (C) to run on a supercomputer cluster (Grid) based on data from random sampling of servers in IPv4 address space
- Developed meaningful visualizations of data with many independent parameters (SciPy/NumPy/Matplotlib).

PUBLICATIONS

[1] **Emma Strubell**, Luke Vilnis, and Andrew McCallum. **Training for Fast Sequential Prediction Using Dynamic Feature Selection**. In *NIPS Workshop on Modern Machine Learning and NLP (NIPS WS)*, Montreal, Quebec, Canada, December 2014.

[2] Benjamin Roth, **Emma Strubell**, Katherine Silverstein, and Andrew McCallum. **Minimally Supervised Event Argument Extraction using Universal Schema**. In *4th Workshop on Automated Knowledge Base Construction (AKBC)*, NIPS '14, Montreal, Quebec, Canada, December 2014.

[3] Benjamin Roth, **Emma Strubell**, John Sullivan, Lakshmi Vikraman, Katherine Silverstein, and Andrew McCallum. **Universal Schema for Slot-Filling, Cold-Start KBP and Event Argument Extraction: UMassIESL at TAC KBP 2014**. In *Text Analysis Conference (Knowledge Base Population Track) '14 Workshop (TAC KBP)*, Gaithersburg, Maryland, USA, November 2014.