

Yoshinari Fujinuma

Office Address

Environmental Design 201
315 UCB, 1060 18th Street
Boulder, CO 80309

Contact

fujinumay@gmail.com
<http://akkikiki.github.io/about/>

Education

University of Colorado Boulder, USA August 2016 - Present
Advisor: Michael J. Paul, Jordan Boyd-Graber; 3rd year Computer Science PhD student

University of Tokyo, Japan Sep. 2014
M.S. in Information Science and Technology
Advisor: Akiko Aizawa; Thesis: Detecting Upsurge of Emotion Using Characteristic Expressions in Tweets

International Christian University, Japan March 2012
B.A. in Computer Science and Mathematics
Advisor: Grant Pogosyan; Thesis: Fault-Tolerant Packet Routing Algorithms on Hypercube Networks

Professional Experience

Research Assistant, University of Colorado Boulder, USA Aug. 2016 - Present

- Co-PI: Jordan Boyd-Graber
- Created cross-lingual embeddings using character-level LSTM to exploit orthographically similar words across different language pairs and to transfer resources in high-resource languages into low-resource languages.
- Proposing a new intrinsic evaluation measure for linear mapping-based methods cross-lingual embeddings applied to document classification and bilingual dictionary induction.
- Proposing to include diversity-promoting regularizer to create better monolingual embeddings for cross-lingual embeddings.

Applied Scientist Intern, Amazon.com, USA May 2018 - Aug. 2018

- Built a cross-lingual slot tagger using cross-lingual embedding as the feature for LSTM-CRF with language-adversarial auxiliary function (gradient reversal).
- The trained cross-lingual slot tagger is superior to the baseline trained only on target language only when the training data size is small.

Software Engineer, Amazon/A9.com, Japan Oct. 2014 - Aug. 2016

- Built a JA/ZH language detector for search queries with recall > 0.8 .
- Built a semi-supervised CRF-based Japanese query label tagger with $F1 = 0.91$.

Software Engineer Intern, Amazon/A9.com, Japan Nov. 2013 - Feb. 2014

- The first search engineer to intern at Amazon Japan.
- Detected JA-EN transliteration pairs with $F1 > 0.9$ using EM algorithm.

Part-time Engineer, Atilika, Japan Aug.- Nov. 2013, Apr.- Sep. 2014

- Built a dictionary with 8000+ product names for JA named entity extractor.
- Built a EN to JA transliteration generator. Achieved around 60% in accuracy@1.

Software Engineer Intern, Cookpad, Japan July 2013 (one month)

- Built a recommendation system for related topics on the community site. Accomplished approximately 1.5% of the whole user click rate.

Publications

- Yoshinari Fujinuma, Jordan Boyd-Graber, Michael J. Paul: “A Resource-Free Evaluation Metric for Cross-Lingual Word Embeddings based on Graph Modularity”, Annual Meeting of the Association for Computational Linguistics (ACL), Jul. 2019
- Dasha Pruss, Yoshinari Fujinuma, Ashlynn R. Daughton, Michael J. Paul, Brad Arnot, Danielle Albers Szafr, Jordan Boyd-Graber: “Zika discourse in the Americas: A multilingual topic analysis of Twitter”, PLOS ONE, May 2019
- Yoshinari Fujinuma, Alvin Grissom II: “Substring Frequency Features for Segmentation of Japanese Katakana Words with Unlabeled Corpora”, International Joint Conference on Natural Language Processing (IJCNLP), Nov. 2017 (short paper)
- Yoshinari Fujinuma, Hikaru Yokono, Pascual Martínez-Gómez, Akiko Aizawa: “Distant-supervised Language Model for Detecting Emotional Upsurge on Twitter”, The 29th Pacific Asia Conference on Language, Information and Computation (PACLIC), Nov. 2015
- Yoshinari Fujinuma: “Detecting Japanese-English Transliteration Pairs in Search Query and Click-through Logs”, Amazon Machine Learning Conference, May 2015

Current Projects **Intrinsic Evaluation Measure for Cross-lingual Embeddings** 2017 - present

- Developing a graph-based intrinsic measure to evaluate the quality of cross-lingual embeddings.

Diversity-Promoting Regularizer for Cross-lingual Embeddings 2018-present

- Proposing to use a regularizer inspired from determinantal point process to ameliorate the so-called “strange geometry” of skip-gram models.

Selected Past Projects

Multilingual Topic Model on Zika tweets 2017 - 2018

- Investigating whether Polylingual LDA (PolyLDA) outputs both monolingually and cross-lingually coherent outputs given small number of aligned tweets across multiple languages.
- We translated around 1% of whole tweets and run Poly LDA to capture and summarize tweets in English, Spanish, and Portuguese.

Finite State Transducer (FST) for Kuromoji 2015

- Replaced a double-array trie to an FST to build a dictionary for Kuromoji, a java-based Japanese tokenizer used in Lucene, Solr, and Elastic Search.
- Available at <https://github.com/atilika/fst>

Academic Honors

Dean’s Fellowship (CU Boulder) 2016
Best Bachelor thesis in CS and Math (Gödel Foundation Prize) 2012

Computer and Language Skills

Languages: Proficient: Python, Java; Intermediate: C++,
Pig, Go
Software: PyTorch, Git, Vim, L^AT_EX, MySQL
English: TOEFL iBT 101 (2015)
Domain-specific: machine learning, natural language processing