Yoshinari Fujinuma

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

Contact
fujinumay@gmail.com
http://akkikiki.github.io
Twitter: @akkikiki

Education

University of Colorado Boulder, USA — Aug. 2016 - Present
Advisor: Michael J. Paul, Jordan Boyd-Graber; 4th year Computer Science PhD student

University of Tokyo, Japan — Sep. 2014
M.S. in Information Science and Technology; Advisor: Akiko Aizawa

International Christian University, Japan — Mar. 2012
B.A. in Computer Science and Mathematics; Advisor: Grant Pogosyan

Professional Experience

Teaching Assistant, University of Colorado Boulder, USA — Jan. 2020 - Present
- TA for "CSCI 2270 Data Structures"

Research Assistant, University of Colorado Boulder, USA — Aug. 2016 - Dec. 2019
- Proposed a new intrinsic measure to evaluate cross-lingual word embeddings. We show that this measure is correlated to multiple cross-lingual downstream tasks including document classification and bilingual lexicon induction.
- Investigating on encoding local structural bias using graphs for refining cross-lingual word embeddings.

Applied Scientist Intern, Amazon.com, USA — May 2018 - Aug. 2018
- Built a cross-lingual slot tagger for a dialogue system using cross-lingual embedding as the feature for LSTM-CRF with a language-adversarial objective function.
- 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 the 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 an EN to JA transliteration generator. Achieved accuracy@1 $\approx 60\%$.

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

- Mozhi Zhang*, Yoshinari Fujinuma*, Michael J. Paul, Jordan Boyd-Graber: "Why Overfitting Isn't Always Bad: Retrofitting Cross-Lingual Word Embeddings to Dictionaries", Association for Computational Linguistics (ACL), 2020

- Mozhi Zhang, Yoshinari Fujinuma, Jordan Boyd-Graber: "Exploiting Cross-Lingual Subword Similarities in Low-Resource Document Classification", Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI), 2020
- Yoshinari Fujinuma, Jordan Boyd-Graber, Michael J. Paul: "A Resource-Free Evaluation Metric for Cross-Lingual Word Embeddings based on Graph Modularity", Association for Computational Linguistics (ACL), 2019
- Dasha Pruss, Yoshinari Fujinuma, Ashlynn R. Daughton, Michael J. Paul, Brad Arnot, Danielle Albers Szafir, Jordan Boyd-Graber: "Zika discourse in the Americas: A multilingual topic analysis of Twitter", PLOS ONE, 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), 2017
- 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), 2015

*denotes equal contribution

Current Projects

Refining Cross-Lingual Word Embeddings 2019 - present

- Proposing to use graph convolutional network to refine cross-lingual embeddings.

Selected Past Projects

Intrinsic Evaluation Measure for Cross-Lingual Embeddings 2017 - 2019

- Developed a graph-based intrinsic measure to evaluate the quality of cross-lingual word embeddings.
- PDF link: `https://www.aclweb.org/anthology/P19-1489`

Multilingual Topic Model on Zika tweets 2017 - 2018

- Investigated whether Polylingual LDA outputs both monolingually and cross-lingually coherent topics given small number (1%) of aligned tweets.
- PDF link: `https://doi.org/10.1371/journal.pone.0216922`

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.
- Code available at `https://github.com/atilika/fst`

Academic Services

- Program Committee: Workshop on Noisy User-Generated Text (W-NUT) 2019
- Secondary Reviewer: ACL 2019, EMNLP 2017, WWW 2017

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, LaTeX, MySQL
English:	TOEFL iBT 101 (2015)
Domain-specific:	machine learning, natural language processing