CHRIS KEDZIE

CURRICULUM VITAE

phone +1 (925) 323-1837

email kedzie@cs.columbia.edu

web www.cs.columbia.edu/~kedzie/

LinkedIn · Github · Google Scholar

Research Statement

I am interested in computational models of natural language generation (NLG) and understanding (NLU). My current work focuses on designing language generation algorithms that faithfully express the communicative goals of a dialog agent. My recent contributions in this area include developing NLG models that faithfully follow dialog plans and self-training methods that improve the semantic correctness of generated utterances.

Some of my research interests include:

- **Text Generation:** Deep neural network models of text generation, paraphrase, and summarization.
- **Faithful/Controllable Generation in Dialog:** Conditional generation of natural language from formal meaning representations, arbitrary data, or other representations of dialog agent belief states, with an emphasis on ensuring the correctness of the generated text with respect to model inputs.
- **Inductive Bias in Text Summarization:** Understanding when humans find text or data interesting, salient, or otherwise remarkable, and building models to do the same in the context of text summarization.

I also apply machine learning to natural language data to extract signals from the web for social scientists.

Education

Columbia University

2014-pres. Ph.D., Computer Science
Advisor: Kathleen McKeown
Thesis: *Salience Estimation and Faithful Generation: Modeling Methods for Text Summarization and Generation*

2014-2018 M.Phil., Computer Science

2012-2014 M.S., Computer Science

Loyola Marymount University

2004-2008 B.A., Music & Recording Arts Double Major
Within Music major, double concentration in:
Classical Guitar Performance
Music Theory/Composition

Research Experience

Research Intern, **Facebook Research**, Summer 2017.

I interned with Umut Ozertem in the Applied Machine Learning (AML) group, focusing on click bait and hate speech detection. During my project, I developed an adversarial training method to improve performance of our convolutional neural network text classifiers. I implemented this architecture in PyTorch and Caffe2, and integrated it into the internal FBLearner system for tracking and running experiments. My project is currently in use by several teams in the AML group, and is currently under submission for a patent.

Research Intern, **Microsoft Research**, Summer 2015.

I interned with Fernando Diaz in New York City, continuing our collaboration on streaming news summarization. I developed scalable summarization systems to provide users with brief updates of news events as they were unfolding. We participated in the Temporal Summarization Track of the 2015 Text Retrieval Conference, where we were the top performer, and invited to give a talk.

Conference Proceedings

Chris Kedzie and Kathleen McKeown. Controllable Meaning Representation to Text Generation: Linearization and Data Augmentation Strategies. *Under submission.*

Language Generation

Chris Kedzie and Kathleen McKeown. A Good Sample is Hard to Find: Noise Injection Sampling and Self-Training for Neural Language Generation Models. *INLG 2019.* ***(Best Paper Award).***

Katy Gero, **Chris Kedzie**, Jonathan Reeve, and Lydia Chilton. Low-Level Linguistic Controls for Style Transfer and Content Preservation. *INLG 2019.*

Automatic Summarization

Yanda Chen, **Chris Kedzie**, Suraj Nair, Petra Galuscakova, Rui Zhang, Douglas Oard, and Kathleen McKeown. Cross-language Sentence Selection via Data Augmentation and Rationale Training. *Under submission.*

Chris Kedzie, Kathleen McKeown, and Hal Daumé III. Content Selection in Deep Learning Models of Summarization. *EMNLP 2018.*

Chris Kedzie, Fernando Diaz, and Kathleen McKeown. Real-Time Web Scale Event Summarization Using Sequential Decision Making. *IJCAI 2016.*

Chris Kedzie, Kathleen McKeown, and Fernando Diaz. Predicting Salient Updates for Disaster Summarization. *ACL 2015.*

NLP for Social Science

Philipp Blandfort, Desmond U. Patton, William R. Frey, Svebor Karaman, Surabhi Bhargava, Fei-Tzin Lee, Siddharth Varia, **Chris Kedzie**, Michael B. Gaskell, Rossano Schifanella, Kathleen McKeown, Shih-Fu Chang. Multimodal social media analysis for

gang violence prevention. *ICWSM 2019.*

Serina Chang, Ruiqi Zhong, Ethan Adams, Fei-Tzin Lee, Siddharth Varia, Desmond U. Patton, William R. Frey, **Chris Kedzie**, Kathleen McKeown. Detecting gang-involved escalation on social media using context. *EMNLP 2018.*

Workshops

David Wan, Zhengping Jiang, **Chris Kedzie**, Elsbeth Turcan, Peter Bell, Kathleen McKeown. Subtitles to Segmentation: Improving Low-Resource Speech-to-Text Translation Pipelines. *LREC 2020 Workshop on Cross-Language Search and Summarization of Text and Speech.*

Douglas W. Oard, Marine Carpuat, Petra Galuscakova, Joseph Barrow, Suraj Nair, Xing Niu, Han-Chin Shing, Weijia Xu, Elena Zotkina, Kathleen McKeown, Smaranda Muresan, Efsun Selin Kayi, Ramy Eskander, **Chris Kedzie**, Yan Virin, Dragomir R. Radev, Rui Zhang, Mark J. F. Gales, Anton Ragni, and Kenneth Heafield. Surprise Languages: Rapid-Response Cross-Language IR. *EVIA 2019.*

Chris Kedzie, Kathleen McKeown, and Fernando Diaz. Summarizing disasters over time. *Bloomberg Workshop on Social Good (with SIGKDD) 2014.*

Talks

2019	Natural Language, Dialog and Speech Symposium. The New York Academy of Sciences. **(STAR Talk 2nd Place Prize)**
2018	Agolo.
2016	Machine Learning and Friends Lunch. UMass. Amherst.
2015	Text Retrieval Conference (TREC), Temporal Summarization Track.
2014	Text Retrieval Conference (TREC), Temporal Summarization Track. KDD at Bloomberg: Data Frameworks Track.

Demos

2016	*Monitoring Large Scale Disasters.* Data Science Day @ Columbia University Data Science Institute.

Doctoral Consortium

2016	*Extractive and Abstractive Event Summarization over Streaming Web Text.* IJCAI.

Summer Schools

2016	Deep Learning Summer School. University of Montreal.

Community Service

2016–pres.	Organizer. Columbia NLP Reading Group.
2014–2017	Organizer. NLP talks at Columbia.

Teaching Experience

Fall 2019	Guest Lecture. Deep Learning for NLP, Columbia University.
Fall 2018	Guest Lecture. Deep Learning for NLP, Columbia University.
Spring 2014	Teaching Assistant. Semantic Technologies in IBM Watson. Instructor: Alfio Gliozzo.