

CHRIS KEDZIE

PERSONAL INFORMATION

Made in California, January, 29th 1986

web www.cs.columbia.edu/~kedzie

email kedzie@cs.columbia.edu

github <https://github.com/kedz>

phone +1 (925) 323 1837

ABOUT

I am currently a third year Ph.D. student in the Department of Computer Science at Columbia University, working in the field of Natural Language Processing under Prof. KATHLEEN McKEOWN. My research has focused on predictive models for large scale and streaming news summarization. I am currently working on text generation for multi-document and streaming news summarization. In particular, I am interested in incorporating linguistic constraints, e.g. discourse coherence, with neural network sequence models to generate more natural, informative, and reliable summaries.

PUBLICATIONS

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

*Proceedings of the
25th International
Joint Conference
on Artificial
Intelligence*

Abstract: We present a system based on sequential decision making for the online summarization of massive document streams, such as those found on the web. Given an event of interest (e.g. "boston marathon bombing"), our system is able to filter the stream for relevance and produce a series of short text updates describing the event as it unfolds over time. Unlike previous work, our approach is able to jointly model the relevance, comprehensiveness, novelty, and timeliness required by time-sensitive queries. We demonstrate a 28.3% improvement in summary F_1 and a 43.8% improvement time-sensitive F_1 metrics.

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

*Proceedings of the
53rd Annual
Meeting of the
Association for
Computational
Linguistics*

Abstract: During crises such as natural disasters or other human tragedies, information needs of both civilians and responders often require urgent, specialized treatment. Monitoring and summarizing a text stream during such an event remains a difficult problem. We present a system for update summarization which predicts the salience of sentences with respect to an event and then uses these predictions to directly bias a clustering algorithm for sentence selection, increasing the quality of the updates. We use novel, disaster-specific features for salience prediction, including geo-locations and language models representing the language of disaster. Our evaluation on a standard set of retrospective events using ROUGE shows that salience prediction provides a significant improvement over other approaches.

TALKS

Nov. 2015 Learning 2 Summarize: TREC 2015

*Temporal
Summarization
Track, TREC 2015*

Abstract: In this talk, I present an overview of our participation in the temporal summarization track at the 2015 Text Retrieval Conference. Most of the available training data for this task consists of static judgments on returned

updates, making it difficult to make use of sequential predictions in a learned model. I show how we used learning based search (SEARN, Learning2Search, LOLS) to sample realistic runs over the training streams and learn from dynamic features like previous update decisions and rolling stream observations. Our resulting system is able to build an event summary in an online fashion avoiding latency penalties while still outperforming retrospective approaches (e.g. clustering).
Joint work with FERNANDO DIAZ.

July 2015 Predicting Salient Updates for Disaster Summarization

*Association of
Computational
Linguistics*

I presented our long-paper submission at the 2015 meeting of the Association for Computational Linguistics in Beijing, China. See associated paper abstract.

Nov. 2014 Columbia U. at TREC: Temporal Summarization

*Temporal
Summarization
Track, TREC 2014*

Abstract: In this talk, I present an overview of our participation in the temporal summarization track at the 2014 Text Retrieval Conference. Our submission was one of the top overall submissions for this track. Our performance gain came largely from our precision in the summary update selection stage; I outline the details of our salience regression model and affinity propagation clustering architecture, including their effect on our scores. I also address our current system shortcomings, especially our inability to explicitly control for redundancy.

Joint work with FERNANDO DIAZ & KATHLEEN MCKEOWN.

Aug. 2014 Summarizing Disasters Over Time

*Bloomberg Data
Frameworks Track,
KDD 2014*

Abstract: We have developed a text summarization system that can generate summaries over time from web crawls on disasters. We show that our method of identifying exemplar sentences for a summary using affinity propagation clustering produces better summaries than clustering based on K-medoids as measured using Rouge on a small set of examples. A key component of our approach is the prediction of salient information using event related features based on location, temporal changes in topic, and two different language models.

Joint work with FERNANDO DIAZ & KATHLEEN MCKEOWN.

DOCTORAL CONSORTIUM

July 2016 Extractive and Abstractive Event Summarization
over Streaming Web Text, 25TH INTERNATIONAL JOINT
CONFERENCE ON ARTIFICIAL INTELLIGENCE

DEMOS

April 2016 Monitoring Large Scale Disasters, DATA SCIENCE
DAY

*Columbia
University's Data
Science Institute*

During crises such as natural disasters or other human tragedies, information needs of both civilians and responders often require urgent, specialized treatment. Monitoring and summarizing important information during such an event remains a difficult problem. We present a system for monitoring online news for such disasters. Given a query: e.g. "Hurricane Sandy," our system analyzes the web, and produces a sequence of updates, brief textual descriptions about the current state of the event, as that event unfolds over time. We use novel, disaster-specific features for generating updates, including geo-locations and language models representing the language of disaster. Our demo will allow users to see updates generated for pre-run queries including:

Hurricane Sandy, the Boston Marathon bombing, and 40 other large scale disasters.

WORK EXPERIENCE

MSR-NYC Summer 2015 Research Intern, MICROSOFT RESEARCH
I interned with Fernando Diaz at Microsoft Research 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 unfold. Our work was submitted to the Temporal Summarization Track of the 2015 Text Retrieval Conference, where we were a top performer and invited to give a talk.

Columbia University Spring 2014 Teacher's Assistant, COLUMBIA UNIVERSITY
I was the TA for the class *Semantic Technologies in IBM Watson*, taught by IBM researcher ALFIO GLIOZZO. The class covered the various inner workings of the Jeopardy playing computer. My responsibilities included teaching several lectures on foundational natural language processing tasks and problems, and an overview of the semantic web. Along with Dr. GLIOZZO, I helped guide and supervise the various student projects, one of which led to a publication at EMNLP 2014.

*Stimmung
stimmung.tv* 2008–2011 Composer's Assistant, STIMMUNG — New York
Performed audio engineering/mixing/editing and sheet music preparation for staff composers in a busy commercial music and sound post-production studio. Posted and presented work to clients. Provided general office support and correspondence. Organized and archived audio and video assets. Coordinated asset delivery to clients/post-production services. Worked on many CLEO and Emmy award winning commercial campaigns including several Super Bowl spots for such clients as: Coca-Cola, Mercedes-Benz, Kia, Levi's, and Monster.com. In addition to commercials, I also helped produce music for several independent films, documentaries, and television shows including *Reagan* (HBO), *The Rising: Rebuilding Ground Zero* (Discovery Communications), and *Journey to the Stars* (Hayden Planetarium, American Museum of Natural History).

*KXLU 88.9FM
kxlu.com* 2006–2008 Production Director, KXLU — Los Angeles
Worked with station directors and staff to plan concerts and events in the Los Angeles area, as well as the annual fundraiser. Supervised implementation of a new website. Coordinated the recording and broadcast of all live and pre-recorded performances and interviews at the station. Managed and researched equipment upgrades for the KXLU Production Studio.

EDUCATION

*Doctor of
Philosophy* 2014–Present Columbia University
Natural Language Processing · Dept. of Computer Science ·
Fu Foundation School of Engineering & Applied Science
Adviser: Prof. KATHLEEN MCKEOWN
Description: I am a third year Ph.D. student, working with Prof. MCKEOWN on event understanding from text data. My research has focused on automatic news summarization, using trainable models in a streaming news setting. For the past two years, I have participated in the Temporal Summarization Track at the Text Retrieval Conference (TREC) and have been invited to present both times. I am generally interested in regression, ranking, and optimization for content selection, especially when applied to automatic summary generation. With Prof. MCKEOWN and FERNANDO DIAZ at Microsoft Research, I have applied these techniques to the domain of man-made and natural disaster news.

<i>Master of Science</i>	2013-2014	Columbia University	GPA: 3.87 · <i>Natural Language Processing</i> · Dept. of Computer Science · Fu Foundation School of Engineering & Applied Science Adviser: Prof. KATHLEEN McKEOWN Description: I continued to pursue my interests in natural language processing, in addition to machine learning and statistics. In Prof. McKEOWN's lab I worked on question-answering (QA) for the DARPA BOLT (Broad Operational Language Translation) project. My research focused on unsupervised methods of text similarity and the application of semantic web/linked open data for QA.
	2012-2013	Columbia University	GPA: 3.95 · <i>Post Baccalaureate Studies</i> · School of Continuing Education Description: While taking introductory courses in computer science, I also worked as a research assistant for Prof. KATHLEEN McKEOWN and her student, SARA ROSENTHAL. Responsibilities included annotating research corpora for supervised learning systems, developing web crawlers to extract user discussions from online forums, and building research corpora for studies in automatic influence and agreement detection in natural language.
<i>Bachelor of Arts</i>	2011	Baruch College, CUNY	GPA: 4.0 · Continuing & Professional Studies Description: I took two classes on Java and Oracle SQL development.
	2004-2008	Loyola Marymount University	GPA: 3.34 · <i>Music/Recording Arts Double Major</i> · College of Communication and Fine Arts/School of Film and Television Description: In my undergraduate degree, I pursued interests in both classical music and sound design/mixing for film. Within the music department, I concentrated on music theory/composition as well as guitar performance, culminating in two senior theses, an original composition, <i>String Quartet for Space Travel</i> , and a guitar recital, featuring works by Antonio Lauro, Roland Dyens, Leo Brouwer, Miguel Llobet, Antonio Vivaldi, and others. Within the film department, I scored and sound designed/mixed many student films (<i>The Cannibal Ad</i> , Golden Hamster(Best Overall) Award and 1 st Place for Narrative Short, 2005 Northwest Projections Film Festival; <i>Lily</i> , Best Sound, Best Film, 2007 LMU School of Film and Television "Film Outside the Frame Festival.").

COMPUTER SKILLS

<i>Application Areas</i>	Automatic Summarization, Text Clustering, Text Representation & Feature Learning, Machine Learning, Data Mining, Web Scraping
<i>Languages (Adept)</i>	English, C/C++, PYTHON, JAVA, PERL, HTML, L ^A T _E X
<i>Languages (Familiar)</i>	Latin, MATLAB, X10, JAVASCRIPT, SQL, SPARQL, LINUX, BASH/shell scripting

OTHER INFORMATION

<i>Interests</i>	Experimental Music · Pop Music · Punk Rock
------------------	--

April 10, 2016