Tweet Summarization:
An Annotated Bibliography

Marcus Hughes (jmh3@williams.edu)
Williams College

April 24, 2017

References

[1] F. C. T. Chua and S. Asur, "Automatic summarization of events from social media," in *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, 2013.

> A very sophisticated approach to identify topics using a graph model with a decay rate. Specific to twitter as well. Unfortunately, they do not provide any summary examples!

[2] W. Li, W. Xu, C. Yuan, M. Wu, and Q. Lu, "Extractive summarization using inter- and intra- event relevance," in *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, ACL44*, 2006, p. 369376.

> TO DOWNLOAD AND TO READ

[3] A. Nenkova and L. Venderwende, "The impact of frequency on summarization," in *Microsoft Research, Redmond, Washington Tech. Rep. MSR-TR 2005-101*, 2005.

[4] S. Sripada, V. G. Kasturi, and G. K. Parai, "Multi-document extraction based summarization," in *CS 224N Final Project Report*, 2013.

> They employ three different approaches to create summaries. Their paper is a nice overview and includes many features to consider. They discuss "stack decoder" formulations, clustering forumlations, and graph based formulations.

[5] A. A. Submission, "Argument mining in twitter: Recognizing premise tweets for claim hashtags," *In prep.*

Hashtags can be used to identify premises. They use different features (e.g. discourse indicators, informativeness, sentiment, tone) to identify whether tweets are premise tweets or not. These are then summarized. I could provide a better summarization perchance.

[6] L. Vanderwende, H. Suzuki, C. Brockett, and A. Nenkova, "Beyond sumbasic: Task-focused summarization with sentence simplification and lexical expansion," *Information Processing and Management*, vol. 43, no. 6.

TO READ

[7] W. Xu, R. Grishman, A. Meyers, and A. Ritter, "A preliminary study of tweet summarization using information extraction," in *Proceedings of the Workshop on Language in Social Media (LASM 2013)*, June 2013, pp. 20–29.

Authors claim that their system produces summaries that are more concise and news-worthy than SumBasic as evaluated by human judges. Their graph-based appraoch uses entities, event phrases, and connections between tweets. This builds off a previous work by [2] on news articles. They provide a wonderful table, Table 1, of previous works I can use as reference. They begin by identifying events in tweets. They emphasize it is important to exploit the redundancy in tweets to find meaning. They map named entities and events together in a weighted graph; edge weight is the number of tweets the nodes co-occur in. They then use a modified version of PageRank. They use human evaluation.

[8] E. Yulianti, S. Huspi, and M. Sanderson, "Tweet-biased summarization," *Journal of the Association for Information Science and Technology*, vol. 67, no. 6, pp. 1289–1300, 2015.

Describes how one must modify summarization to account for tweets. What makes a good summary.