Exploring and Ranking Articles by Optimizing Multiple Semantic Features

Jason Cohn
Northwestern University
Intelligent Information Lab
Ford EDC, 2133 Sheridan Rd., Evanston, IL USA
jasoncohn@u.northwestern.edu

Larry Birnbaum
Northwestern University
Intelligent Information Lab
Ford EDC, 2133 Sheridan Rd., Evanston, IL USA
l-birnbaum@northwestern.edu

ABSTRACT

Whether it's a social media system populating a news feed or a user searching for content to share, decisions are constantly being made on the basis of semantic information. Topic, sentiment, and preference are among the many semantic dimensions of content that humans and machines must carefully weigh when prioritizing media consumption. Though easy for humans, such natural language tasks are nontrivial computationally. In this paper, we present a novel technique for sorting a corpus of news articles based on two competing semantic objectives. Solving this multi-objective optimization problem yields a pareto front with a finite set of solution articles. Iterating on the remaining data, we construct solution sets in tiers of successive pareto fronts. Our technique allows for exploration of the tradeoffs between semantic concepts, yielding cogent and trope-like results.

INTRODUCTION

Semantic information processing is integral to a system which aspires towards any degree of natural language understanding. That is, to a system that is able to reason about concepts and meaning in a humanlike manner. A content curation system based entirely on semantic understanding would represent a major leap forward in computation.

Semantic information is widely neglected in content curation systems, as is the majority of the natural language in an article or post. When Facebook's NewsFeed algorithm recommends content, it is relying almost entirely on metrics based on interactions in the social network itself, such as number of likes, the people involved in the post, and those people's closeness to you [2]. This system represents the standard for content curation across social media platforms.

METHOD

To illustrate our sorting technique, we built simple representations of semantic dimensions for sentiment and topic. Our representations draw inspiration from lexical
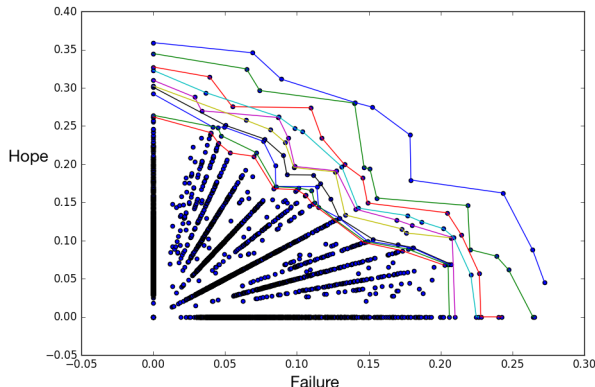

Figure 1. Tiers of pareto fronts rank news articles starting from the top blue line, the first front. 10 iterations shown

techniques for sentiment analysis, especially as construed in the EmoLex [1] system. EmoLex crowdsourced a set of words that participants felt represented any of eight chosen emotions including anger, disgust and joy. The frequency of crowdsourced annotations allowed the system to assign weights to each word as well. [1]

In our research, we similarly represent sentiment and related semantic concepts as lists of words. These include models for $failure$, $hope$, $success$, among others. For the sorting task, a pair of concepts are selected as semantic objective functions. We then solve the resulting multi-objective optimization problem, yielding a pareto frontier. Our algorithm then iteratively builds pareto frontiers from the remaining articles, assigning ranks to articles based on the iteration in which they are included in a pareto frontier.

CONTRIBUTIONS

In this paper, we introduce a novel technique for operating on semantic concepts to perform a task whose output proved to be cogent and meaningful. The sorting algorithm takes as input, representations of semantic concepts, and extrapolates them across a sizable corpus. This new way of approaching content curation shows that we can improve the ability of social network news feeds to analyze natural language in a humanlike fashion. It is imperative that we develop ways to humanize existing algorithms which are all too often enigmatic and non-transparent to users.

REFERENCES

1. Mohammad, S. M., and Turney, P. D. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, Association for Computational Linguistics (2010), 26–34.

2. Rader, E., and Gray, R. Understanding user beliefs about algorithmic curation in the facebook news feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM (2015), 173–182.