BlogVox2: Sentiment Detection in Political Blogs

Sandeep Balijepalli, Justin Martineau & Tim Finin

Work flow of the political opinion system

**Objective**

Present a political summary of blogger opinions.

**1. Contents Aggregator**

New posts, from the RSS feeds obtained from our political urls database, are sent to the sentence chunker.

**2. Filters for sentence screening:**

Un-opinionated sentences are filtered out using filters.

A. **Pattern Matching Filter**

Customized Pattern matching techniques along with parts of speech tagging are applied.

B. **Naïve Bayes**

The Filters are trained from the internet movie review dataset and from the results of the pattern matching filter.

**3. Lucene Indexer**

Sentences passing through the filters are indexed into one of multiple indices based on post date.

**4. Trend Analysis**

Trends are calculated based on index posts.

A. **Top Topics**

- The most frequently discussed.
  - Eg: Bush, Iraq Hillary, Obama.

B. **Hot Topics**

- Terms that experienced a sudden increase in volume.
  - Eg: Virginia tech, Yelsin.

**System Architecture**

- **Database**
  - Get Political Urls from Database
  - Gets Res. from urls
- **Contents / Documents**
  - Check for new content
- **Sentence Chunker**
  - Sentences passing through the filters are indexed into one of multiple indices based on post date.
- **Filters**
  - A. Pattern Matching Filter
  - Customized Pattern matching techniques along with parts of speech tagging are applied.
  
  B. Naïve Bayes
  - The Filters are trained from the internet movie review dataset and from the results of the pattern matching filter.
- **Lucene Index**
  - Index, opinionated sentences
- **Trend Analysis**
  - Sorts results by polarity strength
  - Divide results by Blogger Party

**Future Work**

- **1. User Interaction**
  - User needs Political Information
- **2. Query**
  - User enters query term. Query term is boosted to favor phrases over disjoint query pieces.
- **3. Results**
  - Results displayed in graphs and by sorted posts.

**Future Work**

- Increase pattern matching data set.
- Improve training set for Naïve Bayes filter and tailor it politics.
- Examine other domains.