Scalable Detection of Sentiment-Based Contradictions

Mikalai Tsytsarau
University of Trento
March 28, 2011

Scalable Detection of Sentiment-Based Contradictions

Mikalai Tsytsarau
University of Trento

© WILMAR PHOTOGRAPHY.COM

Themis Palpanas
University of Trento

Kerstin Denecke
L3S Research Center

Scalable Detection of Sentiment-Based Contradictions
Contradictions, what are they?

- **Contradictions** in text are situations where 'two sentences are extremely unlikely to be true together'

- Contradictions may be of different types, for example:
  - antonymy: *hot* - *cold*, *light* - *dark*, *good* - *bad*
  - negation: *nice* - *not nice*, *i love you* - *i love you not*
  - mismatches: *the solar system has 8 planets* - *there are 9 planets*
  - sentiments: *i like this book* - *this reading makes me sick*

- **Sentiment Contradictions** may occur due to:
  - diversity of views
  - change of views
There are many services where users publish their opinions: blogs, wikis, forums, social networks and others.

Sentiment analysis is used to:
- learn customers attitude to a product or its features
- analyze people's reaction to some event

Such problems require scalable sentiment aggregation, which is:
- diversity-preserving
- time-aware
Contradiction Detection Pipeline

1. Text Extraction
2. Topic Identification
3. Opinion Extraction
4. Opinion Aggregation
5. Contradiction Extraction

Web pages → Texts → Topics → Opinions → Statistical Values → Contradictions
Formulating the Problem

- **Sentiment** $S$ is a real number in the range $[-1, 1]$, reflecting opinion polarity.

- **Aggregated Sentiment** $\mu S$ is the mean value over sentiments in the collection.

- **Simultaneous Contradiction.** when two groups of documents express a very different sentiment on the same topic, *in the same time interval*.

- **Change of Sentiment.** when two groups of documents express a very different sentiment on the same topic, *but in consequitive time intervals*.
Contradiction Preserving Sentiment Aggregation

Contradiction Detection

Contradiction Detection

Contradiction Detection
Contradiction Preserving Sentiment Aggregation

- **Raw Sentiments** $S_i$
- **Aggr. Sentiment** $\mu_S = \frac{1}{n} \sum_{i=1}^{n} S_i$
- **Sentiment Variance** $\sigma_S^2 = \frac{1}{n} \sum_{i=1}^{n} (S_i - \mu_S)^2$
- **Contradiction** $C = \frac{\vartheta \cdot \sigma_S^2}{\vartheta + (\mu_S)^2} W$

- We calculate contradiction by combining Aggr. Sentiment and Sentiment Variance.
- If Aggregated Sentiment close to 0, the contradiction is high.
- The larger the variance, the higher the contradiction.
We generated time series of raw sentiments with half of them being Gaussian noise.

We can see an aggregated sentiment, which is first positive and later changes to negative.

On the contradiction series we get peaks of contradiction only at the time points $t_1$ and $t_2$.

They correspond either to simultaneous contradiction ($t_2$) or to change of sentiment ($t_1$) and can be distinguished by change of signs of $\mu_s$ just before and just after peak points.
We need a scalability on the number of topics and time interval length, yet the efficiency of access and update.

We need to analyze time series of contradiction level using different granularities online.

Smaller time windows allow us to detect more simultaneous contradiction.

Larger ones reveal opposite opinions, which are sparse across time.
We have the following types of queries:

- **Adaptive-threshold queries**
  
  output nodes $C > \rho \cdot C_{\text{parent}}$

- **Fixed-threshold queries**

  output nodes $C > \rho$

- **Single-topic retrieval**

  topic is specified

- **All-topic retrieval**

  retrieve all
We applied our algorithms to a range of data sets:
- drug reviews collected from the DrugRatingz.com website
- comments to YouTube videos from L3S
- comments on postings from Slashdot.com

User evaluation was performed in two independent stages, during each step we asked users to find sentiment contradictions:
- in the first stage, by looking at trends of positive and negative sentiments
- in the second stage, by providing a plot of the contradiction level

We measured time and number of clicks, needed to find one contradiction
\[ \Delta T = \frac{T_2}{T_1} \quad \Delta N = \frac{N_2}{N_1} \quad \text{smaller is better} \]

We also asked users to evaluate the effectiveness of each approach
\[ \Delta D = \frac{D_2}{D_1} \quad \text{each value was in the range [1-5], larger is better} \]

Precision was calculated as a fraction of clicks, finding contradiction
\[ \Delta P = \frac{P_2}{P_1} \quad P_1 = \frac{N_1}{N_C} \quad P_2 = \frac{N_2}{N_C} \quad \text{larger is better} \]

Then, we compared their improvements for each dataset
Usefulness Evaluation - Workflow

Stage 1

Stage 2

Sentiments (Stage 1)

Sentiments (Stage 2)

Just say "NO" ladies! My ARNP prescribed it and three months later: headaches 24/7, hair l ...
I have been taking Yaz for 2 1/2 years, it has been very effective in preventing pregnancy ...
The three months I took Yaz were probably the three worst months of my life. I tried Yaz a ...
I am almost 19 years old and Yaz was the first birth control I have ever been on. I am cur ...
Ladies, you MUST take a Vitamin B supplement with Yaz! (I take B- ...
I was on yasmine for one year and things were fine. My doctor switched me to yaz and I dec ...
I am with most of the other woman on how awful Yaz has made me feel. I am also taking Lexa ...
Usefulness Evaluation - Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Topic name</th>
<th>Size</th>
<th>$\Delta D$</th>
<th>$\Delta T$</th>
<th>$\Delta N$</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$\Delta P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug Ratingz</td>
<td>Ambien</td>
<td>60</td>
<td>1.50</td>
<td>0.60</td>
<td>0.88</td>
<td>0.70</td>
<td>0.81</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>Yaz</td>
<td>300</td>
<td>1.58</td>
<td>0.93</td>
<td>0.78</td>
<td>0.75</td>
<td>0.95</td>
<td>1.32</td>
</tr>
<tr>
<td>Slashdot</td>
<td>Int. control</td>
<td>159</td>
<td>1.17</td>
<td>0.89</td>
<td>0.58</td>
<td>0.37</td>
<td>0.63</td>
<td>2.14</td>
</tr>
<tr>
<td>YouTube</td>
<td>Zune HD</td>
<td>472</td>
<td>2.07</td>
<td>0.68</td>
<td>0.62</td>
<td>0.36</td>
<td>0.61</td>
<td>2.09</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td>1.58</td>
<td>0.77</td>
<td>0.72</td>
<td>0.55</td>
<td>0.75</td>
<td>1.69</td>
</tr>
</tbody>
</table>

- Stage 2 (based on contradictions) has showed improvement on all measures
- The largest improvement in precision was achieved for Slashdot and YouTube
- Time improvement was small for a dataset with many contradictions (Yaz), but it was large for dataset with infrequent contradictions (Ambien)
- In all cases, our tool allowed to reduce the search space (number of clicks)
Scalability Evaluation - Experimental Setup

- Syntetic dataset contained randomly generated sentiments for 1,000 topics
- We generated sets of 25 single-topic and all-topics queries
- In queries we used granularities and topics drawn uniformly at random
- We measured the time needed to execute these queries against the database as a function of the time interval, and the granularity of the time windows.
- We report results for both the fixed threshold and the adaptive threshold.

```sql
select c1.topic_id, c1.timeBegin, c1.timeEnd from contradictions c1
join contradictions c2 on c1.topic_id = c2.topic_id and c2.granularity = c1.granularity + 1 and (c1.timeBegin is between c2.timeBegin and c2.timeEnd
where c1.contradiction >= c2.contradiction * @threshold and c1.granularity = @window and (c1.timeBegin is between startDate and endDate or c1.timeEnd is between startDate and endDate);
```
Scalability Evaluation - Results

- Linearly scales on the length of time series
- Fixed threshold shows lower answering time
- Increasing granularity makes answering faster
- All-Topic queries approx. by two orders slower
Contradictions were initially defined as textual inference and analyzed using linguistic technologies:

(De Marneffe et al., 2008; Harabagui et al., 2006).

Contradictions can be distinguished on a large scale by calculating the entropy or clustering accuracy:

(Choudhury et al., 2008; Varlamis et al., 2008).

Contradictions can also be analyzed using visual inspection of opposite sentiments:

(Chen et al., 2006; Liu et al., 2005).

Contradictive sentiments can be useful sources for opinion summarization:

(Kim and Zhai, 2009; Paul et al., 2010).
Thanks for your attention!