Event summarization from news and tweets correlation

Cavallari Sandro
Giglio Marco
Morettin Paolo

ABSTRACT

This report we consider the task to extract a summary of the main event that cause a shift on the opinion of the Twitter users. In particular we refer to those shift as *sentiment shift* and we try to capture the event that cause those sentiment shift. In this work we present different language-independent and semantic-less meteorologies and we try to figure out the pros and cons of those methods.

Introduction

The main goal of this work is to found a summary of the events that cause some sentiment shift. While NLP technique for event extraction are mature, their performance on tweets inevitably degrades, due to the inherent sparsity in short texts. Since Tweet contains heterogeneous language structure and are at most long 140 character, extracting event form tweets text is quite tricky: instead linking tweet to news allow us to extract event form news text that have a well formed structure and contains more text respect to tweet.

Due to the fact that out dataset is composed by labelled tweet, we manage to extract news form the New York Times archive basing the research on tweet's label. Starting form a good Tweet-News classification, allowed us to correlate tweet and news that belong to the same macro-topics avoiding to correlate object of different arguments.

In particular document compression is one of the main topic in *NLP* research field: usually this task is achieved using abstractive methods and language dependent technique. Our work, rather, want to be as much as possible language independent, so we propose solutions for this problem using technique based on:

- a simple Bag of Word combined with a Tf-idf approach

- a Latent Semantic Indexing that uses a SVD mathematical technique [?]

- a N-Gram Graph [?]

Related Work

RelatedWork

Problem Definition

ProblemDefinition

Proposed Approach

ProposedApproach

Experimental Evaluation

ExperimentalEvaluation

Conclusions

1. REFERENCES

[1] Indexing by Latent Semantic Analysis, Scott Deerwester , Susan T. Dumais*, George W. Furnas, Thomas K. Landauer and Richard Harshman

[2] Automatic Summarization from Multiple Documents : N-Gram Graph, George Giannakopoulos and Ncsr Demokritos, June 2009