Project Proposal

Name: _____

Date: _____

Introduction

Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication.

A Research Question

This research intend to determine a user's mood when he or she post a tweet or message on social network.

Existing Solutions and Limitations

Existing approaches to sentiment analysis can be grouped into four main categories: keyword spotting, lexical affinity, statistical methods and concept-level techniques.

Each of these approaches has its limitations. The basic disadvantage of keyword spotting technique is the dependency on the presence of obvious affective words in the text. For instance, the emotion "sadness" cannot be derived from the sentence "I lost my money" as the does not specifically mention the word "sad".

Lexical affinity operating solely on the word-level can easily be tricked by sentences such as "I avoided an accident" (negation) and "I met my girlfriend by accident" (connotation of unplanned but lovely surprise).

In order to leverage statistical methods such as naive Bayes, support vector machine. One must first model the unstructured text into a computable representation, normally, in a form of vector or matrix. Because of the diverse of human language, the curse of dimensionality problem is often encountered.

My Solutions

I propose to improve sentiment analysis by utilizing the information from social network environment. The first information is comment. I do so for two reasons. First, comment information is easily obtainable. Second, and more importantly, comments from friends always have hold similar attitude. It is the idea of the principle of homophily [1]. An study [2] found some evidence of homophily for both positive and negative sentiment among MySpace Friends.

The second information is previous content. Most sentiment analysis algorithms use simple terms to express sentiment. However, differing contexts make it extremely

difficult to turn a string of written text int a simple positive or negative sentiment. Therefore, it is necessary to consider the last several tweets as an useful context.

Experiment Design

Data Collection

I planned to adopt the straightforward approach to creating a labeled test set, namely, extract an arbitrary user's tweet from Weibo (A Twitter like Chinese social network) and label the tweets manually.

Expected Outcomes

References

[1] LAZARSFELD, P. F., AND MERTON, R. K. Friendship as a social process: A substantive and methodological analysis. In *Freedom and Control in Modern Society*, M. Berger, T. Abel, and C. Page, Eds. Van Nostrand, New York, 1954, pp. 18–66.

[2] THELWALL, M. Emotion homophily in social network site messages. *First Monday 15*, 4 (2010).