COLUMBIA UNIVERSITY – CUNY

Multimodal Sentiment Analysis

Logistic Progression – Jessica Ouyang and Joe Ellis

October 21, 2013

1 OVERVIEW

1.1 TASK

In this project we propose to take a multi-modal approach to sentiment analysis, and create the capability to extract sentiment from a variety of different sources, including videos, pictures, and text. Sentiment Analysis is a widely studied area of text analysis [4], but recently some work has been completed on visual sentiment analysis, such as SentiBank [2]. We propose to fuse information from visual and text signals to achieve a more complete representation of sentiment, and more accurately classify sentiment within a variety of sources. Specifically, we plan to classify Youtube news clips as containing either positive or negative sentiment. We propose to also study the difference in sentiment between varying mediums. For example, the change in sentiment on a subject between Twitter, Youtube, and Broadcast Video News may change drastically. We propose to explore these differences and propose new models for multi-modal sentiment analysis.

1.2 DATA USED

We will use a variety of on-line data sources for this project, as well as manually downloaded cable and broadcast news stories. We plan to harvest news clips from the CNN Youtube channel; each clip consists of a single story from a news broadcast. We propose to utilize social media sites such as Twitter, Facebook, and Instagram, to gain supplementary text and image data that can be processed for sentiment. If needed, we may also use the past year's worth of on-line news articles and news stories through the NewsRover project [3].

1.3 IDEAS ON TECHNIQUES

We plan to use co-training to classify videos. We will use one classifier based on the content of the video itself – facial expressions, gestures, and transcript – and one based on metadata – the video's title, description, Youtube comments, and related posts on social media. Metadata from social media can be gathered by searching for posts published within a few days of the video upload date that contain keywords from the video title and description.

1.4 WHY IS IT COOL?

This work builds on a very popular portion of research in a way that has not currently been explored. Multimodal analysis has shown promise in a variety of fields and sensor fusion techniques have become widely used. As we move more toward high-bandwidth data sources such as video and audio content, much of the sentiment that we create will be tied up in mediums other than text. Therefore, the lucrative field of sentiment analysis would benefit from the creation of a framework for multimodal data processing.

2 SUBJECT CONTRIBUTIONS

2.1 NLP CONTRIBUTION

This core of this project will be sentiment analysis, and the medium in which sentiment analysis is the most thoroughly developed is in text. The NLP novelty within this project is the ability to automatically combine text information from multiple different text sources (Twitter, news transcripts, and Youtube titles, descriptions, and comments). Combining these sources in interesting ways could add a novel portion to the typical NLP processing pipeline.

2.2 ML CONTRIBUTIONS

The core of this portion of the contribution is the fusion of features extracted from different modalities. These features are calculated from different spaces, and therefore can not be easily combined. Therefore, we look to find intelligent ways to fuse these different features, and this should be a contribution to the ML community.

2.3 WEB TECHNOLOGIES

We plan to create a program that automatically analyses content from multiple sources of available on-line web data. These sources could include, but are not limited to, Youtube and Twitter. We hope to create programs to process the public stream content that arrives from the accounts.

3 MEMBER CONTRIBUTIONS

We will collaborate on gathering the data and on training our models. We plan to split responsibility for feature extraction based on our backgrounds.

3.1 JOE ELLIS

I will work on the visual and audio portions of the research proposal, focusing on gesture and visual depiction of sentiment. Some features that can be used for visual sentiment are things such as gestures, scene, etc. I also plan to utililze my knowledge of some areas of audio processing to also be able to gleam sentiment analysis from that audio that exists within the videos.

3.2 JESSICA OUYANG

I will work on text-based features extracted from the video transcripts, titles, descriptions, Youtube comments, and social media posts. I have some experience in sentiment detection in Twitter posts based on the work of Agarwal, Biadsy, and McKeown [1].

REFERENCES

[1] Apoorv Agarwal, Fadi Biadsy, and Kathleen R Mckeown. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 24–32. Association for Computational Linguistics, 2009.

[2] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *ACM Multimedia*, October 2013.

[3] Brendan Jou, Hongzhi Li, Joseph G. Ellis, Daniel Morozoff, and Shih-Fu Chang. Structured exploration of who, what, when, and where in heterogeneous multimedia news sources. In *ACM Multimedia*, Grand Challenge, October 2013.

[4] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January 2008.