Affect Analysis in Large-Scale Online Text Communication Datasets

Nan-Chen Chen[1], Cecilia R. Aragon[1], Laurie Beth Feldman[2, 3], Judith F. Kroll[4]
{nanchen, aragon}@uw.edu, lfeldman@albany.edu, jfk7@psu.edu, aragon@uw.edu

[1] University of Washington, [2] The University at Albany-SUNY,
[3] Haskins Laboratories, [4] Pennsylvania State University

Affect Analysis in Online Text Communication

Online text communication, such as chats, blogs, forums, is widely used in both social and work settings. As more people use online text communication, it has become a fertile research site for understanding human behavior and interaction in a variety of fields, including psychology, social science, and computer science.

Previous studies (Aragon & Williams, 2011) have pointed out that affect plays a significant role in distributed collaboration. Other studies (Isen, Daubman, & Nowicki, 1987) also showed that affect has a positive influence on creativity, which is important in problem solving. As more communication takes place online, understanding how affect influences online communication becomes critical. Our goal is to support analysis of affect in large-scale online text communication.

Visual Analytics Tools

Visual analytics tools refer to visualization tools that aim to support analytic reasoning and sense-making. Visual representations of information leverage the high bandwidth visual input channel, and enabling interactivity with tools helps users use their vision for analysis. Figure 1 shows an example of a visual analytics tool.

There has been previous work building computational models for detecting and analyzing affect, but the chaotic essence of human communication makes it hard for programs to catch subtle emotional interactions. We will enable social scientists to leverage their knowledge about humans during the process of analysis. Visual analytics tools help make large-scale datasets manageable for these researchers to process.

Affect Labeler of Expressions (ALOE)

Affect Labeler of Expressions (ALOE) is open-source machine learning-based software for labeling affect in text. The tool was developed at the University of Washington (Brooks et al., ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW 2013). In this project, ALOE will be used to help label affect in our dataset as part of our data processing pipeline.

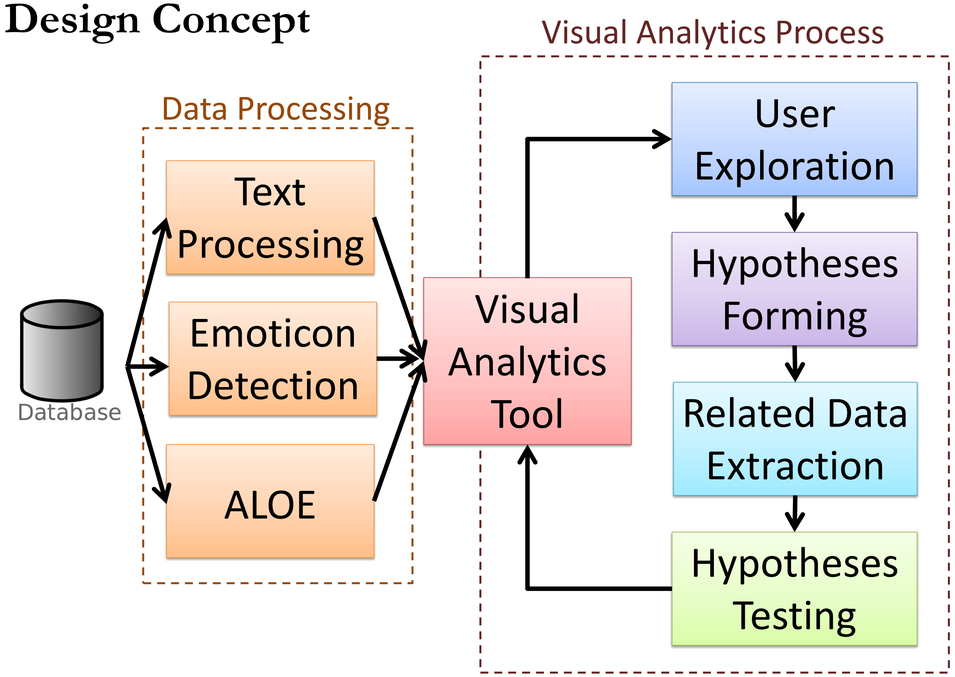
Research Question

How can we integrate text processing and machine learning in a visual analytics tool for affect analysis in text communication?

Scratch Dataset

The dataset we are going to use in our study is **Scratch Research Data** from the Scratch Online Community (http://scratch.mit.edu/) created by the Lifelong Kindergarten Group at the MIT Media Lab. Scratch is software designed to enable novices, especially children, to learn programming and computational thinking by creating stories. The Scratch Online Community is a website where users can create projects for their Scratch stories and collaborate with other users. The dataset includes the first five years (2007-2012) of data and contains a variety of meta information about the projects. We will focus on users' comments on the projects, about 8 million comments.

Design Concept



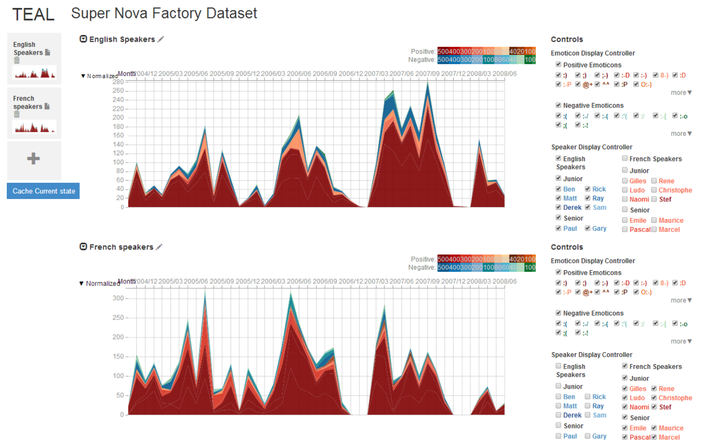


Figure 1. A visual analytics tool we previously built for emoticon analysis in another dataset. The two charts represent emoticon use of two different groups. From the charts we can see the two groups have different emoticon usage patterns. Though it is still a prototype, a few hypotheses of emoticon usage behaviors have been formed during analysts' interaction with it.

Research Plan

- In the first stage, we will focus on building the visual analytics tool with several rounds of user testing.
- In the second stage, we will focus on integrating the data processing components with visual interfaces.
- In the last stage, we will use the system we build to do affect analysis in our data set.

Future Goals

- Make the system online, open source, and open to the public.
- Use the system to form theories of affect in online text communication.
- Apply the system to other datasets.

For more information, contact Nan-Chen Chen at nanchen@uw.edu or Cecilia Aragon at aragon@uw.edu or visit http://depts.washington.edu/sccl

Scientific Collaboration & Creativity Lab
HCDE Human Centered Design & Engineering
dub
UNIVERSITY of WASHINGTON
(((Haskins Laboratories)))
UNIVERSITY AT ALBANY State University of New York
PENNSTATE University Park