Annotating Flames in Usenet Newsgroups

A corpus study by Melanie Martin

Introduction:

This study was undertaken as part of a research program, directed by Dr. Janyce Wiebe, aimed at learning to recognize subjective language in text automatically. ***Flames* in Usenet newsgroups, or on email listservs, are personal attacks containing hostile or abusive language.** Flames often provide an example of extreme subjectivity in natural language, in particular, negative evaluative language. We hypothesized that because of their extreme nature, flames might be relatively easy to recognize and might provide clues for recognizing subtler cases of subjectivity. In addition, it would be highly desirable to have an automatic system that would recognize flames, so that a user could choose whether or not to read, or to send a flame.

The corpus:

On October 4, 1999 we received a file from Computing and Networking at NMSU containing the top 100 newsgroups, in terms of volume, from the NMSU Usenet feed, with alt.binaries and alt.sex removed. On October 18, 1999 we received a second file containing the top 25 newsgroups, in terms of volume, in the comp and sci categories.

From each of four categories in the Usenet hierarchy: alt, comp, rec, and sci, we randomly chose 10 newsgroups. Then 10 threads were randomly chosen from each newsgroup, with the thread length cut off at six messages per thread. The concatenation of these messages is the corpus.

The corpus contains 1140 Usenet newsgroup messages. It was divided, preserving category balance, into a training set of 778 messages and a test set of 362 messages.

The task:

The annotators were instructed to mark a message as a flame if the "main intention of the message is a personal attack, containing insulting or abusive language." A given message could be classified as either a flame or a non-flame, along with a certainty rating from 0 to 3 (3 being most certain).

During the training phase, two annotators, M and R, participated in multiple rounds of tagging, revising the annotation instructions as they proceeded.

A number of policy decisions were made in the instructions, dealing primarily with included messages (part or all of a previous message, included in the current message as part of a reply). Some additional issues addressed in the instructions were who the attack was directed at, nonsense, sarcasm, humor, rants, and raves.

SAMPLE DATA

Xref: news.NMSU.Edu soc.religion.quaker:24379
<ANN flame="y" cert="2" /ANN>

Gfirenzi:
How old are you? We need to know for the record!

Actually, you should get your pastor Bob to post some messages here.... he sounds like an interesting chap. What type of church do you attend?

Incidentally, as long as you ignore any diversity, never listen to anybody else but your pastor, and never engage in any serious thought, you need not worry about being lead astray.

Sheshh, and we are called intolerant! I'm laughing and crying at the same time.

Oh well, anybody for some satanic rituals?
eric
--
eric s volkel

gfirenzi@my-dejanews.com wrote in message
<7d159n$1b6$1@nnrp1.dejanews.com>...
>I talked to my minster, Bob. Bob said i didnt understand his message. he
>was talking about how alot of puritans killed and did evil things in the name
>of christianity. And they killed people they called witches and sometimes
>the quakers that they didnt like. I showed him this internet news thing and
>he didnt like it. he said that i'm to young to be talking to people that can
>mislead me. he said that some of you people that profess to know jesus are
>full of hate and intollerence. bob said that this isnt like Jesus and
>Quakers are sopossed to like peace and and be in concensus, whatevet that
>means. so im sorry for calling you withches but i still pray that Jesus
>heals your hearts from all your meanness, because thats of the devil.

Results:

During the testing phase, M and R independently annotated the test set, achieving a kappa value on these messages of 0.69. A third annotator, L, trained on 492 messages from the training set, and then annotated 88 of the messages in the test set.

The pairwise kappa values on this set of 88 are:
M & R: 0.80;
M & L: 0.75;
R & L: 0.79;
average pairwise kappa of .78.

The distribution of flames to non-flames in the data is highly skewed in favor of non-flames. Thus percentage agreement results are high, as expected with such a skewed distribution. Spertus (1997) reports 98% agreement on non-inflammatory messages and 64% agreement on inflammatory messages. Our percentage agreement results are comparable. For example, the percentage agreement for M and R on the 362 messages in the testing phase was 92%.

The pairwise percentage agreement on the set of 88 messages:
M & R: 93%;
M & L: 91%;
R & L: 91%;
average pairwise percentage agreement of 92%.

Conclusions and future work:

This study provides evidence for the viability of document-level flame annotation. It has also created an annotated corpus, suitable for using supervised learning techniques to develop a flame recognition system. I plan to build a flame-recognition system in the future.

In a subsequent study, M and R also annotated the 362 test-set messages at the flame-element level. Flame elements are defined as the smallest element, or group of words, in a sentence, or message, which captures the flameyness (generally negative evaluative subjectivity). Results of this study are reported and used in Wiebe et al (2001).

Ellen Spertus: *Smokey: Automatic Recognition of Hostile Messages*, Innovative Applications of Artificial Intelligence (IAAI) '97. Also presented at the Eighth Annual Meeting of the Society for Text and Discourse, July 31, 1998.

Janyce Wiebe, Rebecca Bruce, Matthew Bell, Melanie Martin and Theresa Wilson: *A Corpus Study of Evaluative and Speculative Language*. 2nd SIGdial Workshop on Discourse and Dialogue, Aalborg, Denmark, September 1-2, 2001.