

CHAPTER #

## **BOO-HOORAY AND AFFECTIVE APPROACHES TO ETHICAL TEXTUAL ANALYSIS**

*Carson Reynolds*

*Department of Information Physics and Computing*

*University of Tokyo*

*+81 3-5841-6937*

*carson@k2.t.u-tokyo.ac.jp*

### **Abstract**

In *Language, Truth, and Logic* Ayer took the surprising position that ethical judgments, exhortations, and descriptions are neither true nor false, but are emotive statements. While his argumentation was in support of the logical positivist position, it suggests a novel method for computational systems to recognize utterances regarding ethics. Namely, expressions of emotion can be likened to “boos” or “hoorays” issued from a metaphorical crowd in moral response. This paper presents a simple natural language processing system that searches for terms and categorizes the text accompanying these terms as a “boo” or “hooray,” making use of a list of emotional terms compiled by Cowie et al. and orientations recorded by Whissel and Plutchik. The system uses this bag-of-words and a search engine to assign emotive scores to terms of the user's choosing. The contribution of this work is a primitive technique for computers to ethically evaluate textual queries.

### **Keywords**

Computer Ethics, Affective Computing, Natural Language Processing, Emotivism, Emotion Recognition

## COMPUTATIONAL EMOTIVISM

Emotivism, is a position that “judgments of value” are statements having neither truth nor falsity. More complexly, Ayer argues in Chapter 6 of *Language, Truth, and Logic* that a variety of statements regarding ethics cannot be the domain of philosophy [1952]. He describes four classes:

1. “propositions which express definitions of ethical terms, or judgments about the legitimacy or possibility of certain definitions”
2. “propositions describing the phenomena of moral experience, and their causes”
3. “exhortations to moral virtue”
4. “actual ethical judgments”

Of these, Ayer contends that only the first is “ethical philosophy.” The remainder he argues can be dealt with by the social sciences or classed as “mere pseudo-concepts.” In his view we “cannot argue ... the validity of these moral principles. We merely praise or condemn them in light of our own feelings.”

Recent research in affective computing [Picard, 1997] has pointed at a variety of methods for using computers to recognize information dealing with feelings. For instance, Liu et al. have demonstrated a system that analyzes text using common-sense reasoning to assess the (positive or negative) valence of text [2002].

In light of this recent work on computational processing of information dealing with feelings, a sort of “Computational Emotivism” is possible in which statements are analyzed for their content to determine if writers referring to the statement are “praising” or “condemning.” Or, more colloquially, if a group of documents is shouting “boo” or cheering “hooray” with regard to a statement.

## RELATED APPROACHES

A variety of different approaches have been considered for trying to make computers behave in an ethical manner. These include Weld and Etzioni's work to include the notion of "harm" into a planner to create ethical "softbots" [Weld, 1994]. Eichmann proposed an ethic for Internet agents and spiders to limit bandwidth abuse [Eichmann, 1994]. More germanely, LaMuth described an expert system with “affective language analyzer” that guides “motivational determination” of an artificial intelligence system [2003]. Additionally, there has been much work regarding textual analysis of affect that differs in approach, but often seeks the common goal of recognizing the sentiment or affective content of text. For instance, Pang et al. applied scores related to positive or negative sentiment [2002].

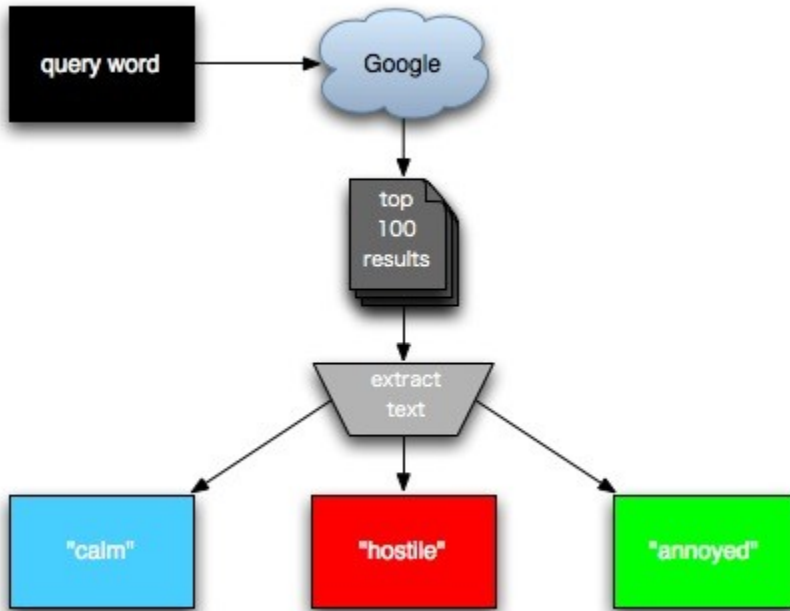


Figure 1: Boo-hooray processes web pages resulting from a Google search for emotional words.

### SYSTEM DESIGN

To better understand how boo-hooray compares to these existing systems, some detail will be provided about how the system operates. The boo-hooray system provides users with a search interface for words or phrases. These expressions are first sent to a search engine (Google) to find the 100 most relevant web pages. These pages are then downloaded into a cache and stripped of hypertext markup yielding a plain text document. These cached documents are aggregated into a common file representing the metaphorical crowd.

This set of pages representing the shouts of the metaphorical crowd are in turn processed to arrive at a list of the words sorted by frequency of occurrence. These words are then filtered to see if any have an associated emotional orientation. This is conducted by taking a list of “emotional words” compiled by Cowie et. al. that have been associated with emotional orientation. Emotional orientations have been assigned to words by subjects in psychology experiments conducted by Plutchik and Whissel [1989]. Segregating the frequency-sorted words into “boo” and “hooray” categories using emotional orientation, yields a set of emotional words associated with the original query and a ratio between the “boo” words and “hooray” words.

An informal metric for how “good” or “bad” a particular query is perceived is to take the sum of the “hooray” words and to subtract the sum of the “boo” words that occur in the search result. This number, which is dubbed “cheer,” is positive when there are more frequently occurring emotional words with positive orientations and is negative when there are more frequently occurring emotional words with negative orientations.

#### EXAMPLE OF USE

To better understand how the system operates and can give indications about good or bad sentiment regarding some query an example is in order. To test whether individuals had perceived an email program as “good” or “bad” a search with the name of the program “emotemail” was conducted. The following “boo” and “hooray” words were found:

<b>Boo Word</b>	<b>Boo Frequency</b>	<b>Hooray Word</b>	<b>Hooray Frequency</b>
critical	21	content	102
bored	9	happy	30
confused	7	interested	28
surprised	4	ready	19
angry	3	patient	9
disappointed	3	satisfied	2
rejected	3	accepting	1
annoyed	2	agreeable	1
bitter	2	calm	1
contrary	2	pleased	1
irritated	2	sympathetic	1
suspicious	2		
ambivalent	1		
awed	1		
disgusted	1		
jealous	1		
uncertain	1		

Table 1: Boo and hooray word frequency for emotemail query.

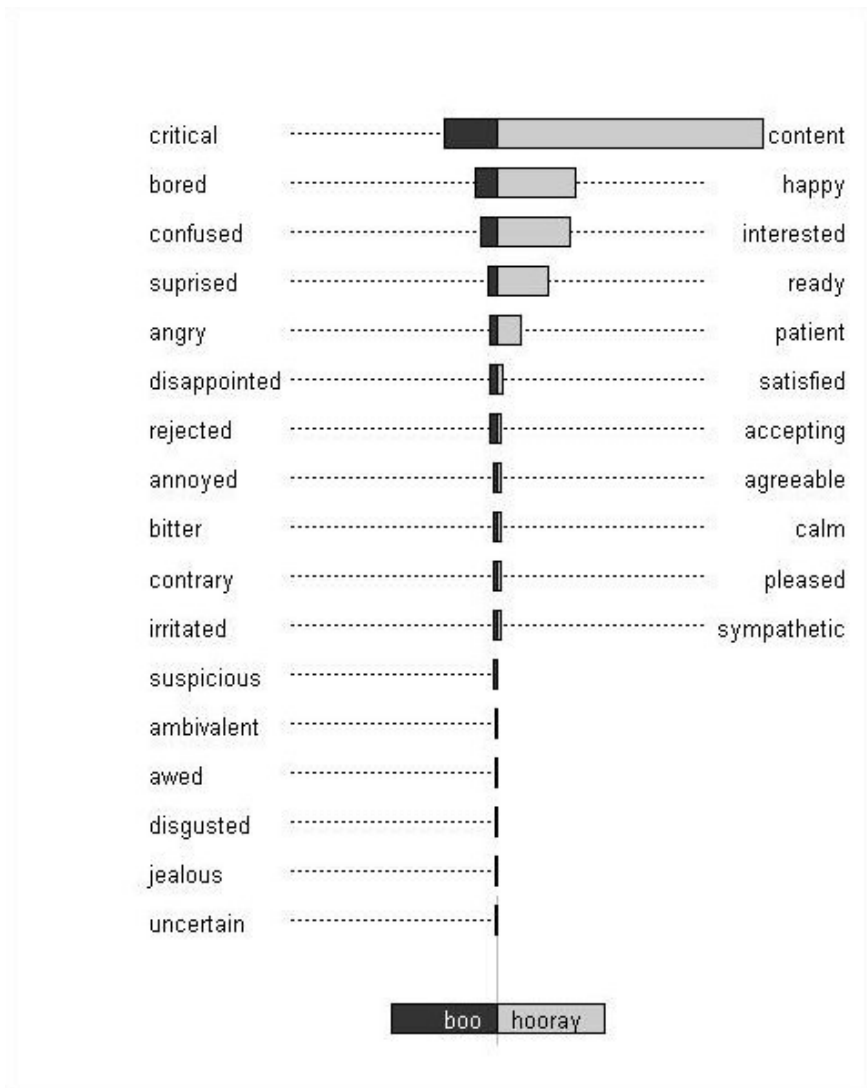


Figure 2: A histogram representation of the results of emotemail query.

The “cheer” associated with this query is 130, suggesting that the term is somewhat positively viewed in the documents deemed relevant by the search engine. As a point of contrast, if we search for a term that is intuitively more menacing such as genocide, then the results are markedly different:

<b>Boo Word</b>	<b>Boo Frequency</b>	<b>Hooray Word</b>	<b>Hooray Frequency</b>
rejected	8	content	8
critical	7	interested	8
confused	5	calm	5
contrary	5	ready	5
angry	4	pleased	3
surprised	4	sympathetic	3
bitter	3	generous	2
uncertain	3	satisfied	2
discouraged	1	delighted	1
disgusted	1	happy	1
indignant	1	trusting	1
stubborn	1		
suspicious	1		
vengeful	1		

Table 2: Boo and hooray terms from genocide query.

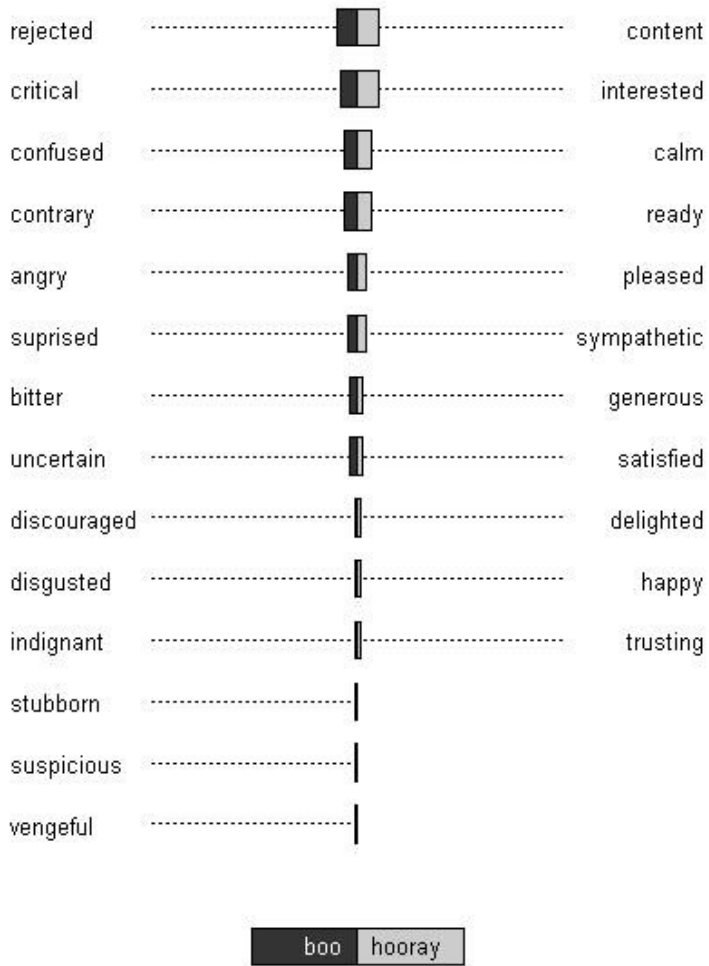


Figure 3: A histogram representation of the results of the genocide query.

The “cheer” associated with genocide is -6, indicating more negative than positive emotional words. Furthermore, there was a much smaller number of emotional words overall in the websites from the genocide query.

### IMPLICATIONS

Having seen an example of boo-hooray's operation, this paper will now discuss some implications of such a system. If we are to take the link between and emotional words and ethical statements at face value, then we have seen a crude

system that would allow individuals or programs to evaluate statements as either a boo or hooray.

What would such a system mean for emotivism and logical positivism as an enterprise? Emotivism (as we saw before) is dismissive of the comprehensibility of ethical statements. While a statement like “we have invaded Iraq” can be verified as either an analytic or synthetic statement, emotivism suggests statements like “we should invade Iraq” cannot be analyzed since it uses the normatively loaded “should.”

However, using the alternative explanation of emotivism (namely that ethical statements are expressions of emotion) boo-hooray has provided a different mechanism for the analysis of ethical statements.

Briefly, logical positivism uses a criterion to judge the meaningfulness and comprehensibility of statements. This criterion, the verification principle, requires that all well-formed statements be logically analyzable. Ayer argued that ethical statements are themselves meaningless essentially because they are not verifiable.

Boo-hooray does not change the verifiability of ethical statements in a logical positivist framework. However, in the less rarefied realms of the social sciences it provides a different variety of analysis. Boo-hooray produces a representation of public ethical opinion.

## CRITICISMS

There are a number of criticisms, however which ought to be considered when thinking about boo-hooray. These criticisms are of two varieties: technical and philosophical.

Technically, boo-hooray is not particularly innovative. From the stand point of natural language processing, the system uses a very naïve approach. The system basically employs a “bag-of-words” that is used to assess web pages.

This has a number of short-comings. One of which is that the system does not distinguish negative utterances from positive utterances. Namely, if a web page talks about being “not happy” or “not very excited” the system incorrectly assumes the author is happy and excited.

A more enlightened approach might try to analyze the sentences in which the emotional terms reside. We might ask what part of speech the words perform.

Thus terms that are homonyms such as “kind” (meaning both nice and a variety) can be analyzed when used as adjectives and not as nouns.

Another interesting problem with the system is the corpus of text used. The system's evaluations are limited to the top results provided by a Google search. This means that for a term to be analyzable a set of web pages must exist on the topic. Moreover, these web pages provide boo-hooray's only perspective while printed books and spoken word might offer a very different one. Boo-hooray then, is myopic seeing only the terms and words which are available in a convenient web based form. As such, it encodes a rather serious bias.

Furthermore, the system makes use of English emotional words. Other languages are currently excluded from its identification of ethical judgments. It would be much improved by including emotional words from more commonly spoken languages such as Chinese, Hindi, or Spanish.

Although, one may wonder if other cultures have different restrictions on emotional expression than those that occur in English language culture. On-line censorship on a variety of topics by countries such as Iran and China provides some evidence that the assumptions used by boo-hooray do not hold across cultures. Similarly, Japanese culture is known to favor masking displays of emotion, which would certainly impede the operation of boo-hooray.

Another problem with boo-hooray is that the orientations used by Plutchik Whissel do not map directly to approval and disapproval. This means that some words such as "awe" and "surprise" under the current scheme have a negative connotation. A much better approach would be to run a variation on the experiments used to construct the emotional orientation in which the words are explicitly separated into boo and hooray words in a context evoking ethical approval or disapproval.

This however, brings us to another set of problems with boo-hooray that are philosophical in nature. One might ask: “Do two categories cover the entire breadth of human emotion?” More extremely, “do **any** number of categories cover the entire breadth of emotion experience?”

There is little agreement on which (if any) models of emotion do the best job of describing the phenomena. Currently, a number of categorical models, different axes (such as valance and arousal), and emotional orientations have been attempted. However, none of these models does perfect justice to the range and subtlety of human emotions, which may be mixed, conflicting or very hard to express. Indeed, functional models based on the underlying neurological

behavior may be needed and even these may not capture the qualia of emotional experience itself.

What exactly is emotional experience? A precise answer to that question would certainly be worthy of a mind greater than my own. However, I can say that emotional experience is not words. What I mean is that the words that are used to express emotion are not the same thing as the emotions themselves. They are a mediation of emotion and as such have some distance from the thing that they represent. A writer of a web page may be writing about emotions which he or she is presently experiencing or experienced some time ago or have not even experienced. But in any of these cases, the writer cannot directly take his or her emotions and record them in one-to-one correspondence as English words.

In this case, boo-hooray suffers from a serious flaw. It cannot analyze emotions directly but only highly mediated emotional utterances which may be so divorced from the originating emotional experience as to introduce a skewing. But then any approach which relies on linguistic analysis of emotions will suffer from some problems due to the process of expressing emotional experience.

There are indeed very deep questions about to what degree computational analysis is able to comprehend human language. The ability to store and transform symbols suggests some elements of the process of comprehension. However, users of software which attempts to translate one language to another are likely to be keenly aware of the limitations of machine comprehension of text. Currently, human readers have a number of advantages in understanding a piece of text which computers have yet to perfectly imitate. These include social contextualization and common-sense reasoning skills along with linking of words not just to symbols but to a large number of memories.

And still there is another set of philosophical criticisms that can be leveled. These have more to do with the ethical approach of boo-hooray. In all likelihood, a large number of ethicists reject emotivism as a metaethical stance since it is dismissive of ethics. For those who do accept emotivism, the following argument might be relevant. Is not Ayer's analysis implying that ethical judgments cannot be analyzed precisely because they are emotional? Namely, in the work we have veered out of the realm of philosophy into that of psychology (satisfying neither).

Those who reject emotivism might rightly point out that logical positivism is largely outdated. Since its heyday much work has been done to revise and improve upon logical positivism's ideas. Popper's re-framing of positivism in

terms of a principle of falsifiability comes to mind as one of the more important [1963].

It may then be worthwhile to reconsider emotivism from the perspective of falsifiability as opposed to verifiability. We might then state that an ethical statement must be one which is disprovable. Thus, if one were to say “I am the rightful king because of the mandate of heaven” it would not be falsifiable. However a statement like “The U.S. civil war was bad because more Americans died in it than any other war” may be falsifiable by virtue of the testability of the component following the “because.”

If this were the case then we might have a procedure for separating ethical statements into falsifiable and non falsifiable varieties, but we still lack a procedure for separating statements regarding ethics from other statements.

#### FUTURE WORK

As we have seen, boo-hooray is subject to a number of criticisms and limitations. There are some technical steps which can be taken to improve upon this initial prototype.

Foremost among these would be to make the system available for use on the web. Currently, the system takes the form of a set of Python, Sed, Awk, and Bash shell based programs for text processing. These programs can be made available through a CGI (common gateway interface) web application.

As was discussed above, boo-hooray's sophistication as a text processing system could easily be improved. Modeling the text as a series of  $n$ -grams would allow statistical analysis of the co-occurrence of query terms and emotional language [Brown et al., 1992]. Extracting parts of speech could help focus the system's analysis to emotional language used to describe a query term.

Another related area for improvement is the use of more robust metrics than frequency of occurrence. For instance, the notion of salience, which normalizes the number of occurrences of a word in a corpus could be employed [Whitman et al., 2003]

#### CONCLUSION

Boo-hooray is presented here as an example of a system which explores a method for identification and analysis of statements regarding ethics. The system should be viewed as a philosophical experiment or conversation piece, an artifact around which criticism and debate regarding the nature of ethics can take place.

## REFERENCES

- Ayer, A. J. (1952) *Language, Truth, and Logic*, Dover Publications, New York, NY.
- Brown, P. F., deSouza, P. V., Mercer, R. L., Pietra, V. J., and Lai, J. C. (1992) Class-based  $n$ -gram models of natural language. *Computational Linguistics*, 18, 4 (Dec. 1992), 467-479.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001), *Emotion Recognition in Human-Computer Interaction*. *IEEE Signal Processing Mag.*, Jan 2001, 32-80.
- Eichmann D. (1994) Ethical web agents. *Computer Networks and ISDN Systems*, 3, 3-13.
- Liu, H., Lieberman, H., and Selker, T. (2002) Automatic affective feedback in an email browser. In MIT Media Lab Software Agents Group Technical Report SA02-01.
- LaMuth, J. E. (2003) Inductive Inference Affective Language Analyzer Simulating AI, United States Patent # 6,587,846.
- Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proc. EMNLP 2002*, 79-86.
- Popper, K. R. (1963). *On the Sources of Knowledge and of Ignorance*. In *Conjectures and Refutations* New York: Harper & Row.
- Picard, R. W. (1997) *Affective Computing*. Cambridge MA:MIT Press.
- Plutchik, R. (1980) *Emotion: A Psychoevolutionary Synthesis*. New York: Harper & Row.
- Weld D and Etzioni O. (1994) The first law of robotics (a call to arms). *Proceedings of the 12th National Conference on Artificial Intelligence*, 1042-1047. Seattle, Washington.
- Whissel, C.M. (1989). *The dictionary of affect in language*. *Emotion: Theory, Research and Experience: Vol. 4, The Measurement of Emotions*, R. Plutchik and H. Kellerman, Eds. New York: Academic.
- Whitman, B., Roy, D., and Vercoe, B. (2003) Learning word meanings and descriptive parameter spaces from music. In *Proceedings of the HLT-NAACL 2003 Workshop on Learning Word Meaning From Non-Linguistic Data - Volume 6 Human Language Technology Conference*. Association for Computational Linguistics, Morristown, NJ, 92-99.