User Evaluation of a System for Classifying and Displaying Political Viewpoints of Weblogs

Alice Oh and **Hyunjong Lee** and **Youngmin Kim**
alice.oh@cs.kaist.ac.kr, {lhjgine,creatcross}@kaist.ac.kr
Computer Science Department
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea

Abstract

This paper presents a Web-based user evaluation of a system for classifying and presenting political viewpoints of blog posts. The system is based on a classification model trained using a supervised learning algorithm, and the data set consists of recent posts from blogs that are self-identified as a liberal or a conservative viewpoint. We first discuss the classification process. Then, with a prototype system for retrieving and classifying political blogs, we look at how the classification results can be presented to users in order to improve the blog search experience. We describe an online user study with 15 users, and the study shows that users preferred the search results page that clearly shows the political viewpoint classification.

Introduction

Much of the Web content has been based on facts, and thus the Web search engines concentrated on retrieving Web documents based on topics such as tf-idf, supplemented with page importance information such as PageRank (Brin & Page 1998). On the other hand, user-generated and unmoderated content, such as blogs, do not always focus on facts. A quick blog search on Google (http://blogsearch.google.com) using the query "barack obama" brings up top results that include commentaries about his cabinet picks and opinions about Obama's religious activities. This is in stark contrast to a general Web search on google (http://www.google.com) using the same query which shows top results including a Wikipedia entry and official websites. Furthermore, the language used in blogs, in terms of subjectivity and formality, differs much from other Web documents (Mishne 2006).

We hypothesize that this is a reflection of the needs and intentions of users when they write, read, and search blogs. In the traditional Web search, user intentions typically fall in three categories: navigational, transaction, and informational (Broder 2002). A thorough study of blog logs by (Mishne & de Rijke 2006) revealed that much of users' intentions in blog search are informational, but the users' interests are quite different for blogs, shown by marked differences in query terms used. (Hearst, Hurst, & Dumais 2008) advocate a faceted search interface, with the assumption that when users read or search for blogs, they look beyond topics and the traditional definition of relevance; they look for different styles, attitudes, and other non-topical dimensions. Our work looks at one of the non-topical dimensions, viewpoint, with the goal of starting a discourse about the significance of non-topicality in weblogs, both in analyzing and presenting them to users, especially in an attempt to reach beyond opinion and sentiment mining.

	Dataset 1		Dataset 2	
	Cons	Liberal	Cons	Liberal
Docs	1218	1386	2066	2489
Spanish Docs	19	53	197	241
Empty Docs	30	48	17	31
< 50 Wds	162	208	217	283
Words	577236	604774	841227	1055801
Avg Sent/Post	22.7	20.2	18.0	17.4
Avg Wds/Post	473.9	436.3	407.2	424.2

Table 1: Detailed statistics of data sets 1 and 2.

Political Blog Data Collection

To learn a model for predicting political viewpoints of posts using supervised learning algorithms, we used two blog catalog sites, blogcatalog.com and blogarama.com that provide blog directories that have blogs tagged conservative or liberal. We collected more than 1000 URLs of conservative and liberal blogs, but among those, we use 286 liberal blogs and 117 conservative blogs, keeping only blogspot blogs and recently active blogs. Further, we filtered out posts that are not written in English, and also posts that are very short (empty or fewer than 50 words). We built two data sets, the first set for testing our classification system, and the second for carrying out a user study. For the first data set, we used as keywords "obama," "mccain," and "election," and dates between 2008-09-30 and 2008-11-29. For the second dataset, between 2008-11-19 and 2009-01-13, and using the search keywords: "senate", "congress", "election", "obama", "president", "israel", "gaza", "middle east", "hamas", "economy", "bailout", "bush", "iraq", "afghanistan", "tax". Table 1 shows the detailed statistics about the two data sets.

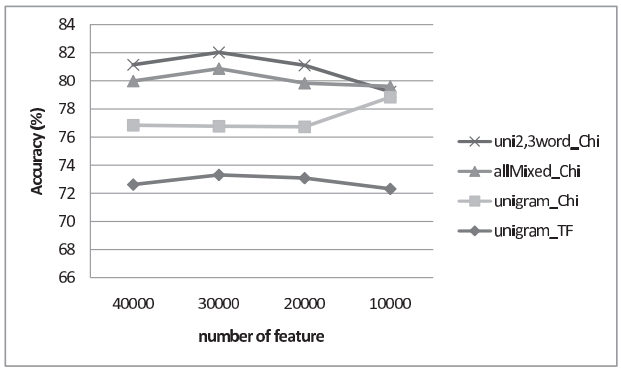

Figure 1: Classification results using SVM and 1-word, 2-word, 3-word term frequencies, co-occurring word statistics, and χ^2 feature selection.

Classification of Blog Posts

Using data set 1, we designed and implemented two classification algorithms: naive Bayes and support vector machines. Our results are better than one similar system (Malouf & Mullen 2008) and worse than others: (Lin, Xing, & Hauptmann 2008) and (Jiang & Argamon 2008), all using supervised training based on similar features. However, it is difficult to compare the three systems because the data, as well as problem formulation, are quite different.

We use three basic term frequency features: single word, 2-word phrases, 3-word phrases, and word co-occurrence. There are more than 900,000 features in our training data, and we used χ^2 feature selection (Manning, Raghavan, & Schütze 2008) to decrease the number of features.

For our SVM experiment, using LIBSVM (Fan, Chen, & Lin 2005), we tested the χ^2 feature selection with the baseline of single-word term frequencies. We tested with feature sizes from 1000 to 49000 with a step size 1000. The χ^2 feature selection improves classification accuracy, peaking around 8000 features at 79%. Next, we compared the different feature sets using the (χ^2) feature selection from, varying the size of the total features from 10000 to 40000. The four feature sets are:

- unigram_tf: single words, selected by term frequency
- unigram_chi: single words, selected by χ^2
- uni2,3word_chi: 1, 2, and 3-word phrases, selected by χ^2
- allmixed_chi: 1, 2, 3-word phrases, and co-occurrence features, selected by χ^2

Figure 1 shows the results of the experiment. The uni2,3word_chi feature set shows the best classification performance at 82% accuracy.

We performed similar experiments with the Naive Bayes classifier using the same feature selection method but with only the general unigram features and co-occurrence unigram features . The feature sets are as follows, and the number in parentheses are the f-scores for NB classification using those feature sets.

- multinomial (75.13): all unigram term frequency features

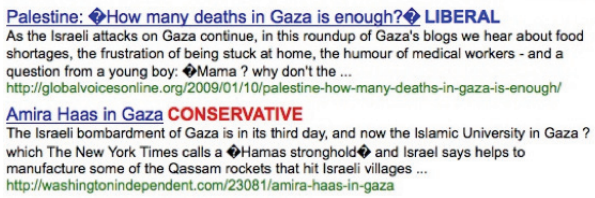

Figure 2: A "tagged view" of the search results where every result is tagged with the political viewpoint information.

- χ^2 (74.65): unigram features selected by χ^2 values
- co-occurrence (63.73): unigram features of only sentences where the keywords appear
- voting (74.81): best 2 out of 3 of the three features
- sum (74.34): taking the sum of the first three features

Showing Political Viewpoints to Users

One important question about blog classification based on political viewpoint is how we can use the results of classification to improve the experience of blog readers and searchers. (Hearst, Hurst, & Dumais 2008) poses a related question of "What should blog search look like?" and argues that a blog search is different from a general Web search, and the search interface should look different. Motivated by these questions, we designed and conducted a user study with a prototype blog retrieval system based on our political viewpoint classification presented in earlier sections.

Blog retrieval is usually categorized into two groups based on the type of the query: ad-hoc queries and filtering queries. Ad-hoc queries are just like the common Web searches where a user types in a query every time he/she performs search. Filtering queries are performed when a user specifies a number of topics he/she is interested in, and the blog search engine automatically generates queries, periodically looking for new blogs that may serve the user's interests. Our system is designed for the ad-hoc retrieval, which may be a smaller of the two types, according to (Mishne & de Rijke 2006), comprising only 19% of all queries, but a much larger portion, 70%, of unique queries.

We designed and implemented three different views of the results page for an ad-hoc blog query. The first view is the baseline and is almost identical to the results page view of a Google blog search (http://blogsearch.google.com). It does not show any political viewpoint information. The second view, shown in Figure 2, is a slight variant of the first view, where an additional viewpoint information is shown in capital letters and different colors ("LIBERAL" in blue, "CONSERVATIVE" in red) next to the link for each result. The third view, shown in Figure 3, is a two-column view, with the "liberal" posts on the left and "conservative" posts on the right. This simple, yet obvious, design of the three different views for the results page allows us to conduct a user study to ask whether analyzing and showing political viewpoint of a blog post is useful for an ad-hoc blog retrieval system.

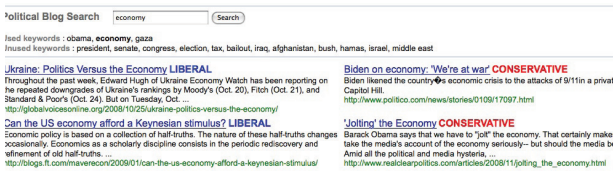

Figure 3: A "two-column view" of the search results where the classified results are presented in two columns, each column representing one of the two political viewpoints.

User Study

For the user study, we wished to simulate a real-time Google blog search as much as possible, so we collected a more recent data set, shown in Table 1, and used those for the prototype blog retrieval task. The data were collected using a set of keywords that we wished to let users type in as queries. Because those keywords represent different topics, we built five different classification models by clustering the keywords by topic and building different models for the different topics. For simplifying the clustering process, we used keyword co-occurrence, and after analyzing keyword co-occurrence, we made five groups of keyword: {obama, president}, {senate, congress, election}, {economy, tax, bailout}, {Iraq, Afghanistan, bush}, {hamas, Israel, gaza, middle east}. F-scores for classification on these groups are, 74.8, 72.0, 76.6, 73.2, and 80.4, respectively.

We recruited 15 subjects who are students and employees at an American university for a Web-based study, and we paid them $15 each for participating. They were first asked to give their political views (optional), then after a brief introduction, began the study by typing in search queries using one of the keywords we provided. Each subject saw all three views of the results page, but they were randomly assigned to an ordering of the views. For each results page, we asked them to freely explore the results, just as they would for their everyday Web searches. Before the fourth query, we showed them the images of the three views they already saw, and asked them to choose a preferred view among them. After examining the fourth results page, they went on to answer questions about the search experience. The questions were:

1. Of the three views, which did you choose for your last search? Why?
2. (For each view) Please rate, on a scale of 1 to 5, the following characteristics:
   (a) Easy to find posts you want to read
   (b) Satisfied with the search results on the results page
   (c) Satisfied with the posts you chose to read
3. How often do you read blogs on US politics?
4. What are your reasons for reading blogs?

Results and Discussions

As it shows in Table 2, 9 out of 15 users chose the two-column tagged view. We interpret this, combined with the reasons users stated, to mean that showing political viewpoint information for each blog post is helpful for users of ad-hoc blog searches. A more carefully designed experiment would have a two-column view with no viewpoint information, to rule out the possibility that users chose the two-column view because it simply allows them to view more results in a more compact view. We will definitely consider that for our next user study.

View	#	Reasons
plain	5	I don't like the tags. They're too polarizing It provided a less biased option I like to decide for myself what is lib/cons
tagged	1	liked the tags
two-col	9	Easier to sort out ones you want to read Nice to see comparison I like to read a couple from each side I like to know something about the author

Table 2: Users' preferences among the three different views and representative reasons for those preferences.

	plain	tagged	2-col
ease-of-use	**2.73**	2.27	**1.93**
satisfied-results	**2.66**	**2.06**	2.2
satisfied-posts	2.33	2.26	2.2

Table 3: Averages of users' ratings, on a scale from 1 (good) to 5 (bad), on the questionnaire.

Table 3 shows how users responded to the questions. ANOVA shows that none of the three questions are significantly different for all three conditions, but pairwise t-test shows that the numbers in bold are significantly different. We can see the pattern that users prefer either the tagged view or the two-column tagged view, but to get a more definite answer, we would need to have more subjects.

Another set of results we looked at are the number of clicks on the posts classified as "liberal" or "conservative". As Table 4 shows, liberal users clicked on more liberal posts than conservative posts for the tagged and the two-column views. We interpret this to mean that, when users have information about the political viewpoint of a post, they make more informed decision about which posts to click. It would be more useful to know whether they were satisfied with each click, via an explicit feedback, or an implicit feedback such as the amount of time they spent looking at that post.

Lastly, we asked users why they read blogs (multiple options and write-ins allowed), and nine out of fifteen users said that they read for a variety of viewpoints, and six users said they read for a specific viewpoint, and seven users said they read for other users' comments. As we consider more factors for blog search interfaces, we would need to carry out a larger scale user study to reveal their behavior, needs, and intentions, but this is an adequate first step that gave us some insights for our next study.

Related Work

There are three lines of work that are related to the present research. They are opinion or sentiment classification, clas-

View	Post	LibU (7)	ConsU (1)	OtherU (5)
Plain	Lib	11	1	12
	Cons	15	1	3
Tagged	Lib	13	1	6
	Cons	10	3	4
2-Col	Lib	30	3	9
	Cons	17	4	11
Total	Lib	50	5	27
	Cons	42	8	18

Table 4: Number of liberal and conservative posts users clicked on for each view by user type: LibU (liberal users), ConsU (conservative users), OtherU (other users).

sification of political articles, and blog retrieval. Opinion and sentiment mining has been very actively researched, and (Pang & Lee 2008) offer an excellent survey of the field. (Yu & Hatzivassiloglou 2003) is one of the earliest works on sentiment classification, and (Chesley *et al.* 2006) is an example of sentiment classification in blogs. However, political viewpoint classification is very different from sentiment classification. Using SentiWordNet (Esuli & Sebastiani 2006) on the words with highest χ^2 values in our data, 86 out of 100 words have O (objective) score of 1, meaning those words do not carry much sentiment.

There has been some recent work on political view classification. One of the earlier works is (Lin 2006), using the bitterlemon corpus consisting of Israeli and Palestinian authors' editorials. (Malouf & Mullen 2008) is a similar line of work, classifying users of an online community according to viewpoint. (Lin & Cohen 2008) and (Kale 2007) present political leaning classification based on link analysis. (Jiang & Argamon 2008) describe something that is closest to our work, but they classify blogs rather than individual blog posts.

There are some good examples of user interfaces for blog retrieval. BlogPulse (http://www.blogpulse.com) is a website that follows and shows the current "trend" in blogs. (Gregory *et al.* 2007) describe a system for visualizing blogs, and (Fujimura *et al.* 2006) have developed a multi-faceted blog search engine with an interesting user interface.

Conclusion and Future Work

We presented a prototype system for classifying and presenting political viewpoints of blog posts. We believe that non-topical dimensions, such as viewpoint, are important for both analyzing and presenting blog posts in an ad-hoc blog search system. Our user study showed some promising results of a simple but useful interface for blog search, and we will further experiment with classification and user interfaces for non-topicality in blogs and other social media.

References

Brin, S., and Page, L. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*.

Broder, A. 2002. A taxonomy of web search. *ACM SIGIR Forum* 36(2):3–10.

Chesley, P.; Vincent, B.; Xu, L.; and Srihari, R. 2006. Using verbs and adjectives to automatically classify blog sentiment. *AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs*.

Esuli, A., and Sebastiani, F. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of LREC*.

Fan, R.; Chen, P.; and Lin, C. 2005. Working set selection using second order information for training support vector machines. *The Journal of Machine Learning Research*.

Fujimura, K.; Toda, H.; Inoue, T.; and Hiroshima, N. 2006. Blogranger—a multi-faceted blog search engine. *Proceedings of the WWW 2006 3nd Annual Workshop on the Weblogging Ecosystem*.

Gregory, M.; Payne, D.; McColgin, D.; Cramer, N.; and Love, D. 2007. Visual analysis of weblog content. *Proceedings of ICWSM*.

Hearst, M.; Hurst, M.; and Dumais, S. 2008. What should blog search look like? *Proceedings of Search and Social Media Workshop, CIKM 2008*.

Jiang, M., and Argamon, S. 2008. Exploiting subjectivity analysis in blogs to improve political leaning categorization. *Proceedings of ACM SIGIR 2008*.

Kale, A. 2007. Modeling trust and influence on blogosphere using link polarity. *Proceedings of International Conference on Weblogs and Social Media (ICWSM)*.

Lin, F., and Cohen, W. 2008. The multirank bootstrap algorithm: Semi-supervised political blog classification and ranking using semi-supervised link classification. *CMU LTI Technical Report*.

Lin, W.; Xing, E.; and Hauptmann, A. 2008. A joint topic and perspective model for ideological discourse. *Proceedings of the 2008 ECML*.

Lin, W. 2006. Identifying perspectives at the document and sentence levels using statistical models. *Proceedings of the 2006 NAACL*.

Malouf, R., and Mullen, T. 2008. Taking sides: user classification for informal online political discourse. *Internet Research*.

Manning, C. D.; Raghavan, P.; and Schütze, H. 2008. Introduction to information retrieval.

Mishne, G., and de Rijke, M. 2006. A study of blog search. *Proceedings of ECIR '06*.

Mishne, G. 2006. Information access challenges in the blogspace. *Proceedings of International Workshop on Intelligent Information Access*.

Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Information Retrieval*.

Yu, H., and Hatzivassiloglou, V. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. *Proceedings of the 2003 EMNLP*.