

Weblogs for Market Research: Improving Opinion Detection using System Fusion

Deanna J. Osman, John L. Yearwood, and Peter W. Vamplew

School of Information Technology and Mathematical Sciences – University of Ballarat

P.O. Box 663, Ballarat Victoria 3353, Australia.

Email: (d.osman, j.yearwood, p.vamplew)@ballarat.edu.au

Abstract— Searching for opinions on a specific product or service within blogs is a new frontier for market researchers. This research investigates the use of system fusion methods to improve Mean Average Precision (MAP) results achieved by the Text REtrieval Conference (TREC) Blog06 participants and reports the improved MAP results. It is hypothesized that diversity of the inputs is vital to maximising the MAP improvements. This is shown in the improvement in MAP values achieved by some of the participant's ranked lists. The growth in the number of blog authors who write valuable opinions about their life experiences has led to an unsolicited resource of opinions on products, politics and services. In 2006, TREC collected blogs and set a task of detecting opinions on given topics to their participants, reporting the results using MAP.

Index Terms— Fusion, Opinion Detection, TREC, Weblogs.

I. INTRODUCTION

The number of people regularly accessing the Internet is reported to have grown by 244.7% worldwide between 2000 and 2007 [1]. One area recording a high level of growth on the Internet is weblogs (blogs). In December 2007 a blog tracking company, Technorati, Inc., reported that it was monitoring 112.8 million blogs worldwide [2], up from 4.2 million in October 2004 [3].

Along with the growth in the number of blogs on the Internet, there is a growth in interest in the content of blogs, particularly opinions within blogs. The majority of blog authors surveyed by Lenhart & Fox (2006) [4] indicated that their reason for blogging is to share their knowledge, skills and life experiences. Often bloggers will express their opinions about products, events and people which impact their lives.

These un-solicited opinions could prove invaluable for market research by organisations who wish to gauge reactions to products and services. For example, negative opinions about a competitor's product may provide a competitive edge for a new design or governments could search blogs for qualitative research (opinion polls) regarding new policies or upcoming elections. Small businesses, who do not have a large 'market research' budget, could gain access to millions of people who potentially have an opinion relating to them.

Most users will not read all documents returned by a search

engine. Jansen et al. [5] found that 58% of users do not read more than the first page of a list of relevant documents. Therefore, the aim of this research is to use a ranked list as input and float the documents with relevant opinions to the top of the list and the remaining documents to the bottom of the list. The resulting list with the opinion-bearing blogs at the top would be returned to the user.

In 2005/06, the Text REtrieval Conference (TREC) created a blog collection (Blog06), comprising of 3.2 million blog posts and comments. The tasks in 2006 and 2007 for the Blog06 collection included an 'Opinion Retrieval Task' where participants retrieved blogs expressing an opinion on one of 50 given topics. Participants could submit up to five runs¹, which included retrieved documents expressing an opinion on a given topic [6]. A total of 56 runs consisting of the top 1,000 opinion-bearing documents for each topic² were submitted by the 14 Blog06 participants. Of these, the top 100 documents from 27 runs were combined with the top 10 documents from the remainder of the runs to create a list of blog documents to be assessed by TREC assessors [6].

Once the assessments were available, the Mean Average Precision³ (MAP) value was calculated for each run, measuring the precision of retrieval of documents relevant to the given topic (irrespective of whether an opinion existed on the topic within the document) and the documents containing an opinion on the given topic. The results were published in Ounis et al. (2006) [6]. MAP is the standard reporting method of TREC corpora. Precision (P) is calculated using formula (1a) and formula (1b) calculates average precision (AP). Mean Average Precision is the mean of the AP value for 50 topics.

$$R_i = \begin{cases} 0 & \text{if document}_i \text{ is not relevant} \\ 1 & \text{if document}_i \text{ is relevant.} \end{cases}$$
$$(a) P_i = \frac{\sum_{j=1}^4 R_j}{i} \quad (b) AP = \frac{\sum_{i=1}^N R_i P_i}{N} \quad (1)$$

The level of agreement achievable by human assessors, when assessing the existence of an opinion on a given topic,

¹ A 'run' is a ranked list of relevant documents submitted to TREC by the participants.

² TREC provides the participants with 50 topics for this task.

³ The calculation of MAP is explained in the document located at http://trec.nist.gov/pubs/trec14/appendices/CE_MEASURES05.pdf

was studied by Osman et al. (2007) [7]. The level of agreement achievable by multiple assessors was calculated to gain insight into the level of accuracy which could reasonably be expected from an opinion detection system [7]. The low level of agreement between the assessors raised questions about the use of single assessors⁴ in opinion detection tasks. A higher level of agreement between the assessors and the TREC assessments was obtained, when the majority vote of committees formed from the assessors, was compared to TREC assessments [7]. The results obtained led to experimentation in using the output of the TREC participant's systems as input for system fusion methods, hypothesizing that improved MAP results could be obtained using a 'majority vote' of ranked lists. The system fusion methods applied in this research are not particularly sophisticated as the focus of this research was on evaluating the suitability of system fusion in opinion detection research.

Extant research has found that there is little overlap in the documents retrieved between different information retrieval (IR) systems [8], [9]. Fusion has been used in both IR [9], [10] and classifier [11], [12]. In this paper we apply several system fusion methods to take advantage of the differing content in the lists to improve the MAP values achieved by the individual TREC Blog06 participants, and hypothesize that diversity within the inputs is vital to improving achievable results. Related work is reviewed in section II, with the various methods of system fusion being described in section III. The results of the experiments are discussed in section IV while section V presents conclusions and future work.

II. RELATED WORK

A. Opinion Detection

The task of opinion detection has previously been investigated in a non-blog context. Editorial and Letter to editor articles were assumed to be opinion-bearing, while Business and News articles were categorised as non-opinion-bearing [13], [14]. These documents were used to create a list of opinion-bearing and non-opinion-bearing words and opinion scores⁵ [13], [14]. Synonyms and antonyms were used to expand the list [14], [15].

The original Wall Street Journal articles in each category were not evaluated to determine the validity of the original [13], [14]. Newswire articles are written using a formal structure, using proper English without slang and word abbreviation. On the other hand, blogs can be an informal form of communicating, using Internet abbreviations and slang [16]. The resulting list of terms and opinion weights was used as training data in the opinion detection task in Blog06 [17].

TABLE I
AN EXAMPLE OF THE VOTES METHOD

Run 1		Run 2		Votes result		
Doc Num	Rank	Doc Num	Rank	Votes	Doc Num	Rank
d ₈	1	d ₃	1	2	d ₁	1
d ₉	2	d ₈	2	2	d ₂	2
d ₆	3	d ₉	3	2	d ₃	3
d ₂	4	d ₂	4	2	d ₆	4
d ₁	5	d ₁	5	2	d ₈	5
d ₃	6	d ₆	6	2	d ₉	6
d ₇	7	d ₁₂	7	2	d ₁₀	7
d ₁₀	8	d ₄	8	2	d ₁₂	8
d ₁₂	9	d ₁₀	9	1	d ₄	9
d ₅	10	d ₁₁	10	1	d ₅	10
d ₄		d ₅		1	d ₇	11
d ₁₁		d ₇		1	d ₁	12

B. System Fusion

Fusion is a common term used for combining multiple IR systems or queries in an attempt to improve the performance of individual systems [18]. Many IR researchers have used fusion techniques in an attempt to improve retrieval results and address the two primary errors that can occur in IR: (1) low-ranking relevant documents and (2) high-ranking non-relevant documents [10]. System fusion poses one possible solution to these errors by taking advantage of training methods for various systems [19].

There are several methods of system fusion, including similarity merge and weighted sum [20]. Similarity merge uses the similarity scores assigned during the retrieval process to merge the ranked lists, attempting to float the relevant documents to the top of the resulting ranked list [10], [19], [21]. An alternative to this method is weighted sum [9], [22], where experts assign weights indicating the relevance of a document to a query and these weights are combined. The ranked lists used as input for the research reported in this paper come from a diverse range of sources. As such, the similarity scores were not considered appropriate input for fusing runs from differing TREC participants. Instead the ranks were used as inputs [19].

There are positional or majoritarian rank aggregation methods [23]. Positional assigns a score to an item depending on the rank of the item, while majoritarian uses a pair wise comparison of ranked items. The fusion methods applied in this research include one of each of these fusion methods as well as a voting system that does not differentiate between documents according to their rank. They are described in the following section.

III. METHODOLOGY

Three methods of system fusion were developed and tested to determine a method of improving the MAP results for various systems using existing results. A voting system where an occurrence of a document within the list counted as one vote is explained in subsection III-A. A second method (Inverse Rank Merge (IRM)) that takes into account the rank of the document within the list is described in subsection III-B, and a

⁴ TREC used single assessors in the assessment of the Blog06 document collection.

⁵ Opinion scores indicate how strongly a word expresses an opinion.

majoritarian method (V/IRM) for combining the votes and IRM method is detailed in subsection III-C. These three methods share the same limitation to different degrees (described in subsection III-D). The system fusion methods (described below) were applied using the 56 runs submitted by the 14 Blog06 participants as input. MAP was calculated for the three system fusion runs and is reported in the results section (IV).

The original MAP results detailed in Ounis et al. (2006) [6] report two measures: (1) documents relevant to the query topic, and (2) documents with an opinion on the query topic. The results reported in this research paper include the MAP results for documents containing an opinion on the given topic only, although similar results were observed when the fusion methods were applied to the documents relevant to the topic. The three system fusion methods (votes, IRM and V/IRM) are detailed below.

A. Votes

Table I demonstrates the votes method where the occurrence of a document in the run is a single vote for that document, irrespective of the rank within the original list. The top 10 documents in two hypothetical runs are shown with the rank for each document. The votes are summed for each document and sorted into ascending order and a rank is allocated accordingly. For example: d_8 is listed in both Run 1 and Run 2 – accumulating two votes, while d_4 is listed in Run 2 and not Run 1 – accumulating one vote.

B. Inverse Rank Method

Table II demonstrates the IRM method where points are allocated to each document in a ranked list depending on that document's rank. For example, in a list of 10 documents, the document ranked one is allocated 10 points and the document ranked 10 is allocated one point. This method is designed to take into account the rank allocated by the original systems. For example: d_8 is ranked 1 in Run 1 (10 points) and ranked 2 in Run 2 (nine points) – accumulating 19 points, while d_4 is ranked 8 in Run 2 (three points) and not ranked in the top 10 documents in Run 1 (0 points) – accumulating three points.

C. V/IRM

The V/IRM method combines the votes and IRM method results by taking the average rank of each document as the score for that document. Table III demonstrates the V/IRM system fusion method being applied to the two hypothetical runs used to demonstrate the previous two methods.

D. Limitation within the methods

The examples shown in this section are all limited by the small number of runs included (two). However, the small number of runs highlights a limitation that may occur using these methods where documents obtaining the same number of votes/IRM/scores record different ranks. This occurs less often in the IRM and V/IRM methods compared to the votes method.

TABLE II

UN EXAMPLE OF THE IRM METHOD						
Run 1		Run 2		IRM result		
Doc Num	Rank	Doc Num	Rank	Points	Doc Num	Rank
d_8	1	d_3	1	19	d_8	1
d_9	2	d_8	2	17	d_9	2
d_6	3	d_9	3	15	d_3	3
d_2	4	d_2	4	14	d_2	4
d_1	5	d_1	5	13	d_6	5
d_3	6	d_6	6	12	d_1	6
d_7	7	d_{12}	7	6	d_{12}	7
d_{10}	8	d_4	8	5	d_{10}	8
d_{12}	9	d_{10}	9	4	d_7	9
d_5	10	d_{11}	10	3	d_4	10
d_4		d_5		1	d_5	11
d_{11}		d_7		1	d_{11}	12

TABLE III
UN EXAMPLE OF THE V/IRM METHOD

Run 1		Run 2		V/IRM result		
Doc Num	Rank	Doc Num	Rank	Score	Doc Num	Rank
d_8	1	d_3	1	3.0	d_2	1
d_9	2	d_8	2	3.0	d_3	2
d_6	3	d_9	3	3.0	d_8	3
d_2	4	d_2	4	3.5	d_1	4
d_1	5	d_1	5	4.0	d_9	5
d_3	6	d_6	6	4.5	d_6	6
d_7	7	d_{12}	7	7.5	d_{10}	7
d_{10}	8	d_4	8	7.5	d_{12}	8
d_{12}	9	d_{10}	9	9.5	d_4	9
d_5	10	d_{11}	10	10.0	d_7	10
d_4		d_5		10.5	d_5	11
d_{11}		d_7		11.5	d_{11}	12

However, this limitation becomes more evident when combining individual participant's results where the number of runs included is low. This limitation has not been addressed within the scope of this paper, instead the ties in the output lists remained in the order in which they appeared (random). However, possible solutions are discussed in the future work section (V).

IV. RESULTS AND DISCUSSION

The system fusion methods described in section III were applied to the 56 runs submitted by the TREC participants and MAP values were calculated for the three newly created runs. All three methods resulted in improved MAP results compared to the best MAP result of any individual run, with the IRM method achieving the highest MAP result. Figure 1 illustrates the MAP value for the three system fusion runs and each individual run submitted.

TABLE IV
MAP FOR IRM SYSTEM FUSION METHOD OF INDIVIDUAL RUNS ACHIEVING HIGHER MAP

Number of individual runs included	MAP
All runs	0.2915
Top half	0.29722
Top third	0.2857
Top ten	0.2480

The best MAP achieved by the participants was 0.2052 (Indiana University run entitled woqln2), while the three system fusion methods all achieved higher MAP. The IRM method achieved the highest result (0.2915), the votes method achieved the lowest of the three methods (0.2590), while the V/IRM method was an approximate average of the two other methods (0.2794). As the IRM method achieved the highest MAP result and is least effected by the limitation outlined in subsection III-D, this method was applied to the remainder of the research reported in this paper.

To test the hypothesis that it is the diversity within the runs that achieves the higher MAP results, opposed to the runs achieving the highest MAP results, the runs achieving lower MAP results were removed from the input for the IRM method. Table IV details the MAP result after the IRM method was applied to the highest MAP achieving runs. The method achieved higher MAP when applied to the top half (28) of the runs. However, MAP was lower when the top third (18) and top 10 runs were used as input for the method, whilst maintaining a higher result than any individual run.

To further investigate the impact of diversity on the system fusion performance, the IRM method was applied to individual TREC participant's runs and MAP was calculated for the resulting run. Figure 2 illustrates the MAP achieved by the individual runs and the IRM run for each participant. Eight of the 14 participant's IRM method run improved on their best MAP result, five participant's IRM method run achieved a MAP result lower compared to their best individual run and the remaining participant did not record a MAP result on their runs, therefore the IRM method run did not record a MAP result.

Of the eight TREC participants whose IRM run achieved a higher MAP result, CMU (Carnegie Mellon University) [24] achieved the highest improvement. Table V lists the runs submitted by CMU and the CMU IRM run, detailing the description of each run as provided in Yang et al. (2006) [24]. CMU applied a mixture of inputs and models to create the runs submitted to TREC. Although not all TREC participants described their submitted runs, it seems that most participants applied a single model, therefore it may be expected that they produced less diversity in their runs [17], [25]-[29]:

- University of Arkansas at Little Rock [26] – five runs consisting of a base run, base run plus linkindegree re-ranking, spam detection, opinion re-ranking respectively and the final run is a combination of all of the above;

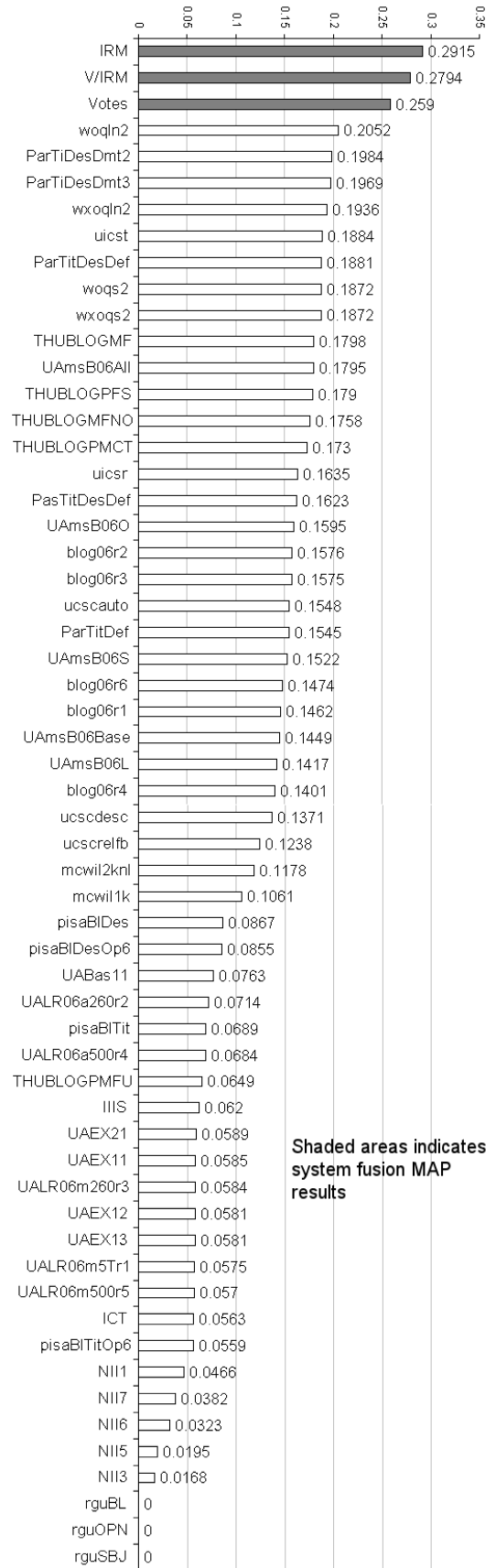


Fig. 1. MAP values for individual runs and result of system fusion methods

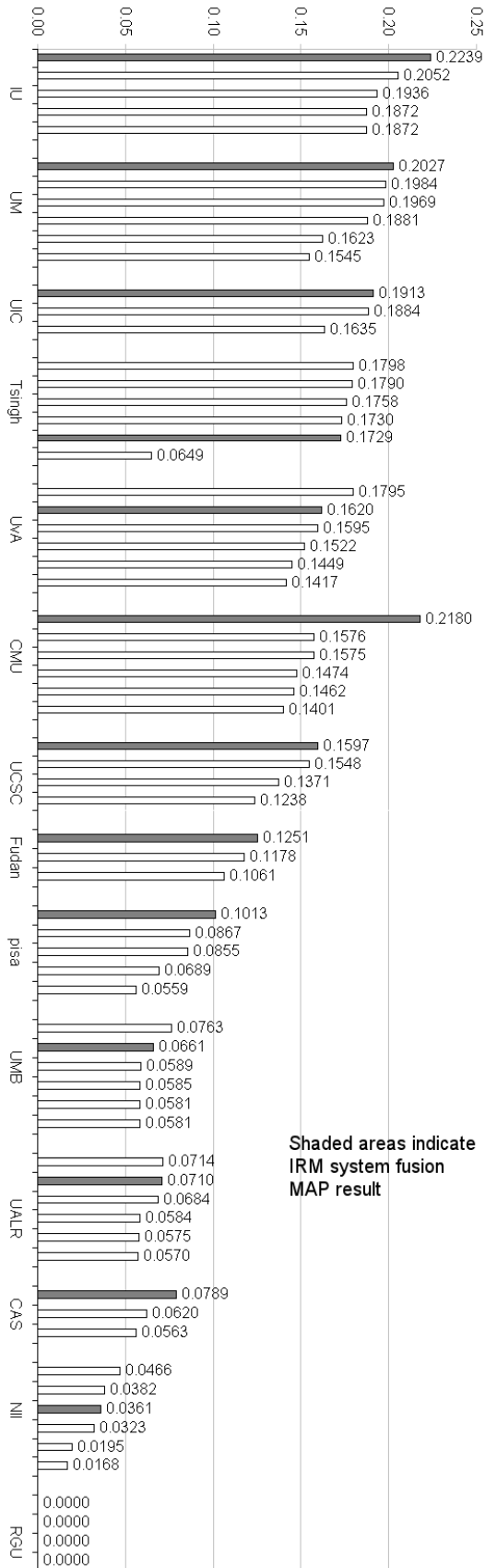


Fig. 2. MAP values for participants runs and result of IRM fusion system method

Name of run	MAP	Description [24]
IRM run	0.2180	
blog06r2	0.1576	Title query, bigram model
blog06r3	0.1474	Title query, trigram model
blog06r6	0.1474	Title + description + opinion words, trigram model
blog06r1	0.1462	Title query, unigram model
blog06r4	0.1401	Title + description + opinion words, unigram model

Note: All runs submitted by CMU listed 'adjective percentage' in their description

- University of Arkansas at Little Rock [26] – five runs consisting of a manual run using five features, automatic and manual runs using 260 and 500 features respectively;
- University of Pisa [27] – four runs with varying queries: title only, title + opinion terms, title + description and title + opinion terms + description;
- University of California, Santa Cruz [28] – three runs. The first two runs used the title and description as query terms respectively, and the third run used the title as the query terms and the description run as pseudo relevance feedback;
- National Institute of Informatics [17] – five runs consisting of a base run, a re-ranking of the base run using a topic-relevance model, a retrieval model and a likelihood probability of opinion terms respectively. The final run made use of the probability that adjectives and adverbs appear more often in opinion-bearing documents and combined this with a query-likelihood model;
- University of Maryland, College Park [29] – five runs consisting of two runs without opinion detection where paragraphs were extracted from the documents with the title and a combination of the title and description used as the query terms respectively. The third run without opinion detection used passages extracted from the documents and a combination of the title and description as the query term. The fourth and fifth runs extracted paragraphs from the documents, used a combination of the title and description as query terms and demoted the non-opinion-bearing paragraphs two times proportionally for the fourth run and three times proportionally for the fifth run.

The IRM run that achieved the lowest result compared to the participant's individual runs was Tsinghua University. The original IRM run achieved a MAP value of 0.1729. However when the individual run that achieved 0.0649 was removed from the input to the IRM run, the new MAP result was 0.1773. This indicates the runs achieving low MAP results will directly affect the IRM run's MAP result, re-enforcing the findings of Fuller et al. (1997) [21] who found that poor quality inputs can lower fusion performance.

V. CONCLUSIONS AND FUTURE WORK

This study shows that improved opinion detection can be achieved using system fusion and higher MAP results are achievable. It is preferable to include runs using diverse methods as input to gain the maximum improvement in MAP, while ensuring that poor quality runs are not included.

Two possible methods of addressing the limitation (described in subsection III-D) of the proposed methods are:

- Average rank – in the example detailed in table I eight documents scored two votes and were ranked 1-8, using average rank would assign all documents scoring two votes a rank of four and all documents scoring one vote would be assigned a rank of 10.5;

- Using scores of original runs instead of rank [20] – this solution may be applied to individual participant's runs to improve their overall MAP results. However, it cannot be applied directly to runs from a mixture of sources as the scoring system applied by each participant may not be consistent. Normalising the scores will be investigated in future research to determine whether scores can be used.

This research has shown that simply removing runs achieving lower MAP values may not be the best method of selecting which runs to include as input for a fusion method. Future research will investigate the best properties to use in selecting which runs should form the input for a fusion method.

REFERENCES

- [1] "Internet world stats - usage and population statistics," December 2007, available: <http://www.internetworldstats.com/stats.htm>.
- [2] "About technorati," December 2007, available: <http://www.technorati.com/about/>.
- [3] A. Rosenbloom, "Introduction: The blogosphere," *Commun. ACM*, vol. 47, no. 12, pp. 30-33, 2004.
- [4] A. Lenhart and S. Fox, "Bloggers: A portrait of the internet's new storytellers," Pew Internet & American Life Project, Tech. Rep., July 2006.
- [5] B. J. Jansen, A. Spink, and T. Saracevic, "Real life, real users, and real needs: a study and analysis of user queries on the web," *Information Processing and Management*, vol. 36, no. 2, pp. 207-227, 2000.
- [6] I. Ounis, M. de Rijke, C. Macdonald, G. Mishne, and I. Soboroff, "Overview of the trec-2006 blog track," in *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.
- [7] D. Osman, J. Yearwood, and P. Vamplew, "Automated opinion detection: Implications of the level of agreement between human raters," 2007, submitted to *Information Processing and Management* (2007).
- [8] D. Harman, "Overview of the first trec conference," in *SIGIR '93: Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 1993, pp. 36-47.
- [9] B. T. Bartell, G. W. Cottrell, and R. K. Belew, "Automatic combination of multiple ranked retrieval systems," in *SIGIR '94: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: Springer-Verlag New York, Inc., 1994, pp. 173-181.
- [10] E. A. Fox and J. A. Shaw, "Combination of multiple searches," in *The 3rd Text Retrieval Conference (TREC-3)*, Gaithersburg, Maryland, 1995, pp. 105-108.
- [11] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble selection from libraries of models," in *ICML '04: Proceedings of the twenty-first international conference on Machine learning*. New York, NY, USA: ACM Press, 2004, p. 18.
- [12] Y.-S. Dong and K.-S. Han, "Text classification based on data partitioning and parameter varying ensembles," in *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*. New York, NY, USA: ACM Press, 2005, pp. 1044-1048.
- [13] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences," in *Proceedings of the 2003 conference on Empirical methods in natural language processing*. Morristown, NJ, USA: Association for Computational Linguistics, 2003, pp. 129-136.
- [14] S.-M. Kim and E. Hovy, "Automatic detection of opinion bearing words and sentences," in *Natural Language Processing - IJCNLP 2005*. New York: Springer, 2005.
- [15] S. Kim and E. Hovy, "Identifying and analyzing judgment opinions," in *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*. New York City, USA: Association for Computational Linguistics, June 2006, pp. 200-207. [Online]. Available: <http://www.aclweb.org/anthology/N06/N06-1026>
- [16] A. Andreevskaia, S. Bergler, and M. Urseanu, "All blogs are not made equal: Exploring genre differences in sentiment tagging of blogs," in *Proceedings of the International Conference on Weblogs and Social Media (ICWSM-2007)*, 2007.
- [17] K. Eguchi and C. Shah, "Opinion retrieval experiments using generative models: Experiments for the trec 2006 blog track," in *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.
- [18] C. C. Vogt and G. W. Cottrell, "Predicting the performance of linearly combined ir systems," in *SIGIR '98: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 1998, pp. 190-196.
- [19] R. Wilkinson, "Using combination of evidence for term expansion," in *Proceedings of the 19th Annual BCS-IRSG Colloquium on IR Research*, J. Furner and D. Harper, Eds. Springer, April 1997.
- [20] K. Yang, N. Yu, A. Valerio, and H. Zhang, "Widit in trec-2006 blog track," in *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.
- [21] M. Fuller, M. Kaszkiel, C. L. Ng, P. Vines, R. Wilkinson, and J. Zobel, "MDS trec6 report," in *Text REtrieval Conference*, 1997, pp. 241-257. [Online]. Available: citeseer.ist.psu.edu/70787.html
- [22] A. Singhal, C. Buckley, and M. Mitra, "Pivoted document length normalization," in *SIGIR '96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 1996, pp. 21-29.
- [23] M. Farah and D. Vanderpooten, "An outranking approach for rank aggregation in information retrieval," in *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 2007, pp. 591-598.
- [24] H. Yang, L. Si, and J. Callan, "Knowledge transfer and opinion detection in the trec2006 blog track," in *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.
- [25] G. Mishne, "Multiple ranking strategies for opinion retrieval in blogs," in *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.
- [26] H. Joshi, C. Bayrak, and X. Xu, "Ualr at trec: Blog track," in *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.
- [27] G. Attardi and M. Simi, "Blog mining through opinionated words," in *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.
- [28] E. Zhang and Y. Zhang, "Ucsc on trec 2006 blog opinion mining," in *The Fifteenth Text REtrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.
- [29] D. Oard, T. Elsayed, J. Wang, and Y. Wu, "Trec-2006 at maryland: Blog, enterprise, legal and qa tracks," in *The Fifteenth Text Retrieval Conference Proceedings (TREC 2006)*, E. M. Voorhees and L. P. Buckland, Eds., Gaithersburg, Maryland, 2006.