Using Subjectivity Analysis to Improve Thread Retrieval in Online Forums*

Prakhar Biyani[1], Sumit Bhatia[2], Cornelia Caragea[3], and Prasenjit Mitra[4]

[1] Yahoo Labs, Sunnyvale, CA, USA
[2] IBM Almaden Research Center, San Jose, CA, USA
[3] Computer Science, University of North Texas, Denton, TX, USA
[4] Qatar Computing Research Institute, Doha, Qatar
pxb5080@yahoo-inc.com, sumit.bhatia@us.ibm.com,
ccaragea@unt.edu, pmitra@qf.org.qa

Abstract. Finding relevant threads in online forums is challenging for internet users due to a large number of threads discussing lexically similar topics but differing in the type of information they contain (e.g., opinions, facts, emotions). Search facilities need to take into account the match between users' *intent* and the type of information contained in threads in addition to the lexical match between user queries and threads. We use intent match by incorporating subjectivity match between user queries and threads into a state-of-the-art forum thread retrieval model. Experimental results show that subjectivity match improves retrieval performance by over 10% as measured by different metrics.

1 Introduction

Apart from asking questions and holding discussions, internet users search archives of online forums for threads discussing topics that are relevant to their information needs. Often, threads sharing common keywords discuss different topics and in such cases, finding relevant threads becomes challenging for users. Consider the following query issued by an internet user to a travel forum such as Trip Advisor–New York: *"best thanksgiving turkey"*. The query is subjective seeking opinions and viewpoints of different users on the quality of thanksgiving turkey served/found at different places in New York. A thread simply containing keywords *"thanksgiving"* and *"turkey"* and not having opinions of users would not satisfy the searcher. Similarly, for queries seeking factual information, threads having long discussions and opinions are likely to be not relevant. Hence, in addition to the *lexical* dimension (i.e., keyword match), search facilities in online forums need to take into account the ***intent*** dimension i.e., the type of information (e.g., opinions, facts) a searcher wants, to improve search. The current work addresses precisely this problem.

We improve an ad-hoc thread retrieval model for an online forum by combining lexical match between query and thread content with the match between searchers' intent and the type of information contained in threads. We focus on an important dimension of searchers' intent which is his preference for *subjective* and *non-subjective* information. Subjective information contains opinions, viewpoints, emotions and other private

* Work performed when Prakhar Biyani was at Pennsylvania State University.

A. Hanbury et al. (Eds.): ECIR 2015, LNCS 9022, pp. 495–500, 2015.
© Springer International Publishing Switzerland 2015

states whereas non-subjective information contains factual material [1]. Specifically, we identify the subjectivity of thread topics and user queries and incorporate the subjectivity match in a state-of-the-art retrieval model for online forums [2] to improve the retrieval performance. To predict thread subjectivity, we use our subjectivity classifier, specifically developed for online forum threads (Biyani et al. [3]). The classifier uses features derived from thread structure, sentiment and dialogue acts [4]. For determining query subjectivity, we use manual subjectivity tags (subjective/non-subjective).

2 Related Work

Subjectivity analysis has been actively researched in opinion mining [5], question-answering [6,7,8,9,10], and finding opinionated threads in online forums [11,12,13]. Stoyanov et al., [6] used subjectivity filter on answers, separating factual sentences from opinion sentences, to improve answering of opinion questions. Li et al., [8] used graphical models to rank answers based on their topical and sentiment relevance to opinion questions. Gurevych et.al., [7] used a rule-based lexicon based approach to classify questions as subjective or factoid. Moghaddam et al., [9] performed aspect-based question answering in product reviews and showed that taking into account the match between opinion polarities of questions and answers improved answer retrieval. Oh et al., [10] improved answering of non-factoid why-questions by using supervised classification for re-ranking answers based on their sentiment and other properties. All these previous works focused on improving question-answering of non-factoid (i.e., opinion) questions in product reviews and community QA sites. In contrast, the current work employs subjectivity analysis to improve an ad-hoc vertical retrieval model for an online forum. We show that using the subjectivity match, retrieval performance can be improved for both subjective and non-subjective queries.

3 Retrieval Model

Here, we discuss how information about subjectivity of threads can be utilized in thread retrieval systems. We use a state-of-the-art probabilistic model for forum thread retrieval [2] as a strong baseline (explained below) and incorporate subjectivity match between queries and threads in the model to see if it helps improve the retrieval performance.

3.1 Probabilistic Retrieval

Bhatia and Mitra [2] used a probabilistic model based on inference networks that utilizes the structural properties of forum threads. Given a query Q, the model computes $P(T|Q)$, the probability of thread T being relevant to Q, as follows:

$$P(T|Q) \stackrel{rank}{=} P(T) \prod_{i=1}^{n} \left\{ \sum_{j=1}^{m} \alpha_j P(Q_i|S_{j_T}) \right\} \tag{1}$$

where: $P(T)$ is the prior probability of a thread being relevant, Q_i is the i^{th} term in query Q, S_{jT} is the j^{th} structural unit in the thread T, α_j determines the weight given to component j and $\sum_{j=1}^{m} \alpha_j = 1$.

Note that the term $\prod_{i=1}^{n} \left\{ \sum_{j=1}^{m} \alpha_j P(Q_i|S_{j_T}) \right\}$ models lexical match between query and thread content. In order to estimate the likelihoods $P(Q_i|S_{j,T})$, we use the standard language modeling approach in information retrieval [14] with *Dirichlet Smoothing* as follows:

$$P(Q_i|S_{j_T}) = \frac{f_{Q_i,j_T} + \mu \frac{f_{Q_i,j_C}}{|j|}}{|j_T| + \mu} \tag{2}$$

Here,
f_{Q_i,j_T} = frequency of term Q_i in j^{th} structural component of thread T,
f_{Q_i,j_C} = frequency of term Q_i in j^{th} structural component of all the threads in the collection C.
$|j_T|$ is the length of j^{th} structural component of thread T,
$|j|$ is the total length of j^{th} structural component of all the threads in the collection C,
μ is the Dirichlet smoothing parameter.

In this work, we set μ to be equal to 2000, a value that has been found to perform well empirically [15]. Thus, the model computes the overall probability of a thread being relevant to the query by combining evidences from different structural units of the thread (title, initial post and reply posts).

3.2 Incorporating Subjectivity Information in the Retrieval Model

In absence of any information about thread's content, subjective threads are more likely to be relevant to subjective queries and vice versa for non-subjective threads. We conceptualize this idea by taking into account the match between subjectivities of threads and queries in addition to the lexical match between them. Specifically, we incorporate the subjectivity match using the term $P(T)$ (in Equation 1) which represents the prior probability of a thread being relevant to a query. We use the following two settings to incorporate subjectivity match between threads and queries into the retrieval model:

1. Subjectivity probability of a thread as its prior relevance probability: For subjective (or non-subjective) queries, a thread's prior probability of being relevant is taken to be its probability of being subjective (or non-subjective). More precisely, for a subjective query, Q_s, relevance score of a thread T is calculated as follows:

$$P(T|Q_s) \stackrel{rank}{=} P(Subj|T) \prod_{i=1}^{n} \left\{ \sum_{j=1}^{m} \alpha_j P(Q_{si}|S_{j_T}) \right\} \tag{3}$$

Here, $P(Subj|T)$ is the probability of thread T being subjective as outputted by the subjectivity classifier. Likewise, for a non-subjective query, the term $P(Subj|T)$ is replaced by $P(NSubj|T)$ which is the probability of thread T being non-subjective. For a thread T, $P(Subj|T) + P(NSubj|T) = 1$.
2. Re-ranking using subjectivity probabilities: A two-step ranking model is used. First, threads are ranked according to their lexical similarity with the query where $P(T)$

is taken as constant for all the threads and then re-ranking of threads (at various ranks) is performed based on their subjectivity probabilities. Basically, for a subjective query, re-ranking is sorting (in descending order) the ranked list of threads based on their subjectivity probabilities. Re-ranking for a non-subjective query is done similarly.

3.3 Getting Subjectivity Information for Threads and Queries

To obtain the subjectivity probability for a thread ($P(Subj|T)$), we used our subjectivity classifier developed previously (Biyani et al. [3]). We used the classifier to get confidence scores for all the threads (of belonging to the subjective class) and used the scores as the subjectivity probabilities. For determining query subjectivity, we took help of human annotators (discussed in Section 4.1).

4 Experiments and Results

4.1 Data Preparation

For our experiments, we used the dataset as used by Bhatia et al. [2]. It consists of threads crawled from a popular online forum: **Trip Advisor–New York** that contains travel related discussions mainly for New York city. It has 83072 crawled threads from the forum, a set of 25 queries and associated relevance judgments. For a query, the dataset has graded relevance judgments: 0 for totally irrelevant, 1 for partially relevant and 2 for highly relevant threads.

For annotating queries as subjective or non-subjective, we took help of three human annotators. First, two annotators tagged all the 25 queries with a percentage agreement and Kappa value of 88% and 0.743 respectively. The third annotator was then asked to disambiguate the tags of the queries on which the two annotators disagreed. Finally, we get 10 subjective and 15 non-subjective queries. Table 1 lists some of the subjective and non-subjective queries.

Table 1. Examples of subjective and non-subjective queries

Type	Example queries
Subjective	best mode of transportation from brooklyn to manhattan; safety in manhattan; best thanksgiving turkey; how safe is new york; how much to tip people
Non-subjective	new york to niagara falls; educational trips in new york; beaches in new york city; winter temperature in new york city; penn station to JFK

4.2 Experimental Setting

To conduct retrieval experiments, we used the Indri language modeling toolkit[1]. While indexing, stemming was performed using Porter's stemmerand stopwords were removed

[1] `http://lemurproject.org`

using a general stop word list of 429 words used in the Onix Test Retrieval Toolkit[2]. The queries and relevance judgments available with the dataset as discussed in Section 4.1 were used for retrieval experiments. For the baseline retrieval model, we used the optimal parameter settings as used in the original work [2]. In order to compare the performance of various retrieval models, we report precision, Normalized Discounted Cumulative Gain (NDCG) and Mean Average Precision (MAP) at ranks 5, 10 and 15.

Table 2. Retrieval results

Model	P@5	P@10	NDCG@5	NDCG@10	MAP@5	MAP@10
Subjective queries						
Baseline	0.56	0.52	0.7745	0.7180	0.7264	0.6662
Top 5 Re-rank	0.56	0.52	**0.8672**	0.7322	**0.8792**	**0.7355**
Top 10 Re-rank	0.58	0.52	0.7433	0.6964	0.7504	0.6873
Top 15 Re-rank	0.6	0.55	0.7370	0.7018	0.7361	0.6956
Subjectivity Prior Model	0.56	0.54	0.8010	**0.7433**	0.7880	0.6882
Non-subjective queries						
Baseline	0.546	0.546	0.6838	0.6988	0.7	0.651
Top 5 Re-rank	0.546	0.546	0.7056	0.7263	0.6688	0.6499
Top 10 Re-rank	0.56	0.546	0.8148	0.7644	0.7475	**0.7078**
Top 15 Re-rank	0.546	0.533	**0.8220**	**0.7658**	**0.7938**	0.6761
Subjectivity Prior Model	0.546	0.546	0.7827	0.7597	0.7518	0.7045
Average						
Baseline	0.552	0.536	0.7201	0.7065	0.7105	0.6572
Top 5 Re-rank	0.552	0.536	0.7703	0.7286	0.7530	0.6842
Top 10 Re-rank	0.568	0.536	0.7862	0.7372	0.7486	**0.6996**
Top 15 Re-rank	0.568	0.54	0.7880	0.7402	**0.7707**	0.6840
Subjectivity Prior Model	0.552	0.544	**0.7900**	**0.7532**	0.7663	0.6980

4.3 Results

Table 2 presents retrieval results for subjective and non-subjective queries, and the overall average result. **Subjectivity Prior Model** denotes the setting where thread's subjectivity probability is used as its prior relevance probability (as explained in Section 3.2). We see that using subjectivity information of threads improves MAP and NDCG values for both subjective and non-subjective queries against the baseline model. We also note that precision values remain almost unchanged (across all the settings). This is an interesting observation as it suggests that subjectivity match does not help much in finding more relevant threads. Instead, it improves ranking of threads by changing relative ordering of ranked threads. MAP takes into account ordering of ranked results and NDCG takes into account ordering and graded relevance (0, 1, 2) of the ranked results. For the re-ranking setting, we see that re-ranking at rank 5 outperforms re-ranking at ranks 10

[2] `http://www.lextek.com/manuals/onix/stopwords1.html`

and 15 for subjective queries. In contrast, for non-subjective queries, re-ranking at rank 15 outperforms the other two re-ranking settings.

5 Conclusion and Future Work

We combined the two dimensions of *lexical similarity* and *intent match* in a forum thread retrieval model and showed that the combination performs better than the model based only on lexical similarity. In future, we plan to explore automatic subjectivity classification of user queries, investigate other dimensions of user intent, and build fully automated thread retrieval systems.

References

1. Bruce, R., Wiebe, J.: Recognizing subjectivity: A case study in manual tagging. Natural Language Engineering 5(2), 187–205 (1999)
2. Bhatia, S., Mitra, P.: Adopting inference networks for online thread retrieval. In: AAAI, pp. 1300–1305 (2010)
3. Biyani, P., Bhatia, S., Caragea, C., Mitra, P.: Thread specific features are helpful for identifying subjectivity orientation of online forum threads. In: COLING, pp. 295–310 (2012)
4. Bhatia, S., Biyani, P., Mitra, P.: Classifying user messages for managing web forum data. In: Proceedings of the 15th International Workshop on the Web and Databases, pp. 13–18 (2012)
5. Liu, B.: Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, 2nd edn., pp. 627–666 (2010)
6. Stoyanov, V., Cardie, C., Wiebe, J.: Multi-perspective question answering using the opqa corpus. In: EMNLP-HLT, pp. 923–930 (2005)
7. Gurevych, I., Bernhard, D., Ignatova, K., Toprak, C.: Educational question answering based on social media content. In: AIE, pp. 133–140 (2009)
8. Li, F., Tang, Y., Huang, M., Zhu, X.: Answering opinion questions with random walks on graphs. In: ACL, pp. 737–745 (2009)
9. Moghaddam, S., Ester, M.: Aqa: Aspect-based opinion question answering. In: ICDMW, pp. 89–96. IEEE (2011)
10. Oh, J.H., Torisawa, K., Hashimoto, C., Kawada, T., De Saeger, S., Kazama, J., Wang, Y.: Why question answering using sentiment analysis and word classes. In: EMNLP-CoNLL, pp. 368–378 (2012)
11. Biyani, P., Caragea, C., Singh, A., Mitra, P.: I want what i need!: Analyzing subjectivity of online forum threads. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 2495–2498 (2012)
12. Biyani, P., Bhatia, S., Caragea, C., Mitra, P.: Using non-lexical features for identifying factual and opinionative threads in online forums. Knowledge-Based Systems 69(0), 170–178 (2014)
13. Biyani, P., Caragea, C., Mitra, P.: Predicting subjectivity orientation of online forum threads. In: Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics, pp. 109–120 (2013)
14. Ponte, J.M., Croft, W.B.: A language modeling approach to information retrieval. In: SIGIR, pp. 275–281 (1998)
15. Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to ad hoc information retrieval. In: SIGIR, pp. 334–342 (2001)