Using automatic keyword extraction to detect off-topic posts in online discussion boards

Nayer Wanas
Cairo Microsoft Innovation Lab
306 Corniche El-Nile
Maadi, Cairo, Egypt
nayerw@microsoft.com

Amr Magdy
Computer and Systems Engineering
Alexandria University
Alexandria, Egypt
amr.magdy@alex.edu.eg

Heba Ashour
Cairo Microsoft Innovation Lab
306 Corniche El-Nile
Maadi, Cairo, Egypt
hebaa@microsoft.com

ABSTRACT

Online discussions boards represent a rich repository of knowledge organized in a collection of user generated content. These conversational cyberspaces allow users to express opinions, ideas and pose questions and answers without imposing strict limitations about the content. This freedom, in turn, creates an environment in which discussions are not bounded and often stray from the initial topic being discussed. In this paper we focus on approaches to assess the relevance of posts to a thread and detecting when discussions have been steered off-topic.

A set of metrics estimating the level of novelty in online discussion posts are presented. These metrics are based on topical estimation and contextual similarity between posts within a given thread. The metrics are aggregated to rank posts based on the degree of relevance they maintain. The aggregation scheme is data-dependent and is normalized relative to the post length.

Categories and Subject Descriptors

H.3.1 [**Information storage and retrieval**]: Content Analysis and Indexing – *abstracting methods, indexing methods*

H.3.3 [**Information storage and retrieval**]: Information Search and Retrieval - i*nformation filtering*

I.5.2 [**Pattern Recognition**]: Design Methodology – *feature evaluation and selection*

General Terms

Algorithms, Experimentation.

Keywords

Content Filtering, Online Discussion Forums, Novelty Detection.

1. INTRODUCTION

Online discussion boards, also known as newsgroups or online forums, have been popular since the early days of the internet. Discussion board users share opinions, experiences, pose questions and search for answers. Forums are considered a rich repository of user generated content that contain vast resources of knowledge. Discussion boards are, however, typically marred with several problems, similar to other forms of user generated content. Amongst the major problems is the limited ability to filter and search the content to meet a specific need. This is due to the nature of these tools, which grant users significant liberties in terms of what content to present, where to place it, and when to post it. In addition, online discussion boards differ from other web-based information resources in that they are organized in tree structures known as threads. The lead post within this thread, called the thread head, initiates the discussion. Subsequent posts present additional content that extends the discussion. This, in turn, implies that knowledge within forums is retained in a sequence of posts within a thread, rather than a specific post. Irrelevant posts that infiltrate the sequence could obscure the ability to isolate nuggets of knowledge.

In order to overcome this problem, it is necessary to detect relevant posts within threads. This translates to the ability to detect which element of the thread is on-topic and which is off-topic. This is a challenging problem, due to the short and fragmented nature of the content which, therefore, allows for only minimal detection of context. In addition, users often take significant liberties in their use of language on forums and conform to the community's terminologies and styles, rendering traditional NLP tools less effective.

In this work we present a novel approach to automatically identify whether a given post P_j is on or off-topic in reference to a thread T_i. This approach is founded in the ability to automatically extract keywords that represent a particular thread, T_i, and distinguish it from all other threads. The extent to which a given post, P_j, contributes to this representative list is used to determine if the post is on or off-topic. The rationale behind this approach is that posts belonging to the same topic would share the same keywords and the assessment of a given post's contribution to the keywords of the thread would indicate its relevance.

2. NOVELITY DETECTION

The analysis of online discussion boards has attracted recent interest focusing on several aspects, including the analysis of structure[4], the examination of social roles and users[1] and computational linguistics. In addition, the automatic assessment of posts has been addressed through a variety of algorithms[2][11]. However, the detection of novelty within the sequence of posts has not been addressed.

Novelty detection in documents has, however, attracted recent interest in the information retrieval community. Fueled by the TREC novelty tracks[5][9][8], the main goal of this work is to allow users to easily access new information from a ranked list. The user query is modeled as the topic and the documents are evaluated based on the novelty they present. In this context, novelty detection event level novelty implies identifying documents that are relevant to the topic and discuss a new, related topic(s)[7]. This is in contrast to novelty at the sentence level, which mandates the provision of new information about the event within the document. In the context of novelty detection in online discussion forums, event level novelty is more appropriate.

CAW2.0 2009, April 21, 2009, Madrid, Spain.

Current work on event level novelty detection is based on event modeling using tools such as vector space models, language models or lexical chains[6]. Clustering is then performed to group similar documents together. New documents are then assigned to the appropriate cluster using a similarity score. When the similarity score is less than an assigned threshold, a new cluster is formed, representing a new event. In addition, temporal proximity has been accounted for in detecting event novelty. The aforementioned approaches assume that a query presents the initial topic and that proper linguistic rules are observed. In addition, larger documents provide enough context to detect similarity amongst them. However, this is not the case with posts in discussion boards. While the thread head dictates the topic of a given thread, the topic is yet to be inferred from the post content. In addition, posts in online discussion boards need not adhere to proper linguistic rules. In addition, concepts being presented are fragmented across several posts.

Our research focuses on detecting novelty in online discussion posts through initially estimating the topic of a given thread. This is achieved while avoiding the use of linguistic features which are replaced by an estimation of the ordered rank of words within the thread. To help establish the relevance of a given post, the structural dependencies between the different posts are also accounted for.

3. ESTIMATING NOVELTY IN ONLINE DICUSSION BOARDS

In order to determine if a given post P_j is on-topic within a given thread, we first need to estimate the topic of the thread. Keywords that distinguish a thread T_i are used as a representation of the topic of T_i. Keywords representing the contribution of each post towards each topic are estimated. The degree of similarity between these two vectors is an indication of the extent to which the post conformed to the thread topic. In addition, the relationship between a given post and the sequence of preceding posts, especially the lead post and the post immediately preceding, are factored into the estimation of the level of similarity between the post and the thread topic.

3.1 Keyword Extraction and Topical Estimation

All the words, less stop words, used in all the posts within each thread are represented as a vector, $W_T(T_i)$. While the word order may be important, the short and less structured nature of the content collectively reduces its impact and makes a bag-of-words approach sufficient. The elements of $W_T(T_i)$ are ordered using term frequency, *tf*, either directly or normalized by the informativeness level of each word, *w*, based on the Binomial Log-likelihood Ratio Test (BLRT) suggested in[10]. The BLRT score is calculated as

$$2\log\frac{L(p_1,k_1,n_1)L(p_2,k_2,n_2)}{L(p,k_1,n_1)L(p,k_2,n_2)}$$

Where $p_i = k_i/n_i$, $p = (k_1+k_2)/(n_1+n_2)$ and

$$L(p,k,n) = p^k(1-p)^{(n-k)}$$

The BLRT is used to test whether a given word has the same distribution in a foreground and background corpus. In order to determine the keywords of a thread, the collection of posts within the thread are considered the foreground and the collection of all other threads are considered the background. Hence, the value of k_1 is the term frequency of the word within the thread and n_1 is the total number of words within the thread. On the other hand, k_2 is the term frequency of the word in all other threads and n_2 is the total number of words in that collection.

The ordered words in the vector $W_T(T_i)$ are considered a topical representation of thread T_i. Similarly, we could consider the post P_j in thread T_i and the aggregation of all other posts within the thread as the foreground and background distributions, respectively. Hence, by using the BLRT score on these two distributions, we can estimate the words representing vector $W_P(T_i,P_j)$ for each post. In turn, the top words in vector $W_P(T_i,P_j)$ are considered the topic of posting P_j within thread T_i.

The cosine similarity between the vectors $W_T(T_i)$ and $W_P(T_i,P_j)$, $Sim(W_T(T_i),W_P(T_i,P_j))$ is considered a measure of how relevant posting P_j is to thread T_i.

3.2 Overlap Level

Posts that remain on-topic will continue to be relevant to the posts preceding them. A measure of the degree of similarity between the posts would give an indication of how relevant a post remains. The more similar posts are to each other, the greater the probability that they are discussing the same topic. This similarity is captured through measuring three values (i) *OverlapPrevious*, (ii) *OverlapHead* and (iii) *OverlapAll*.

3.2.1 OverlapPrevious

This feature measures the maximum degree of overlap between the terms used in posting P_j and posting P_{j-1}. This is achieved through word count normalized by the length of posting P_j,

$$OverlapPrevious(P_j) = \frac{count(\{P_{j-1}\} \in \{P_j\})}{|P_j|}$$

3.2.2 OverlapHead

The first post in the thread remains the most influential element that affects the topic of the discussion. This influence is signified by the need to maintain topical relevance in all posts that follow. To capture this importance, a degree of overlap between the terms used in a posting P_j and posting P_0 is calculated through word count,

$$OverlapHead(P_j) = \frac{count(P_0 \in P_j)}{|P_j|}$$

3.2.3 OverlapAll

An indicator of how relevant a given post, *Pj*, is to all previous posts within the same thread T_i is assessed by estimating the overlap will all posts $P_0 \ldots P_{j-1}$. This is estimated as

$$OverlapAll(P_j) = \frac{count(P_l \in P_j)}{|P_j| * (j-1)} \; \forall\, l < j$$

4. DATA

In this work, we are considering four different data sets representing a variety of online discussions. These data are (i) Slashdot[1], (ii) Myspace[2], (iii) Ciao[3], and (iv) Twitter[4].

[1] http://www.slashdot.org
[2] http://www.myspace.com
[3] http://www.ciao.com
[4] http://www.twitter.com

4.1 Slashdot

Slashdot is a popular technology discussion forum that also integrates an elaborate moderation scheme While Lampe and Resnick[6] suggested a post rating scheme that has shown to be sound, a good portion of the threaded discussion could however pass before users identify the value of its posts. Additionally, latter posts are often overlooked by moderators. Wrongly rated posts were usually not reversed, along with the fact that the quality of the rating is greatly affected by the value of the initial post. Collectively, these factors play a role in the amount of knowledge being surfaced in online discussion forums. The dataset used contains over 140,000 posts in 496 threads from the politics sub-forum. Posts in this section usually revolve around an initial post or contribution, and are generally lengthy, providing a significant amount of content for analysis.

4.2 Myspace

MySpace is a popular community based website that also includes a discussion forum. Similar to Slashdot, Myspace forums revolve around a set of user-defined topics. Moderation on Myspace is based on users flagging inappropriate content. A dataset composed of 380,000 posts from 16346 threads, covering topics of campus life, news and politics, and movies, are selected. Posts are a mixture of long contextual posts and short chat-like posts.

4.3 Ciao

Ciao defines itself as "a multi-million-strong online community that critically reviews and rates millions of products and services for the benefit of other consumers". This, in turn, results in a sequence of posts on a specific product that are independent from one another, rather than a discussion. 20,000 opinions from users reviewing movies are selected. Most opinions are lengthy and contain contextual information regarding the movie being rated.

4.4 Twitter

Twitter is a service that allows users to connect with other individuals. The main thrust behind Twitter is the exchange of information around the question "What are you doing?". The training set is composed from 900.000 posts from about 27000 users. While the messages are short, bounded by a 140 character limit, the social aspect associated with twitter distinguishes it from other online forums. 'Tweets' can be exchanged with a specific user, allowing for the evolution of a discussion around a topic. Posts can therefore be broken down into two sets, (i) threads between two users, and (ii) threads between a specific user and all. This dataset is prepared accordingly and relevance is assessed relative to each set.

5. RANKING RELEVANT POSTS IN ONLINE DICUSSION BOARDS

To rank posts based on the degree of relevance, the keyword similarity and overlap metrics are combined. However, there are two factors to consider, namely the length of the post and the nature of the data.

5.1 Data Dependence

Individual posts in different discussion boards are diverse in nature. While posts in discussion boards such as Slashdot and Myspace are more conversational, posts in Ciao are more independent. Collectively, the posts in these forums are generally lengthy, allowing for the derivation of context. Twitter threads, on the other hand, are made up of short, fragmented messages. This diversity presents a challenge and reflects on the weight each measure plays in ranking posts. Posts with more context allow for better topical estimation, while shorter, fragmented posts rely heavily on the overlap with previous posts.

In the absence of labeled data for training, the suggested scheme for combining the aforementioned metrics is as follows

$$\begin{aligned} relevance(P_j) = {} & w_1 * Sim\big(WT(Ti), WP(Ti, Pj)\big) \\ & * OverlapAll\big(P_j\big) + w_2 \\ & * OverlapPrevious\big(P_j\big) + w_3 \\ & * OverlapHead\big(P_j\big) \end{aligned}$$

In the case of Slashdot, Myspace and Ciao, the values of w_1, w_2 and w_3 are selected as 1, 0, and 0 respectively. While in the Twitter dataset, w_1, w_2 and w_3 are selected as 0.5,1, and 1 respectively.

While the similarity $Sim(W_T(T_i),W_P(T_i,P_j))$ ranges from -1 to 1, the values of overlap are positive numbers ranging from 0 to 1. The normalized relevance is therefore calculated as

$$\begin{aligned} & \overline{relevance\big(P_j\big)} \\ & = \frac{w_1 * \big(Sim\big(WT(Ti), WP(Ti, Pj)\big) * OverlapAll\big(P_j\big) + 1\big)}{2 * (w_1 + w_2 + w_3)} \\ & + \frac{w_2 * OverlapPrevious\big(P_j\big) + w_3 * OverlapHead\big(P_j\big)}{w_1 + w_2 + w_3} \end{aligned}$$

5.2 Lengthiness

The length of a post affects all the measures that are calculated. The longer the post, the more likely it is to contain words similar to other posts and would therefore contribute more to the topical estimation of the thread. This might not be coupled with its relevance to the thread. To overcome this aspect, the metrics are normalized based on the lengthiness of the post. The lengthiness of a given post is estimated as

$$Lengthiness\big(P_j\big) = \frac{|P_j|}{Average\ length\ of\ postings\ in\ thread}$$

Posts shorter or longer than the thread average are normalized by multiplying or dividing the combined measure by the lengthiness, respectively. This is reflected as

$$\begin{aligned} & \overline{relevance\big(P_j\big)} \\ & = \begin{cases} \dfrac{\overline{relevance\big(P_j\big)}}{lengthiness\big(P_j\big)} & \text{if } lengthiness > 1 \\ \overline{relevance\big(P_j\big)} * lengthiness\big(P_j\big) & \text{if } lengthiness \leq 1 \end{cases} \end{aligned}$$

In turn the degree of a post P_j is off-topic is estimated as

$$off - topic(P_j) = 1 - \overline{relevance(P_j)}$$

6. RESULTS

The goal of this work is to provide an indicator of the level by which a given post P_j is off-topic. A set of metrics that aim to estimate the topical relevance of a given post within a thread are suggested. The topical estimation is based on selecting a set of words to represent each thread and individual posts.

6.1 Keyword Representation

The similarity between vectors $W_T(T_i)$ and $W_P(T_i,P_j)$ is used to estimate the relevance of a given post, P_j, to the thread T_i. However, the elements within each vector could be represented with three different values. Words could be represented by (i) a binary value representing their existence or absence in the vector,

(ii) the *tf* of each word in the thread and post respectively, or (ii) the product *tf*BLRT-score* for each word. In addition, pruning the length of the vector would make the algorithm more computationally efficient. In this work we have pruned the ordered word vector to 5%, 10%, 20% and 50% of its size.

6.2 Evaluation

Slashdot, in contrast to the other datasets considered, contains posts that have been manually moderated with labels. In all other datasets, and in the absence of labeled data, the evaluation of the algorithms suggested is based on the level of separation between on-topic and off-topic posts. This separation is assessed based on how they conform to the general behavior of the forum.

6.2.1 Slashdot Data

Almost 32,000 posts in the training dataset have labels, of which just over 1000 posts are labeled as off-topic. In addition to being a minute set within the data, there is no guarantee that these labels are unbiased and represent the community's estimation of this post. This is due to the fact that there is no guarantee that several moderators have rated the post, and that this is consistent across all rated posts. Nonetheless they are used for evaluation. The set of 32,000 posts are collected and recall, F1 and accuracy measures are compared for this set. Five cross-validation sets of balanced data are used to evaluate the performance. Only posts labeled off-topic are considered, while all other labels are assumed to be relevant to the thread. Each set contains 800 posts labeled as off topic in addition to 800 posts labeled otherwise.

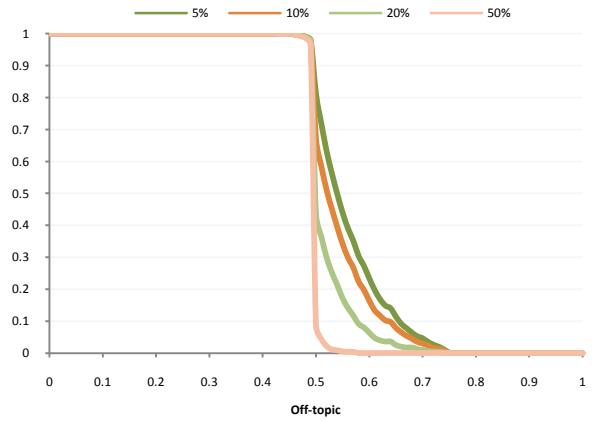

Figure 1: Recall for different keyword vector length

Given the dubious nature of the data labels, recall will be valued more than precision. Analysis of the recall indicates that the performance decreases with the reduction of the number of keywords selected. Figure 1 illustrates the recall on the training set, and indicates the best performance using 5% of the words as keywords. A similar behavior is demonstrated on the F1 measure (Figure 2).

Due to the experimental setup, the accuracy settles at 0.5 on both ends of the range of values of off-topic. The accuracy (Figure 3) generally improves with an increase in the number of keywords used, with the peak being at an off-topic degree of 0.5. The accuracy using *tf*BLRT-score* reduces when more than 20% of the words are selected as keywords, as indicated in Figure 3. It is worth mentioning that the best overall performance over the five cross validation sets is 66.7% using 50% of the words and a vector representation using the *tf*. However, this performance is marred with a high standard deviation. This deviation is reduced using the *tf*BLRT-score*. In addition, it is important to note that the computational and storage requirements dramatically increase with longer vectors. It is worth mentioning that the performance difference based on the three vector representations suggested (binary, *tf*, *tf*BLRT-score*) is limited, and generally insignificant.

The performance is significantly poorer using the cosine similarity alone compared to the normalization using the overlap and length. This is illustrated using the F1 measure in Figure 4. Overall, the off-topic metric using the *tf*BLRT-score* with 5% of the words selected as keywords gives the best performance.

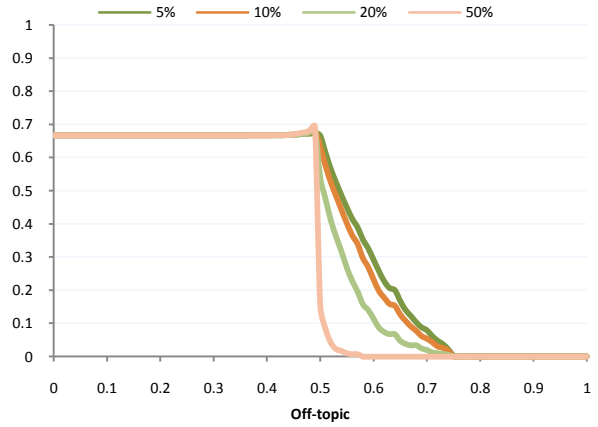

Figure 2: F1 measure for different keyword length

Figure 5 presents the histogram of off-topic values on Slashdot training data. This distribution demonstrates the wide range of off-topic values of this dataset. Using the test set, almost two thirds of the posts are considered off-topic with varying degrees. With the increase in the number of keywords, the range of values decreases and more posts are assessed with a 0.5 off-topic degree.

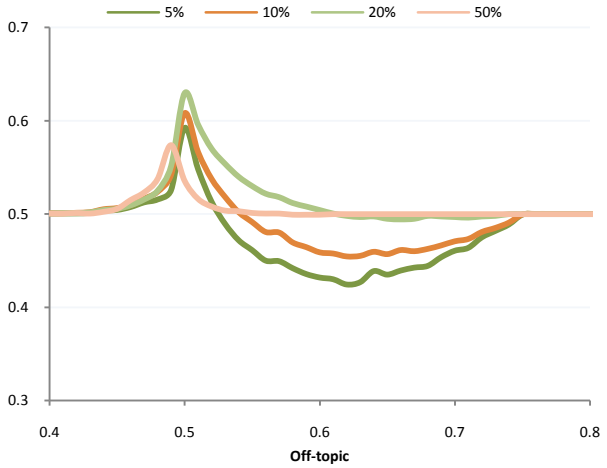

Figure 3: Accuracy with different keyword vector length

6.2.2 MySpace

The evaluation of Myspace data is based on the histogram of the off-topic degree (Figure 7). The histogram illustrates a larger segment of off-topic posts, which is more fitting to the nature of the dataset. Over 28,000 posts, from a total of 33,000+ posts are assessed as off-topic using a threshold of 0.5. Similar to Slashdot, the increase in the number of keywords used implies a limited

range of off-topic values, and hence the distinguishing power of the algorithm reduces.

6.2.3 Ciao

The histogram of the Ciao test data indicates a limited number of off-topic posts. Less that 3000 posts, representing 10% of the total posts, are assessed as being off-topic using a threshold of 0.5. This fits well with the fact that posts are independent reviews of products that rarely veer off-topic.

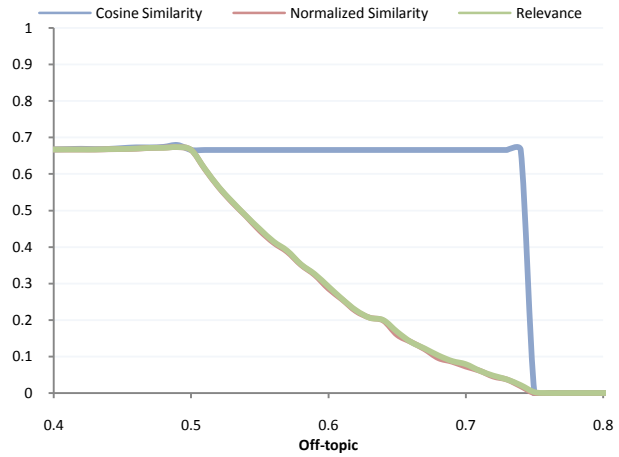

Figure 4: F1 measure using different metrics

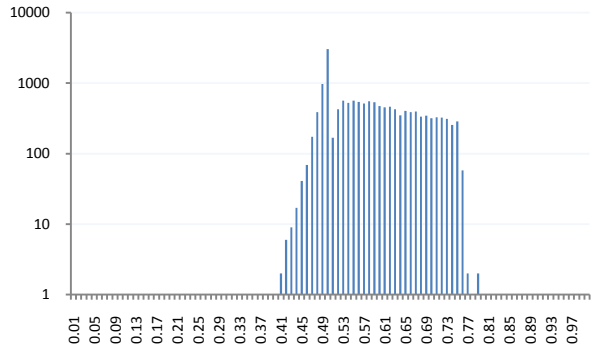

Figure 5: Histogram of off-topic values using *tf*BLRT-score* on 5% keywords on Slashdot test set

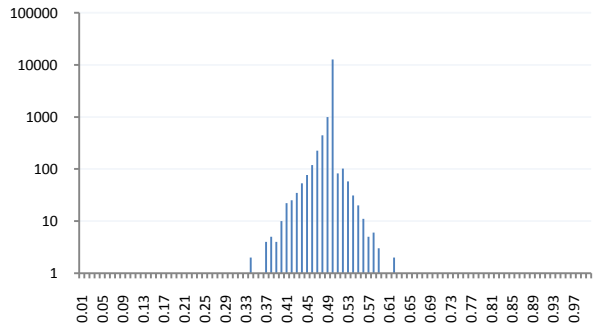

Figure 6: Histogram of off-topic values using *tf*BLRT-score* on 50% keywords on Slashdot test set

6.2.4 Twitter

The Twitter discussion data is extracted from the original data to represent exchange of tweets between two users. The collection posts for each user pairs are considered a thread, with the individual tweets representing the posts. In total, 69,000+ posts, from the original set that contains 169,000+ tweets, organized in 36,000+ threads are selected. Based on this set, the histogram of the number of postings with the same off-topic degree is highly skewed towards the larger values as illustrated in Figure 9. This is befitting the short and un-related nature of the data, which is also illustrated in the large range of off-topic values when using 50% of the words as keywords (Figure 10).

7. Data Challenge

Four approaches have been selected to submit to the data challenge on testing data provided from Slashdot. These approaches are:

Approach 1: A binary decision is based on a threshold of 0.5 on the off-topic degree of posts based on the *tf*BLRT-score* representation. The top 5% keywords are selected.

Approach 2: The top 5% keywords are selected, and the off-topic degree is evaluated using the *tf*BLRT-score*.

Approach 3: The top 50% keywords are selected and the off-topic degree is based on the assessment of the *tf*.

Approach 4: The top 20% keywords are selected to assess off-topic using the *tf*BLRT-score*.

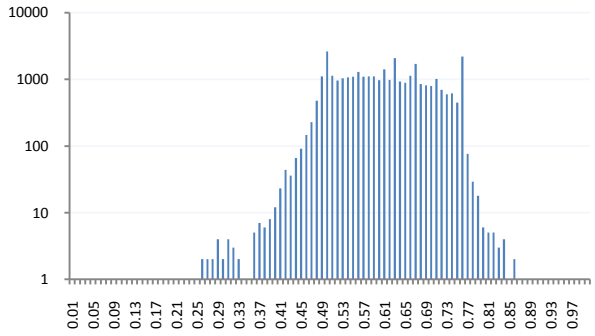

Figure 7: Histogram of off-topic values using tf*BLRT-score and 5% keywords on Myspace test data

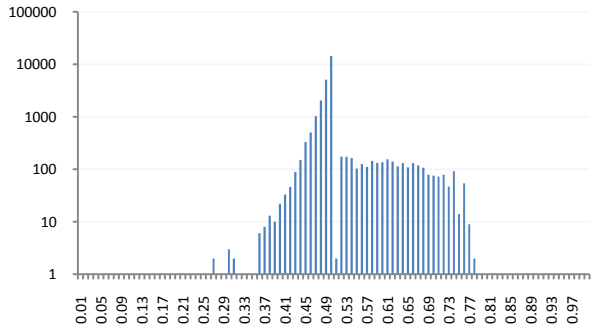

Figure 8: Histogram of off-topic values using tf*BLRT-score and 5% keywords on Ciao test data

8. DISCUSSION AND FUTUREWORK

In this work we present a novel approach to detecting off-topic posts in online discussions. The approach is based on two metrics to detect topical relevance. Topical estimation is based on selecting a set of keywords to represent each thread and individual posts. The keywords are selected using the BLRT informativeness measure, where the words are represented using three different measures, namely (i) a binary value, (ii) *tf*, (iii) *tf*BLRT-score*. The vector is pruned to different lengths of 50%, 20%, 10% and 5%. A measure of the degree of relevance of a given post to the lead post, the preceding post and the collection of all preceding posts is computed. An aggregation of these metrics is normalized based on the post length to provide an off-topic degree for each post.

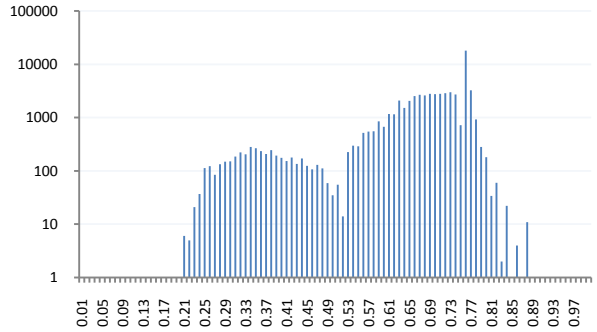

Figure 9: Histogram of off-topic values using *tf*BLRT-score* and 5% keywords on twitter discussions test data

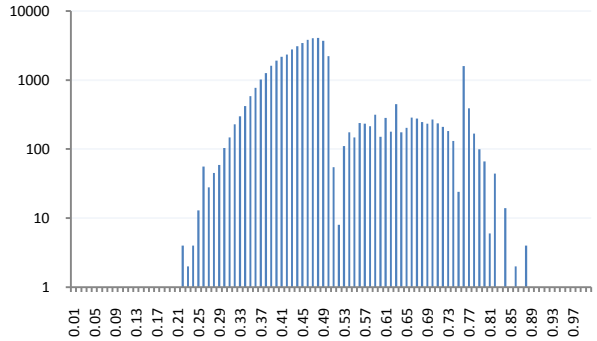

Figure 10: Histogram of off-topic values using *tf*BLRT-score* and 50% keywords on twitter discussions test data

Experimental results on Slashdot data indicate that the vector pruning reduces the processing time for the algorithm at the expense of a small decrease in accuracy. In addition, using the *tf*BLRT-score* to represent the best performance overall considering recall and F1-measure. A threshold value of 0.5 produces an optimal performance for computing a binary decision on off-topic posts.

There are many ways in which this work could be expanded

1. The topical representation of each thread is affected by the individual posts, including those that are off-topic. It would be interesting to implement this algorithm in an iterative fashion, where off-topic candidates are removed and the algorithm is re-applied on the remaining posts. While this approach would be computationally expensive, it is expected to return better estimates.
2. In the presence of labeled data, the aggregation model could be trained using a classifier to produce a more effective scheme.
3. This approach could be used to model users on online discussion forums through detecting their topical affinity.

REFERENCES

[1] Fisher D., Smith, M., and Welser, H., 2006, You Are Who You Talk To: Detecting Roles in Usenet Newsgroups, 39th Hawaii International Conference on System Sciences HICSS-39. - Kauai, HI, USA, IEEE Press, New Jersey, NJ,, January 4-7, 2006. - p. 59b.

[2] Fortuna B., Rodrigues, E., and Milic-Frayling, N., 2007, Improving the Classification of Newsgroup Messages through Social Network Analysis, Conference on Information and Knowledge Management CIKM'07. - Lisboan, Portugal, ACM Press, New York, NY, November 6-8, 2007. - pp. 585-588.

[3] Glance N., Hurst, M., Nigam, K., Siegler, M., Stockton, R., and Tomokiyo, T., 2005, Deriving Marketing Intelligence from Online Discussion, SIGKDD International Conference on Knowledge Discovery and Data Mining KDD'05. - Chicargo, IL, USA, ACM Press, New York, NY, August 21-24, 2005. - pp. 419-428.

[4] Gómez V., Kaltenbrunner, A., and López, V., 2008, Statistical Analysis of the Social Network and Discussion Threads in Slashdot, 17th International World Wide Web Conference WWW2008. - Beijing, China, ACM Press, New York, NY, April 21-25, 2008. - pp. 645-654.

[5] Harman D., 2002, Overview of the TREC 2002 novelty track, NIST Special Publication: SP 500-251, the eleventh text retrieval conference TREC2002. - 2002.

[6] Lampe C. and Resnick, P., 2004, Slash(dot) and Burn: Distributed Moderation in a Large Online Conversation Space, SIGCHI Conference on Human Factors in Computing Systems CHI'04, ed. Debray S., and Peterson, L.. - Vienna, Austria, ACM Press, New York, NY, April 24-29, 2004. - pp. 543-550.

[7] Li X., and Croft, W., 2008, An information-pattern-based approach to novelty detection, Information Processing and Management. - 2008.

[8] Soboroff I., 2004, Overview of the TREC 2004 novelty track, NIST Special Publication: SP 500-261, the thirteenth text retrieval conference TREC 2004. - 2004.

[9] Soboroff I., & Harman, D., 2003, Overview of the TREC 2003 novelty track, NIST Special Publication: SP 500-255, the twelfth text retrieval conference TREC 2003. - 2003.

[10] Tomokiyo T., and Hurst, M., 2003, A Language Model Approach to Keyphrase Extraction, Workshop On Multiword Expressions: Analysis Acquisition And Treatment. - 2003.

[11] Weimer M., Gurevych, I., and Mühlhäuser, M., 2007, Automatically Assessing the Post Quality in Online Discussions on Software, 45th Annual Meeting of the Association for Computational Linguistics ACL2007. - Prague, Czech Republic, ACM Press, New York, NY, June 23-30, 2007. - Vols. Volume P07-2. - pp. 125-128.