DAIICT-LDRP at TREC RTS 2017: Real Time Push Notification and Post Summarization

Sandip Modha *sjmodha@gmail.com*[1],
Chintak Mandalia *chintak.soni*75@*gmail.com*[2],
Shyamal Shah*shah.shyamal*872@*gmail.com*[2],
Sahil Kewlani*kewlaniz@gmail.com*[2],
Bhavya Shah *bhavyashah248@gmail.com*[2],
Deep Doshi *deep3197@gmail.com*[2], and
Prasenjit Majumder *prasenjit.majumder@gmail.com*[1]

[1]DAIICT Gandhinagar Gujarat-382007
[2]LDRP Gandhinagar Gujarat-382016

October 24, 2017

ABSTRACT

This paper present the participation of Information Retrieval and language processing lab(IRLPLAB) at Dhirubhai Ambani- Institute of Information and Communication Technology, at Gandhinagar, India in Real-Time Summarization track TREC 2017. This year TREC RTS offered two tasks. In the first task, that is scenario A, our system will be monitoring continuous posts from Twitter public stream and push the relevant tweet for each interest profile to RTS broker. We have done query expansion using Named entity and hash-tag of past 30 days' tweets. We have calculated relevance score between tweets and expanded interest profile using Okapi BM25 model and customize ranking function. For Scenario B, Email digest, we anticipated summarization problem as a clustering problem. In scenario A, we reported result in terms of Expected Gain EGp (primary metric) =0.1998 and in scenario B we have achieved primary metric nDCG-1 = 0.1324.

Keywords: Social media, BM25, Language Model with JM and dirichlet smoothing clustering, jaccard similarity,

1 INTRODUCTION

Social media, like Facebook, Twitter, Snapchat are some of the massive sources of real-time information. Twitter, a popular microblogging website, which has massive user-generated content due to its large number of registered users. This year TREC 2017 has offered Real-Time Summarization (RTS) track with following objectives: (i) how fast the participating system can deliver relevant and novel tweets based upon the interest profile to mobile assessor via RTS broker through push notification. (ii) How we can generate the day-wise summary of tweet for given interest profiles.

Twitter allowed its registered user to post up to 140-character short message called tweet. Because of this, it becomes challenging for the participating system to calculate its relevancy against the interest profile. In the rest of paper, we will use tweet and post interchangeably. The organizer of TREC 2017 RTS [1] gave 188 interest profiles having topic-id, title, narrative and description.

The two scenarios were: Scenario A [1]: Push notifications: Participating system continuously listens to the Twitter sample stream using Twitter Streaming API [7]. The Twitter streaming API offers an approximately 1% sample of all tweets and is freely available to all registered users. As soon as the system identifies a relevant post against the required profile, it is immediately pushed to broker via a push notification. Push notifications should be relevant, timely and novel. Scenario B[1]: Email digest. Alternatively, a user might want to receive a daily email digest that summarizes what happened that day with respect to the interest profiles. At a high level, these results should be relevant and novel; timeliness is not particularly important, provided that the tweets were all posted on the previous day.

For Scenario A, we pre-processed interest profile, removed all the stop words, and considered only noun, proper noun, and verb using Stanford POS tagger. In the next step, live tweets were collected, pre-processed and cleaned. Then we applied Okapi BM25 to get relevance score between tweets and each expanded interest profile. The threshold for interest profiles were estimated TREC RTS 2016 dataset. To ensure novelty we applied Jaccard similarity between the tweets which were already pushed and current candidate tweet

Scenario B, e-mail digest, the submission was after scenario A in which we had to send a maximum of 100 tweets for each profile each day. At the end of a day, tweets were sorted in descending order of BM25, language model with JM smoothing and dirichlet smoothing ranking function score for each interest profile. All the tweets above the threshold were selected and saved according to the template given by TREC 2017. Again threshold was estimated using TREC 2016 dataset.

The rest of paper is organized as follows: In section2 we discuss related work, In section 3 we define the problem statement for scenario A and B. In Section 4 we discussed the methodology for scenario A. In section 5 we describe the methodology for scenario B. In section 6 we discussed the result and performance evaluation metric. In section 7 we conclude the discussion.

2 RELATED WORK

We started our work by referring TREC MICROBLOG 2015 papers.

CLIP[2] has trained their Word2vec model using 4 years tweet corpus. They used Okapi BM25 relevance model to calculate the score. To refine the scores of the relevant tweets, tweets were rescored using the SVM rank package using the relevance score of the previous stage. Then Novelty Detection is done, where the tweets which are not useful are discarded, this is done using Jaccard similarity.

University of Waterloo[4] implemented the filtering tasks, by building a term vector for each user profile and assigning different weights to different types of terms. To discover the most significant tokens in each user profile, they calculated pointwise KL divergence and ranked the scores for each token in the profile.

3 PROBLEM STATEMENT

Scenario A was mainly real time filtering task. Given an interest profile Q=$\{Q_1,Q_2,..Q_n\}$, and stream of tweets T=$\{t_1,t_2,..t_n\}$ from public sample stream we need to compute the relevance score between tweets and profile R score=f(Q,T). Tweets having R score greater than threshold with respect to profile moved in the set PT= $\{pt_1,pt_2,...pt_n\}$. At most, 10 novel tweets can be pushed to RTS broker per profile per day.

In the Scenario B, summary s=$\{s_1,s_2,...,s_n\}$ has to be formed from relevant tweet RT=$\{rt_1,rt_2...rt_n\}$ where rt i represents relevant tweet for a particular profile. A batch of top 100 ranked tweets per day per interest profile with any two tweets having a similarity of less than threshold sim$(t_1,t_2) < Ts$ is used for Email Digest. All tweets from 00:00:00 to 23:59:59 were eligible candidates for a particular day.

4 METHODOLOGY FOR SCENARIO A:

4.1 Interest Profile Pre-processing

TREC RTS 2017 has given 188 interest profiles. We converted these profiles into query by removing stop words and considering nouns, proper nouns and verbs using Stanford POS tagger.

4.2 Profile (Query) expansion

We have downloaded profile specific tweet and extract top Named entity and hashtags for the each profile.

4.3 Profile Normalization

Interest profiles are consisting of title, sentence long description and paragraph long narrative. Title and description were merged so as to make interest profiles more informative. To increase the relevance, interest profiles were also pre-processed by converting all alphabets to small case and expanding the abbreviations. Also, interest profiles were stemmed.

4.4 Tweet Pre-processing

After gathering tweets, non-English tweets were filtered out. Tweet includes smileys, hashtags, and many special characters. We did not consider retweets and tweet with only hashtag or emoticon or special characters. We also ignored the tweet with less than 5 words and removed all the stopwords from the tweet.

4.5 Relevance Score

To calculate relevance score between tweets and interest profiles, we set weight as 2 for the original term in the interest profile and 1 for the terms added after training the profile. We have used BM25 model for calculating relevance score between expanded interest profile and query.Score is defined as:

$$R\_score = BM25\ Sim(Q\ exp,T)]$$

$$R\_score = \alpha T+\beta U+\lambda E$$

We have estimated α, β, λ using TREC 2016 Qrels.

4.6 Novelty detection

For novelty detection, Jaccard Similarity was used. Temporal feature is used to select most informative tweet from the cluster

$$J(A,B) = (A \cap B)/(A \cup B) + \mathrm{F(T)}$$

where A and B are pushed and current tweets respectively. The highest ranked tweet for each profile was sent to TREC for assessment. Now for next eligible tweet, we calculated it’s similarity with already sent tweet(s) so as to ensure novelty between them. Again a Jaccard threshold was decided and tweets below it were sent. Lower the similarity score, greater is the dissimilarity ensuring more novelty. Also, there were interest profiles where no relevant tweets for some day were found, which were known as SILENT DAYS.

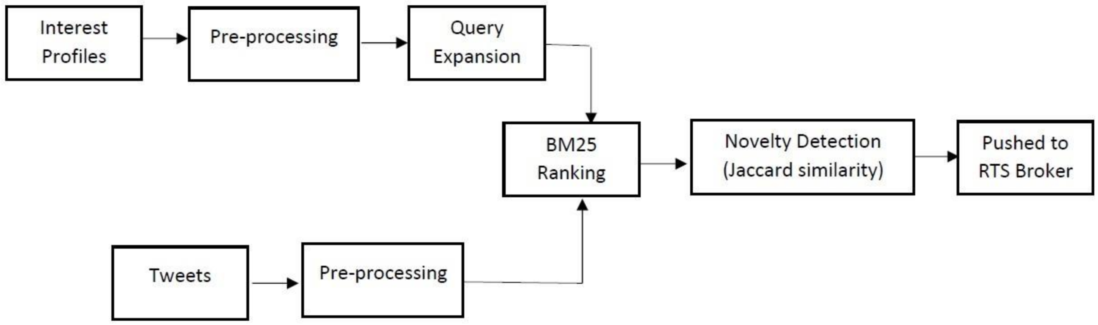

Figure 1: Scenario-A

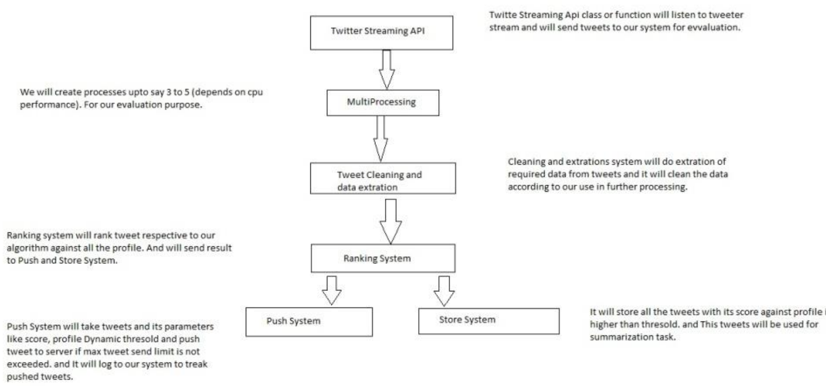

Figure 2 : Flowchart for ldrp run-2

5 METHODOLOGY FOR SCENARIO B:

In this scenario, we had to make a summary up to top 100 relevant tweets for each interest profile. We have submitted 3 runs using ranking function like language model with JM smoothing, dirichlet smoothing Okapi BM25 ranking function to calculate relevance score. We have set 2 thresholds for the experiments. T_s- silent day threshold and relevance threshold T_r Again, these threshold were estimated using TREC 2016 datasets.

6 RESULTS

The evaluation of TREC 2017 Microblog track lasted 10 days, It consisted of 188 interest profiles. During the evaluation time, participants listened to the tweet stream continuously and analyzed with every tweet.

6.1 Scenario A: User in the loop assessments

This live evaluation approach was started in TREC 2017 and promises a number of significant advantages over traditional post hoc batch evaluations because it is able to capture live user assessments. In this method, tweets submitted by participating systems to the RTS evaluation broker are immediately routed to an assessor, where it is rendered as a push notification containing the text of the tweet and the corresponding interest profile. The assessor may choose to judge the tweet immediately, or if it arrives at an inappropriate time, to ignore it. Either way, the tweet is added to a judging queue in a custom app on the assessor's mobile phone, which the assessor can access at any time to judge the queue of accumulated tweets. As the assessor judges tweets, the results are relayed back to the evaluation broker and recorded.

	Relevant	Redundant	Non Relevant	online utility(strict/lenient)	Median of online utility (both)
Irlab-run1	565	122	935	-492/-248	-805/-406
Ldrp-run2	640	197	2015	-1572/-1178	

Table 1: Result of Live User Assessment

6.2 Scenario A: Post Hoc Batch Evaluations

The evaluation methodology was based on pooling. A common pool has been constructed based on scenario A and scenario B submissions. The pool depth was be determined after the evaluation period ended by NIST based on the number of submissions and available resources. The assessment workflow is as follows: First, tweets returned by the systems were assessed for relevance. Tweets were judged as not relevant, relevant, or highly relevant.

Run	EGp	EG1	nCGp	nCG1	GMP.33	GMP.50
Irlab1	0.2065	0.1774	0.1929	0.1638	-0.1156	-0.0696
Ldrp-run2	0.1998	0.1617	0.1932	0.1551	-0.5084	-0.3634

Table 2: Result of Post Hoc Batch Evaluation scenario-A

6.3 Scenario B

In scenario B, nDCGp and nDCG1 score computed of each day for each interest profile and then average across them. We have reported very bad result in scenario-2. We found that lucene misses many hits in the rank list

Run	nDCGp	nDCG1
IRLAB-DAIICT	0.1324	0.0697
IRLAB-LDRP	0.1099	0.0773
IRLAB-LDRP2	0.0995	0.0619

Table 3: Scenario B result

7 CONCLUSION

In this paper, we have calculated relevance score between a tweet and expanded interest profile using BM25 Ranking function and customize ranking function. Novelty detection has been done using Jaccard similarity and temporal feature between previous pushed tweet and tweet to be pushed. We achieved average EG-p = 0.2065.

References

[1] *TREC 2017 Official Guidelines*

[2] Mossaab Bagdouri, Douglas W.Oard. *CLIP at TREC 2016: Microblog and LiveQA.*

[3] Xiang Zhu, Jiuming Huang, Sheng Zhu, Ming Chen, Chenlu Zhang, Li Zhenzhen, Huang Dongchuan, Zhao Chengliang, Aiping Li, Yan Jia *NUDTSNA at TREC 2015 Microblog Track*.

[4] Luchen Tan, Adam Roegiest, Charles L.A. Clarke *University of Waterloo at TREC 2015 Microblog Track*

[5] Luchen Tan, Adam Roegiest, Charles L.A. Clarke, Jimmy Lin *Simple Dynamic Emission Strategies for Microblog Filetering*

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean. *Efficient estimation of word representations in vector space. CoRR, abs/1301.3781, 2013*

[7] Luchen Tan, Adam Roegiest, Jimmy Lin, and Charles L. A. Clarke *An Exploration of Evaluation Metrics for Mobile Push Notifications*

[8] Shamanth Kumar, Fred Morstatter, Huan Liu *Twitter Data Analytics*