MSR KMG at TREC 2014 KBA Track Vital Filtering Task

Jingtian Jiang Chin-Yew Lin Yong Rui
Microsoft Research
Building 2, No. 5 Dan Ling St, Haidian Dist. Beijing, China 100190

{jiji, cyl, yongrui}@microsoft.com

ABSTRACT
In this paper, we present our strategy for TREC 2014 KBA track Vital Filtering task. This task is also known as "Cumulative Citation Recommendation" or "CCR" in 2012 and 2013. Vital Filtering task is to identify "vital" documents containing timely and new information that should be used to update the profile of a given entity (also called a topic). Our strategy for vital filtering is to first retrieve as many relevant documents as possible and then apply classification and ranking methods to differentiate vital documents from non-vital documents. We first index the corpus and retrieve candidate documents by combining entity names and their redirect names as phrase queries. We then learn to rank documents by leveraging four types of feature: 1) time range: the earlier documents get a higher score than the later documents, 2) temporal feature: burst of entity mentions, 3) title/profession feature: the title and profession information around an entity mention, and 4) action pattern: the entity name and its associated verb in the sentence mentioning the entity. A simple global adjustment is applied at the end to further improve system performance. Our experiment results confirm that these features are very effective, especially for action pattern and time range. The system incorporating all the proposed features significantly outperforms the phrase query baseline.

Categories & Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *Information Filtering*; H.3.m [Information Storage and Retrieval]: Miscellaneous – *Test Collections*; I.2.7 [Natural Language Processing] Text analysis – *Language parsing and understanding*

General Terms
Experimentation, Measurement

Keywords
KBA, CCR, Vital Filtering, Action Patterns, Time Range

1. INTRODUCTION
Knowledge Bases (KBs), such as Wikipedia, have been used in many applications including question answering, entity retrieval and entity linking. With the explosion of information on the web, it becomes critical to detect relevant documents and assimilate new information to entities in KBs in a timely manner. However, most KBs are maintained manually by volunteer editors, which are hard to keep up-to-date because of the limit number of editors and the huge number of entities in KBs. [3] indicates that the median time lag between the publication date of the cited articles and the date of the citations are created in Wikipedia is over one year. Moreover, some esoteric entities in KBs do not attract enough attention from the editors. It makes the maintenance more challenging. This gap could be reduced if relevant documents could be automatically found as soon as they are published and then recommended to the editors. Cumulative Citation Recommendation (CCR) was introduced by TREC KBA track to address this problem. A CCR system aims to filter candidate documents for a given set of entities from a stream corpus. CCR task continues this year with the name of “vital filtering”, but uses diversified entities and a larger stream corpus. KBA 2014 has two major changes related to Vital Filtering task: 1) all target entities are selected from a single geographic region centered on the region between Seattle and Vancouver. This region was specially selected to prepare for cross-language Chinese-English KBA in the future. 2) Not all target entities have Wikipedia profiles that can be used by CCR systems. It is a crucial aspect of CCR that the only available data about an entity may be examples from the corpus. To that end, it is ensured that all entities have some rated documents for training in the beginning of the stream. This year, the target entity set is composed of 67 entities. Among them, 37 entities have Wikipedia profiles.

KBA 2014 has augmented the stream corpus of KBA 2013, covering the time period from October 2011 to May 2013. The whole data set is huge but the organizers released a pre-filtered corpus as the official corpus. This corpus is about one tenth of the full corpus.

There is only one sub task of CCR in KBA 2014: vital filtering - treating only vital documents as positive instances and non-vital as negative instances. Vital + useful is just used to illustrate how hard the vital filtering is.

We submitted 7 runs to KBA 2014. All these runs used learning to rank. We used exact match run as our baseline, and our best result outperforms the baseline significantly.

The rest of this paper is organized as follows: section 2 introduces the pre-processing step to further reduce the size of stream corpus and get the candidate documents. We present our observations and features in section 3 and report our results in Section 4. Finally, we conclude and discuss future work in section 5.

2. PRE-PROCESSING
Before presenting the details of vital document filtering, we would like to briefly introduce the pre-processing step in our system, including indexing and retrieval. This step is to retrieve the candidate documents from the big corpus.

2.1 Indexing
In order to process the huge stream corpus more efficiently, we use ElasticSearch to index the official stream corpus and then simulate an hourly stream by issuing queries restricted to each hour in sequential order. ElasticSearch is an open-source, Lucene-based

text search engine[1]. In the data corpus, each document contains many fields. However, in the index we only care about a few fields of each document: stream_id, clean_visible, clean_html, timestamp and source. Table 1 describes the meanings of these fields.

Table 1. Document fields used in MSR KMG CCR system

Field	Description
stream_id	An unique identifier of the document
clean_visible	Plain text content of the document
clean_html	Html source code of the document
source	Source of the document (e.g., news, blog, forum, …)
timestamp	A timestamp measured in seconds since the 1970 epoch

2.2 Retrieval

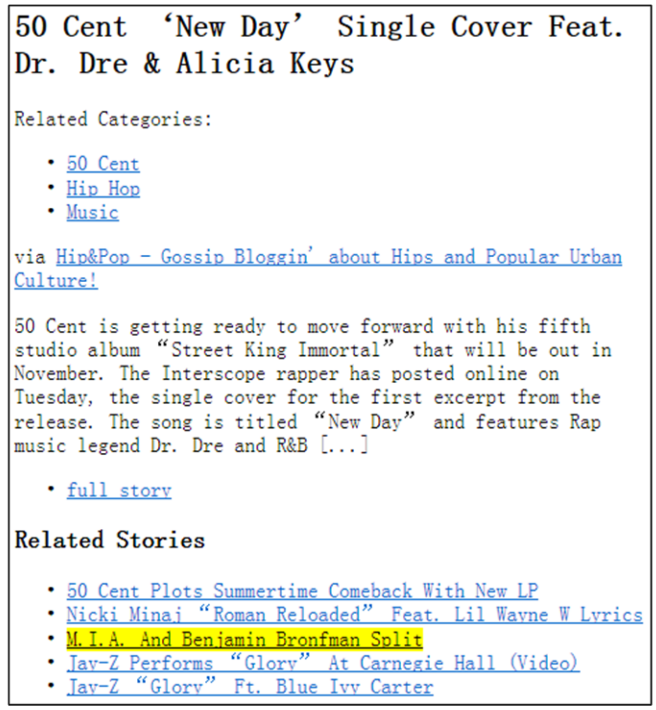
Based on the annotation of KBA 2012, none of the document with zero mention of target entity is annotated as central (vital), and there are only 0.4% of the documents with zero mention of the target entity have been labeled as relevant (useful) [3]. So we filter the corpus to create a compact working set by discarding the documents without any mention of the target entities. In order to do that, we query the built index with high recall queries. We construct a matchPhraseQuery[2] to assure that the retrieved documents should mention the target entity at least once by either its name or surface form and its expanded forms.

For each target entity from the Wikipedia, we extract the redirect[3] names as its surface forms. For example, Geoffrey E. Hinton, who is a machine learning scientist, has the following redirect names: Geoffrey Hinton, Geoff Hinton, Geoffrey E. Hinton, and Geoffrey Everest Hinton. For those target entities that do not have Wikipedia profiles, we treat them as they do not have any other surface forms. We also tried to use part of their names as surface forms, but that would retrieve many irrelevant documents. We construct a matchPhraseQuery with the target entity name and all its surface forms together and make a search against the index. Only the hit documents are processed in the following processing steps.

After the pre-processing step, we reduce the size of stream corpus to a great extent. The retained documents composed the working set for our CCR system. All our approaches were evaluated on this working set. After the pre-processing step, the number of docs needed to be processed decreased from 579,838,246 to 135,594. This makes our CCR system much more efficient.

3. VITAL FILTERING

In KBA 2013 CCR task, the best result is (P=0.206, R=0.736, F1=0.322, SU=0.157), which is not much higher than the baseline (P= 0.190, R=0.824, F1 = 0.310, SU = 0.157) [6]. The low performance indicates vital filtering is a hard problem. In order to find out what the challenges are, we spent much effort to carefully investigate the data in CCR 2013. In the following sub sections, we first report our observations and then introduce features inspired by the observations.


50 Cent 'New Day' Single Cover Feat. Dr. Dre & Alicia Keys

Related Categories:

- 50 Cent
- Hip Hop
- Music

via Hip&Pop - Gossip Bloggin' about Hips and Popular Urban Culture!

50 Cent is getting ready to move forward with his fifth studio album "Street King Immortal" that will be out in November. The Interscope rapper has posted online on Tuesday, the single cover for the first excerpt from the release. The song is titled "New Day" and features Rap music legend Dr. Dre and R&B [...]

- full story

Related Stories

- 50 Cent Plots Summertime Comeback With New LP
- Nicki Minaj "Roman Reloaded" Feat. Lil Wayne W Lyrics
- M.I.A. And Benjamin Bronfman Split
- Jay-Z Performs "Glory" At Carnegie Hall (Video)
- Jay-Z "Glory" Ft. Blue Ivy Carter

Figure 1. Related news could be a vital event (Benjamin Bronfman is the target entity)

3.1 Observations

After a careful investigation, we have the following observations:

1. A vital event can be found in a sentence and only in the sentence within a document. We found that many vital documents contain only one sentence about a target entity, e.g. the related news in a news document. It is annotated as a vital document because it has key information about the target entity. For example, as shown in Figure 1, the related story "M.I.A And Benjamin Bronfman Split" (highlighted in the figure) is a vital event to the target entity "Benjamin Bronfman". It is just a short sentence but not a long paragraph or a document about the entity. Notice that the main news of the document is not about Benjamin Bronfman but about 50 Cents.
2. The time range between a document and the beginning of the event is crucial. A vital event is usually associated with a lot of documents. But according to the definition of vital document, the documents appearing1~3 days after the beginning of an event are no longer vital if they do not have new information about the event. A cluster of similar documents within a specific range of time could be considered as reporting a same event, while the first document in the cluster signals the beginning of the event. Accordingly, the earlier documents are more likely to be vital than the later documents in the same cluster.
3. Temporal information is important. Typically, after an event happened, there will be a lot documents reporting the event at

[1] http://www.elasticsearch.org/

[2] http://www.elasticsearch.org/guide/en/elasticsearch/reference/current/querydsl-match-query.html

[3] http://en.wikipedia.org/wiki/Wikipedia:Redirect

the beginning. And as time goes on, less documents will report the event. Therefore, it is likely to have a burst of mentions of a target entity if a vital event about the target entity occurs. KBA 2013 overview [6] called this “spike”.

4. An acting target entity in a document is a good vital indicator of the document. In the true positive instances of KBA 2013 CCR task, the target entities usually participate in events, i.e. do something or announce something, etc. On the contrary, in most false positive instances, the target entities are just mentioned or involved in general actions, e.g., saying or telling something. Therefore, if a target entity is associated with an *action* (e.g., “Chad Kroeger to marry Avril Lavigne”, “Peter Goldmark won re-election”), then this document is very likely to be vital to the target entity. Otherwise, if we cannot detect any action involved or the action is a general action, then this document is less likely to be vital (e.g., “Ted Sturdevant was happy with the ruling”, “Dan Satterberg said in a statement”).

3.2 Clustering

As we discussed in our observations, many documents are annotated on sentence level but not document level. Thus it would be better to consider the sentence mentioning a target entity instead of the whole document, because the main content could be irrelevant to the event as we showed in Figure 1. So we decide to filter the documents based on sentences. A document may contains more than one sentence that mentions an entity. In this case, we take the max rating of all sentences as the document’s rating. That is, if one sentence is considered as vital, then the document is vital no matter what the ratings of other sentences are.

In addition, an event would be reported by a lot documents, and these documents are almost the same. And as we discussed earlier, among all documents reporting the same event, the earlier documents are considered as vital while the later are non-vital. Thus it requires a good CCR system to put the documents reporting the same event together and give larger scores to early documents and smaller scores to later documents.

Accordingly, after retrieving the candidate documents from the index, we extract all sentences mentioning the target entities, and perform a KNN-like algorithm to cluster the sentences. For a target entity, given a new sentence, we compute the word-based similarity between the new sentence and existing clusters. If the max similarity is larger than a pre-defined threshold, the new sentence is assigned to the corresponding cluster; otherwise, it would form a new cluster by itself. The threshold is decided based on the training data. When it is set to 0.9, the clustering algorithm achieved best results on the training data. In each cluster, the first arrived document was considered as the beginning of the event.

3.3 Features

In this section, we present the features used in our system. Balog et al. [2] has summarized 4 kinds of useful features for CCR, including document features, entity features, entity-document features and temporal feature. The best system in KBA 2013 also adopted these features and enriched them by adding citation similarity feature. However, our experiments show the similarity features are not very effective. For example, as shown in Figure 1, if the related news is a vital event, then the main content is not necessary to be similar to the target entity’s profile. The document’s main content is about a music band’s new single (the document’s title is “50 Cent ‘New Day’ Single Cover Feat. Dr. Dre & Alicia Keys”), which has nothing to do with “Benjamin Bronfman”.

Inspired by our observations in Section 3.1, besides the temporal feature used in KBA 2013, we propose a few new features to differentiate the vital documents from others more effectively. The features used in our CCR system are: *time range, title/profession feature, temporal feature,* and *action patterns*. We describe these features in detail in the following subsections.

3.3.1 Time Range

In the vital definition, it states that one must assess the time lag between the reporting and the event. More specifically, the documents appearing 1~3 days later which talk about that same event are no longer vital, but useful.

Therefore it is intuitive to assume that the later a document is, the smaller score it should get. Therefore, we penalize the later documents in a cluster. In practice, we decreased the feature value of a document over hours. That is, the first document of a cluster would get a feature value of 1.0, and the later documents would get smaller values. That could be expressed by the equation below:

$$\mathrm{tr}(d_i) = 1.0 - (\,h_i - h_0)/72.0$$

Where h_0 is the hour converted from the timestamp of the first document d_0 in that cluster, and h_i is the hour of the ith document d_i. The magic number 72.0 means 3 days: 24 hours x 3 days. The value $\mathrm{tr}(d_i)$ is used as one feature in our systems.

3.3.2 Temporal Feature

When a vital event occurred, a lot of documents would report that event in a short period of time, e.g., several hours or several days. And a vital event should be reported by more documents than a non-vital event. So we believe that the temporal feature would be helpful for vital filtering. Ludovic et al. [4] have tried statistics gathered on a sliding window over the past week as temporal feature, such as the number of entity's mentions in earlier documents. In CCR task, the sliding window is at hour level. Thus in our work, we used a past 10 hour sliding window. Balog et al. [2] found that Wikipedia page view statistics is a useful signal if something vital is happening about the target entity at a given point in time. But we did not use this feature because not all target entities in CCR 2014 have Wikipedia profiles.

The sudden ascend of entity mentions in a stream is called *burst* in our work. The magnitudes of mention of different entities vary according to their popularity. To normalize the gap between different entities, we define a burst value for each entity-document pair. Given an entity e and a document d, the pair’s burst value is defined as below:

$$\text{burst\_value}(e, d) = \frac{m(e,\ h)}{\left(\frac{M(e)}{N}\right)}$$

Where, h represents the hour converted from the timestamp of the document d, and $m(e,\ h)$ is the number of mentions of entity in hour h. $M(e)$ is the total mentions of entity e from the beginning of the stream corpus to hour h, and N is the total hours from the beginning of stream corpus to hour h.

Typically, the burst_value of a vital document would be greater than 10. For some popular entities, a vital document’s burst_value may be bigger. For example, Figure 2 shows the burst_value of the entity “Chad Kroeger” over hours. The two peak burst_value is related to 2 vital events: “Chad Kroeger to marry Avril Lavigne”, and “Details on Chad Kroeger’s wedding were revealed”.

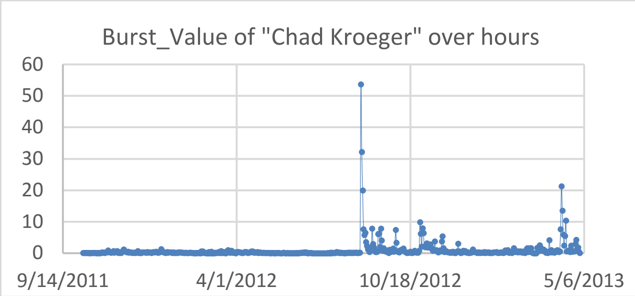


Figure 2. An example of burst_value of "Chad Kroeger"

3.3.3 Title/Profession Feature

Although the basic query can retrieve the documents mentioning the entities, it cannot disambiguate ambiguous entities with a same name or surface forms. For example, the target entity "Bill Templeton" is a coach of a basketball team. But in one of the retrieved documents, the mentioned "Bill Templeton" is not the coach. To disambiguate mentions of different entities, we propose the *title/profession feature*, which makes use of the title and profession information around a mention. Usually, when a person is mentioned, his or her title or profession is also mentioned around the name to let people know who this mention is. For example, "All the other stuff matters not, Lions coach Bill Templeton said.", and "… said Bill Templeton, an organizer for the local chapter of Pennsylvania Association of Sustainable Agriculture (PASA)". Of course, if the mentioned person is very popular, maybe his or her title or profession will be ignored. But that case is rare.

In our work, we used the similarity between a target entity's title/profession and the extracted title/profession of its possible mention as a feature in our systems. To enable this comparison, we need to obtain the title/profession information for each target entity first. To do so, we first built the title/profession dictionaries from Freebase [4]. The title dictionary contains 2,294 titles and the profession dictionary contains 2,440 professions. Then we adopt the word-based n-gram (n=1, 2, 3, 4, 5) inside a [-5, 5] window around a mention. These n-grams found also in the dictionaries form a title/profession vector. At last, we construct the title/profession vector for each target entity using the n-grams from all vital and useful documents in the training data.

In training or evaluation, for each entity-document pair, we first extracted the title/profession vector from the document and then compute the cosine similarity between the extracted vector and the entity's title/profession vector. The cosine similarity is used as a feature in our systems.

3.3.4 Action Patterns

As we have observed, vital documents typically contain sentences which describe target entities take part in events in which the target entities carry out some actions, e.g. scored a goal, read a poetry in public. According to this definition, an entity's action is a key indicator of whether a documents is vital or not. Then the problem is how to detect the action an entity has taken or experienced. We found that if an entity involves in an action, the entity usually appears as the subject or object of the sentence. So we decided to mine triples from sentences mentioning a target entity. If a triple is found and the target entity is the subject or object, then we say this entity takes action in the event.

To mine triples, we use Reverb [7], Reverb is a state-of-the-art open domain extractor that targets verb-centric relations, which have been shown in [8] to cover over 70% of open domain relations. Such relations are expressed in triples <subjective, verb, objective>, which exactly meet the requirements of our task.

We run Reverb on each sentence mentioning a target entity and get the triples. Then for each triple, we use the "entity + verb" (if the entity is the subjective) or "verb + entity" (if the entity is the objective) as an action pattern. Note that the verb is stemmed. For example, from the sentence "Public Lands Commissioner Democrat Peter Goldmark won re-election", the extracted triple is<Peter Goldmark, won, re-election>, and the action pattern is "Peter Goldmark win". In our systems, each action pattern is a binary feature: if the sentence/document has an action pattern, the feature value is 1; otherwise 0.

3.4 Models

Vital filtering task could be considered as a binary classification problem to differentiate vital documents and non-vital documents. It could also be considered as a ranking problem because of the relevance levels, i.e. vital > useful > unknown > non-relevant. As demonstrated by Balog et al., ranking-based approaches have more potential than classification approaches on all evaluation measures [1]. Our experiments also confirm that. Therefore, we concentrated more on ranking-based approaches.

In CCR 2012 and CCR 2013, we found that creating a separate model for each target entity could achieve better results than a general model for all target entities. In CCR 2014, the annotated data are split to ensure each entity has enough training examples. So in our system, we trained a separate ranker for each target entity using the features described in Section 3.3. As the random forest ranking method achieved the best results in CCR 2013, we also adopted random forest implemented in RankLib[5] as our ranking method.

3.5 Global Adjustment

In the final submissions, the score of each document is scaled to (0, 1000]. In the evaluation, F1 will be computed at each confidence threshold and take the maximum F1 as the single score for a system. However, as we trained different rankers for different target entities, different entities may have different confidence thresholds to achieve their best F1. For example, entity A gets its best F1 at threshold 900, while entity B gets its best F1 at threshold 500. And the final F1 is computed at threshold 700. At the threshold 700, entity A's results is not the best; so is entity B. Thus the global F1 might not be the best F1 for some systems.

To get the best global F1, we need to do an exhausted search across different confidence threshold for all target entities. But it is not feasible because there are 67 entities and each entity has 1,000 confidence thresholds. However, we make a simple optimization: we linearly adjust the documents' scores of a target entity using the training documents, make it achieves the best F1 at a given threshold, i.e. 500. Then we could ensure all entities would get the best F1 at threshold 500, and the global F1 would be improved.

4. RESULTS

The primary metric for vital filtering (CCR) is maximum macro-averaged F1 measure. F1 is a function of confidence cutoff. By sweeping the confidence threshold, we obtain a range of precision

[4] http://www.freebase.com/

[5] http://sourceforge.net/p/lemur/wiki/RankLib/

(P) and recall (R) scores for each target entity. After averaging P and R across the set of target entities, we then compute F1 at each confidence threshold and take the maximum F1 as the single score for the system. KBA 2014 also defined another measure to evaluate the performance of a system: max(SU) [5].

We first evaluated the different features. Error! Reference source not found. **shows the changes of the performance when different feature combinations are used. From the results, we can see that *action patterns* is the most effective which improved the performance by 4.5 points. The *time range* is also very powerful as it improves the result by 3 points. *Temporal feature* is also useful. *Title/profession feature* helps but is not as effective as we have expected. We analyze the results and find that most of the target entities are not ambiguous. On the 5 entities that is ambiguous, this feature could achieve nearly 4 point's improvement. *Global Adjustment* is simple, but it really improved the performance. It is even more effective than *title/profession feature* and *temporal feature*. The last row's system setting is corresponding to our best result in**

Table 3. In this system, we set the number of bags to 800 and sub-sampling rate to 1.0 in our model.

Table 2. The performance changes with adding more features. (All measures are reported by official scorer with cutoff-step-size=10.)

Features	avg(P)	avg(R)	max(F(avg(P), avg(R)))	max(SU)
Baseline	0.288	0.953	0.442	0.267
+Time Range	0.342	0.774	0.474	0.349
+Time Range +Temporal Feature	0.367	0.743	0.491	0.367
+Time Range +Temporal Feature +Title/Profession Feature	0.378	0.744	0.501	0.377
+Time Range +Temporal Feature +Title/Profession Feature +Action Patterns	0.447	0.702	0.546	0.464
+Time Range +Temporal Feature +Title/Profession Feature +Action Patterns +Global Adjusting	0.472	0.706	0.566	0.510

Table 3. Results of official runs. (All measures are reported by official scorer with cutoff-step-size=1.)

Run	avg(P)	avg(R)	max(F(avg(P), avg(R)))	max(SU)
Baseline	0.287	0.948	0.441	0.267
TR_P_All_GA	0.436	0.621	0.513	0.454
TR_P_All_GA_1	0.433	0.686	0.531	0.332
TR_P_All_GA_2	0.441	0.674	0.533	0.329
TR_PC_GA_1	0.476	0.588	0.526	0.489
TR_PC_GA_2	0.476	0.588	0.526	0.489
TR_PC_GA_3	0.476	0.588	0.526	0.489
TR_PC_GA	0.476	0.588	0.526	0.489

The results of our final runs are listed in

Table 3. The best system achieved F1=0.533 and SU=0.329. These runs could be divided into two group: the first 3 runs form one group and the remaining runs form the other group. The difference is that the first group used less patterns than the other group. So the first group has higher recall but lower precision. Though the second group's F1 is a little bit lower than the first group, its SU is significantly higher than the first group. In the first group, the difference between different runs is due to different parameters used in the models. The best run used 800 bags and the sub-sampling rate is set to 1.0. Other runs used less bags and smaller sub-sampling rate. In the second group, these runs used the same parameter setting (just more patterns than the best run). But in global adjustment, we adjusted the best F1 to different thresholds, so their performance is the same.

We also did some analysis on the error cases. We found there are 3 main reasons: 1) about 70% cases are due to new action patterns. There are many new patterns that does not exist in the training data. These cases with new patterns were considered as vital in our systems. For example, the patterns like "held by state Rep. Andy Billig" and "Jeff Mangum extend" do not exist in training data. 2) About 20% cases are due to inconsistent annotation. Some documents that were 3 days after the event were also annotated as vital; or for some documents with the same content, a few were annotated as vital while others were annotated as non-vital. For instance, one document reporting "Schools Chief Randy Dorn Announces Re-Election Bid" about 8 days after the event, but it is still annotated as vital. And another document reporting "Randy Dorn Issues Statement on Legalized Weed" is annotated as non-vital even it is just several hours after the event. 3) The remaining are due to the patterns' meaning changed. For example, in training data, a pattern is related to vital documents; but it is related to non-vital documents in the test data. An example is "Bill Templeton say". In the training data, it is related to all vital documents, but in the test data, most documents with this pattern are annotated as non-vital.

5. CONCLUSIONS

In this notebook paper, we presented our observations and features for KBA 2014 Vital Filtering task. In our work, we carefully investigate KBA 2013's data and mainly use four kinds of effective features in KBA 2014: Time Range, Temporal Feature, Title/Profession Feature, and Action Patterns. The experiments results show that these features are effective. Besides, we also make Global Adjustment to achieve a higher F1.In future, we would like to explore new effective features to handle the new action patterns and the change of patterns' meaning. We would also try more models to improve the performance.

6. REFERENCES

[1] K. Balog and H. Ramampiaro. Cumulative citation recommendation: classification vs. ranking. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, SIGIR 2013, pages 941-944, New York, NY, USA, 2013.

[2] K. Balog, H. Ramampiaro, N. Takhirov, and K. Nørvåg. Multi-step classification approaches to cumulative citation

recommendation. In Proceedings of the 10th Conference on Open Research Areas in Information Retrieval, OAIR'13, pages 121-128, Paris, France, 2013.

[3] J. R. Frank, M. Kleiman-Weiner, D. A. Roberts, F. Niu, C. Zhang, C. Re, and I. Soboro. Building an Entity-Centric Stream Filtering Test Collection for TREC 2012. In Proceedings of the Text REtrieval Conference (TREC), 2012.

[4] V. B. Ludovic Bonnefoy and P. Bellot. Lsis/lia at trec 2012 knowledge base acceleration. In Proceedings of the Text REtrieval Conference (TREC), 2013.

[5] S. Robertson and I. Soboro. The trec 2002 filtering track report. In TEXT RETRIEVAL CONFERENCE, 2002.

[6] J. R. Frank, S. J.Bauer, M. Kleiman-Weiner, D. A. Roberts, N. Tripuraneni, C. Zhang and C. Ré. Evaluating Stream Filtering for Entity Profile Updates for TREC 2013. In Proceedings of the Text REtrieval Conference (TREC), 2013.

[7] Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying Relations for Open Information Extraction. In Proceedings of EMNLP 2011.

[8] Michele Banko and Oren Etzioni. 2008. The Tradeoffs between Open and Traditional Relation Extraction. In Proceedings of ACL 2008.