1. Introduction

Nowadays, there is an overload of information available on the internet, from different sources such as news articles, Web documents, blogs, wikis, etc. Articles released from different sources may report and discuss the same event, so those who are interested in one event may end up reading many redundant information. Therefore, a novel sentence mining system is useful to filter out the irrelevant and redundant information while retrieving the relevant and novel sentences. The goal of a novel sentence mining system is to return a complete (i.e. no missing novel information) and concise (i.e. no redundant information) view for a specific topic predefined by a user.

A typical novel sentence mining system has two modules, i.e. categorization and novelty mining (novelty detection). Categorization classifies each incoming document into its relevant topic bin. Then, novelty mining retrieves the sentences with novel information from those relevant documents in each topic bin. Standard document level novelty mining focuses on identifying novel documents from a document stream. However, due to the rich contents of a document, even a highly novel document may contain some redundant information. Therefore, in recent years there have been more studies conducted on novel sentence mining, which is the focus of our paper.

Early research on novel sentence mining was conducted in the TREC Novelty Track from 2002 to 2004 (Robertson & Soboroff, 2002; Soboroff, 2004; Soboroff & Harman, 2003), which provided both the task definitions and the public data. Previous work for novel sentence mining focused on two major areas. The first area focused on how to improve the representation of the natural language sentences, e.g. methods using named entities (Abdul-Jaleel et al., 2004; Collins-Thompson, Ogilvie, Zhang, & Callan, 2002; Ru, Zhao, Zhang, & Ma, 2004; Zhang, Xu, Bai, Wang, & Cheng, 2004), entity patterns (Li & Croft, 2008), synonyms (Eichmann et al., 2004), part of speech (Ru et al., 2004; Zhang et al., 2004; Zhang & Tsai, 2009), etc. The second category of work concentrated on creating or comparing the different novelty metrics. Based on different models chosen, the novelty metrics can be divided into those based on the vector space model (such as cosine similarity) and those based on language models (such as the Kullback-Leibler divergence) (Allan, Wade, & Bolivar, 2003; Zhang, Callan, & Minka, 2002). The vector space model represents sentences by counting words or sequence of words (Ailguliyev, 2009), while the language model represents sentences by modeling the probability distributions of words. Allan et al. reported that the novelty metrics based on the vector space model would continue to perform best even for noisy data (Allan et al., 2003).

Although many papers have reported the good performance of vector space model-based novelty metrics, to our knowledge, no studies have analyzed the preconditions and reasons why certain metrics did well on the data. Further knowledge on the limitations of the different novelty metrics can serve as guidance for choosing suitable novelty metrics for different types of data. This paper attempts to fill this gap by first analyzing the potential pitfalls of two typical novelty metrics: cosine similarity and new word count. We discover that the performance of novelty metrics is affected largely by the ratio of novel sentences of a topic. In an experimental study, we found that the cosine similarity metric performed well for those topics with a high ratio of novel sentences while the new word count metric showed the opposite characteristics. In order to fully utilize the merits of both metrics, a blended metric...
that combines both cosine similarity and new word count is proposed and tested on TREC 2003 and 2004 Novelty Track data. To our knowledge, there is no previous work that combines novelty metrics, which is especially useful for real-time novelty mining of sentences where novelty ratios are unknown beforehand. The experimental results show that the blended metric not only enhances the overall performance of a real-time novel sentence mining system, but also makes the system more flexible to different users’ requirements.

This paper is organized as follows. In Section 2, related work is reviewed and the possible pitfalls of two typical novelty metrics are analyzed. The proposed blended metric is presented in Section 3 and tested in Section 4. Section 5 concludes the paper.

2. Novelty metrics

Intuitively, a novel sentence means that the sentence containing sufficient novel information, based on a user’s history and preferences. Usually, there are two sequential steps in detecting a novel sentence: (i) scoring its degree of novelty by some novelty metric based on its history sentences, and (ii) making a final decision on whether the sentence is novel or not based on whether the novelty score falls above or below a predefined threshold.

2.1. Related work

In the TREC Novelty Track papers (Robertson & Soboroff, 2002; Soboroff, 2004; Soboroff & Harman, 2003), the proposed novelty metrics can be broadly grouped into two categories, i.e. metrics for sentences that are represented by the language model and metrics for sentences that are represented by the vector space model. In the language model, each sentence is represented by a probability distribution of words and the information theoretic metric Kullback–Leibler divergence is usually used to measure the dissimilarity between two probability distributions. In the vector space model, each sentence is represented by a weighted vector, and cosine similarity is a typical metric to calculate the similarity between two vectors. The novelty metrics can also be grouped based on their symmetric properties. Symmetric novelty metrics such as cosine similarity and Jaccard similarity evaluate the similarity between sentences without concern for the sentence order. Asymmetric novelty metrics such as new word count, set difference and overlapping metric can take the order information into account (Zhao, Zheng, & Ma, 2006).

Zhang et al. reported the good performance of cosine similarity for novelty mining at the document level (Zhang et al., 2002). Allan et al. indicated that, although the cosine similarity can work well at the document level, it may not be so effective for the short text like sentences (Allan et al., 2003). Zhao et al. reported that the similarity metrics such as cosine similarity may not be well-suited to the nature of novel sentence mining because the relationship between the current sentence and its history sentences is externally asymmetric, while the asymmetric metrics such as overlapping metric should be more suitable from the theoretical point of view (Zhao et al., 2006). However, their experimental results of overlapping metric did not seem to be better than similarity metrics.

Next, we will analyze the potential pitfalls of two typical novelty metrics including cosine similarity and new word count metrics.

2.2. Potential pitfalls of two typical novelty metrics

2.2.1. Cosine similarity novelty metric

In novel sentence mining, the cosine similarity novelty metric calculates the similarities between the current sentence si and each of its history sentences sj (1 ≤ i ≤ t – 1), which determines the novelty score for si, as shown in Eq. (1).

\[ J_{\cos}(s_i) = \max_{1 \leq j \leq t-1} \cos(s_i, s_j) \]
\[ \cos(s_i, s_j) = \frac{\sum_{k=1}^{n} w_k(s_i) \cdot w_k(s_j)}{|s_i| \cdot |s_j|} \]

where \( J_{\cos}(s) \) denotes the cosine similarity score of sentence s and \( w_k(s) \) is the weight of kth element in sentence weighted vector s. The weighting function used in our work is tf.idf (term frequency multiply inverse sentence frequency) weighting function as defined below.

\[ w_k(s) = tf_{wk} \log \left( \frac{n + 1}{n_{wk} + 0.5} \right) \]

where \( tf_{wk} \) is the frequency of the word \( w_k \) in sentence s; \( n_{wk} \) is the number of sentences, in which the word \( w_k \) appears in the collection; n is the number of sentences in the collection.

To illustrate and analyze the possible pitfalls of the cosine similarity novelty metric, we first give a simple example. Suppose there are three sentences \( s_1 \sim s_3 \) as follows:

\[ s_1 = [A] \quad s_2 = [A, B] \quad s_3 = [A, B, C] \]

where A, B and C are three different terms or words of sentences. If we use cosine similarity to measure the incoming sentence stream \( s_1 \sim s_2 \sim s_3 \), where “…” denotes the incoming order, the first incoming sentence \( s_1 \) is absolutely novel (Zhang et al., 2002). Next, since \( s_1 \) is the only history sentence for \( s_2 \), novelty score for \( s_2 \) is \( J_{\cos}(s_2) = \cos(s_1, s_2) \) as defined in Eq. (1). If we suppose the cosine similarity score \( J_{\cos}(s_1) \) is smaller than some predefined threshold \( \theta \), i.e. \( J_{\cos}(s_2) < \theta \), \( s_2 \) will be predicted as a novel sentence. For the sentence \( s_3 \), again, the cosine similarities between \( s_1 \) and each of the history sentences are calculated and the maximum will be assigned to \( J_{\cos}(s_3) \). Obvious, \( \cos(s_1, s_3) \) is assigned to \( J_{\cos}(s_1) \) for \( \cos(s_1, s_3) > \cos(s_1, s_1) \). If we also suppose \( \cos(s_2, s_2) < \theta \), \( s_3 \) will also be predicted as a novel sentence.

However, if the order of incoming sentence stream is changed, some problems will occur. For example, if the incoming order of these three sentences is reversed, i.e. \( s_1 \sim s_2 \sim s_3 \), the system will also retrieve all three sentences if we use the same assumptions, i.e. \( \cos(s_2, s_2) < \theta \) and \( \cos(s_2, s_3) < \theta \). Obviously, this result makes little sense. Based on these assumptions, it is easy to find that \( s_1 \), \( s_2 \) and \( s_3 \) will always be retrieved no matter how they are arranged.

This problem may be caused by two reasons: (i) pairwise comparison between sentences, and (ii) symmetric properties of cosine similarity, i.e. \( \cos(s_1, s_2) = \cos(s_2, s_1) \). Although the order of the incoming sentences does affect the results of novelty detection, this example reveals that the cosine similarity novelty metric cannot accommodate for the different orders. This potential pitfall of cosine similarity may not appear problematically in the situation where most of the incoming sentences contain enough novel information, e.g. in the case of \( s_1 \sim s_2 \sim s_3 \). However, in other cases where most of the incoming sentences do not contain novel information, cosine similarity may retrieve a sentence just because it is dissimilar to the history sentences. Whether this sentence contains novel information cannot be clearly determined. Therefore, further quantitative analysis for the cosine similarity novelty metric is necessary.

2.2.2. New word count novelty metric

To avoid the problem of the cosine similarity novelty metric, we may select other metrics. Out of many other novelty metrics, the new word count novelty metric shows very diverse characteristics. Using new word count novelty metric, the new sentence is compared with its entire set of history sentences. Furthermore, it is
an asymmetric metric where the so-called “new word” can only be defined based on its history sentences. The new word count novelty metric (Allan et al., 2003) assigns each sentence a count of new words that have not appeared in any history sentence, as defined below.

\[ J_{nwr}(s_t) = |W(s_t) \cap [\bigcup_{s_{t-1}}^s W(s_{t-1})]| \]  

where \( W(s_t) \) is the set of words in the sentence \( s_t \). In the previous example, when the new word count novelty metric is applied to the incoming sentence stream \( s_3 \rightarrow s_2 \rightarrow s_1 \), \( s_2 \) and \( s_1 \) will be predicted as “non-novel” sentences because they do not contain any new words.

Since the new word count novelty metric compares the current sentence with all the words in its history sentences (Eq. (3)), the number of new words will decrease as sentences gradually accumulate. However, a lower value of new word count does not necessarily mean that there is less novel information in a sentence. For example, suppose we define another three sentences \( s_4, s_5 \) and \( s_6 \) below.

\[ s_4 = [A, B, C, D] \quad s_5 = [E, F, G] \quad s_6 = [A, A, F, G] \]  

For the incoming sentence stream \( s_4 \rightarrow s_5 \rightarrow s_6 \), both cosine similarity and new word count novelty metrics can predict \( s_4 \) and \( s_5 \) as “novel”. For \( s_6 \), since it contains no new words, the new word count novelty metric will assign a zero novelty score to \( s_6 \), which cannot be retrieved by any threshold. However, \( s_6 \) has not only a different term frequency of words, but also a different combination of old words which may imply novel information. These subtle differences between sentences cannot be measured by the new word count novelty metric.

3. Blended metric

As analyzed in Section 2.2, cosine similarity and new word count metrics have complementary characteristics, which is a good basis for creating a metric to combine the benefits of both metrics. The goal of the blended metric is to integrate the merits of both metrics and hence generalize better over different topics. From a classification point of view, constructing a blended metric that combines the measurements from multiple metrics is equivalent to constructing a classifier using multiple features rather than only one feature. Since the characteristics of the two metrics differ as we discussed in the previous section, some samples that overlap in the one dimensional feature space could be separable in the two dimensional feature space. Moreover, the decision boundary in the two dimensional feature space which is defined by the blending strategy and threshold can be more flexible than that in the one dimensional feature space.

Two major issues for constructing a blended metric are (i) the scaling problem that ensures different component metrics are comparable and consistent and (ii) the blending strategy that defines the way of fusing the outputs from different component metrics.

3.1. Normalizing the component metrics

The cosine similarity score ranges from 0 (i.e. totally different) to 1 (i.e. exactly the same). The smaller the cosine similarity score, the more novel information in the sentence (Zhang et al., 2002). The value of the new word count novelty metric for any sentence is a non-negative integer such us 0, 1, 2, etc. As such, it is incomparable with the cosine similarity novelty metric. Therefore, the values of the new word count novelty metric can be normalized by the total number of words in the sentence as shown below.

\[ J_{nwr}(s_t) = \frac{|W(s_t) \cap [\bigcup_{s_{t-1}}^s W(s_{t-1})]|}{|W(s_t)|} \]  

where the denominator \(|W(s_t)|\) is the word count of the current sentence \( s_t \). This normalized metric, called new word ratio \( J_{nwr} \), has the range of values from 0 (i.e. sentence with 100% new words) to 1 (i.e. sentence with no new words). Using this normalized metric, the cosine similarity and new word ratio novelty metrics have become both comparable and consistent (i.e. the larger the value of metric is, the less novelty contained in a sentence).

3.2. Blending strategy

After normalizing the component metrics, we formulate the blended metric as shown in Eq. (6).

\[ J_{blended}(s_t) = \alpha J_{nwr}(s_t) + (1 - \alpha) J_{cos}(s_t) \]  

where \( \alpha \) is the blending parameter ranging from 0 to 1. In the experimental study, the proper weights of two component metrics defined by \( \alpha \) must be specified. The best parameter \( \alpha \), which is data dependent, needs to be learnt from some training samples, such as the user’s novelty feedback. However, the novel sentence mining algorithms in the previous work (Allan et al., 2003; Zhang et al., 2002; Zhao et al., 2006) as well as in our work are all unsupervised because the user’s feedback cannot be guaranteed in the practical novel sentence mining system. Instead of discussing the adaptive parameter setting here, a fixed \( \alpha \) is predefined in our experiments. We found the blended metric can obtain generally better results by using the different \( \alpha \) values. The parameter setting with \( \alpha = 0.75 \), which obtains the overall best performance for TREC 2003 and 2004 Novelty Track data is used throughout our experimental study.

4. Experimental study

4.1. Data collections

Currently, the possible public datasets for the experiments of novel sentence mining are TREC 2002 (Robertson & Soboroff, 2002), 2003 (Soboroff & Harman, 2003) and 2004 (Soboroff, 2004) Novelty Track data. In this experimental study, we use TREC 2003 and 2004 Novelty Track data. TREC 2002 Novelty Track data is excluded from the experiments because of its lack of non-novel sentences (90.9% of relevant sentences are marked as novel and 23 out of 50 topics have no sentences labeled as non-novel).

The statistics of data are summarized in Table 1.

<table>
<thead>
<tr>
<th>Data</th>
<th>Topics</th>
<th># Rel. sentences</th>
<th># Novel sentences</th>
<th>( r_{ns} ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREC2003</td>
<td>N1–N50</td>
<td>15557</td>
<td>10226</td>
<td>65.7</td>
</tr>
<tr>
<td>TREC2004</td>
<td>N51–N100</td>
<td>8343</td>
<td>3454</td>
<td>41.4</td>
</tr>
</tbody>
</table>
4.2. Experimental results

As we have discussed in Section 3, the goal of the blended metric is to make use of the merits of different component metrics while overcoming their weaknesses. Therefore, the two main goals of the experiments are (i) finding the “weak zones” for cosine similarity and new word ratio novelty metrics, and (ii) assessing the performance of the blended metric in these “weak zones”. For this purpose, we calculate the ratio of novel sentences \( r_{novel} \) for each topic and create three groups of topics corresponding to three levels (“zones”) of \( r_{novel} \), i.e. \( r_{novel} \in [0, 0.3), (0.3, 0.6] \) and \( (0.6, 1] \) as low, medium and high ratios of novel sentences.

Results reported in the tables and figures are the average of topics at each level of \( r_{novel} \). For each topic, the best \( F \) score over different thresholds uniformly distributed from 0 to 1 is selected as the primary measure, while the corresponding recall and precision are used as the supplemental measures.

Table 2 shows the average best \( F \) score at different levels of \( r_{novel} \) for TREC 2003 and 2004 Novelty Track data, respectively. From Table 2, the first observation is that the cosine similarity and new word ratio metrics have different “weak zones”, as denoted by ▲. The cosine similarity metric performs worst when the ratio of novel sentences is smaller than 0.3, while the new word ratio metric performs worst when the ratio of novel sentences is larger than 0.6. This observation shows that the “weak zones” of two single novelty metrics do not overlap.

Secondly, we can observe that the blended metric performs generally better than both cosine similarity and new word ratio metrics. The blended metric obtains the highest \( F \) scores, 0.8343 and 0.6538 in TREC 2003 and 2004 Novelty Track data, respectively. The second highest \( F \) scores are 0.8323 (cosine similarity in TREC 2003) and 0.6496 (new word ratio novelty metric in TREC 2004). Although the results of the blended metric are only incrementally better, these results prove that different types of single metrics which have their own pitfalls can benefit from fusing the information from each other. This is especially beneficial for real-time detection of novel information, where a priori novel rates cannot be estimated.

Fig. 1 integrates the performance of three novelty metrics over all 100 topics in TREC 2003 and TREC 2004 Novelty Track data, at 10 different levels of ratios of novel sentences. From this figure, it is clear that the \( F \) score of all three metrics are highly correlated with \( r_{novel} \). It can also be observed that cosine similarity performs worst when the ratio of novel sentences is smaller than 0.5. On the contrary, new word ratio performs worst when the ratio of novel sentences is greater than 0.5. The blended metric performs well at all levels of \( r_{novel} \). These results also prove that the “weak zones” for cosine similarity and new word ratio do not overlap, and that

### Table 2

<table>
<thead>
<tr>
<th>Data</th>
<th>Metrics</th>
<th>Best ( F ) score</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( r_{novel} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0–0.3</td>
<td>0.3–0.6</td>
<td>0.6–1</td>
</tr>
<tr>
<td>TREC2003</td>
<td>cos</td>
<td>N.A.</td>
<td>( 0.7209 ▲ )</td>
</tr>
<tr>
<td></td>
<td>nwr</td>
<td>N.A.</td>
<td>( 0.7214 ▲ )</td>
</tr>
<tr>
<td></td>
<td>Blended</td>
<td>N.A.</td>
<td>( 0.7267 ▲ )</td>
</tr>
<tr>
<td>TREC2004</td>
<td>cos</td>
<td>0.4854 ▲</td>
<td>0.6551 ▲</td>
</tr>
<tr>
<td></td>
<td>nwr</td>
<td>0.5246</td>
<td>0.6615 ▲</td>
</tr>
<tr>
<td></td>
<td>Blended</td>
<td>0.5242</td>
<td>0.6657 ▲</td>
</tr>
</tbody>
</table>

Fig. 1. Average best \( F \) scores (above) and the corresponding number of topics (below) in 10 different levels of ratios of novel sentences in TREC 2003 and TREC 2004 Novelty Track data.
the blended metric can utilize the merits of different types of novelty metrics.

Table 3 shows the recall and precision for these three metrics in different levels of ratios of novel sentences. One significant phenomenon is that the best F score of cosine similarity corresponds to high recall and low precision. Also, the cosine similarity metric retrieves more sentences than the other two metrics especially when the novelty ratio is low, as shown in Table 4. This implies that the cosine similarity metric may lead in the situation where high recall is required, such as the case where missing out important information is especially detrimental. On the other hand, the new word ratio metric may outperform in the situation where high precision is required.

4.3. Discussions

For a real-time novel sentence mining system that needs to satisfy different users’ requirements, the flexibility of the system to adapt to different requirements is especially critical. For example, when users do not want to miss any novel information, a high recall system which may retrieve more sentences is desired. When users want to read the most novel sentences first, a high precision system is preferred. It is well known that there is a tradeoff between recall and precision, which can be adjusted by a predefined threshold for novelty scores. However, the threshold can only adjust recall-precision tradeoff externally, while the intrinsic adjustability of recall and precision depends on the flexibility of the novelty metric itself. For example, Section 4.2 showed that cosine similarity is prone to retrieve more sentences, and hence may result in a system exhibiting high recall and low precision, which may not satisfy a precision-oriented requirement well.

The precision-recall (PR) curves for three novelty metrics are shown in Fig. 2. In information retrieval system, PR curves are usually used to compare different algorithms by comparing the areas under the curves, where the larger the area is, the better performance of the algorithm (Davis & Coadrich, 2006). In this experiment, we use the PR curve to compare the flexibility of different novelty metrics by comparing the range of values for both the precision and recall. The wider the range of values is, the more flexibility of the metric.

For the precision (y-axis in Fig. 2), the range of values for the precision of the cosine similarity metric is much narrower than those of new word ratio and blended metrics. For the recall (x-axis in Fig. 2), the range of values of the new word ratio is smaller than those of cosine similarity and blended metrics. The maximum recall obtained by the new word ratio is around 0.93, while the maximum recall obtained by the other two metrics can approach 0.99 (see the amplified graph on the right side in Fig. 2). The blended metric has a wider range of values for both recall and precision.

---

Table 3

<table>
<thead>
<tr>
<th>Data</th>
<th>Metrics</th>
<th>Recall</th>
<th>Precision</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r_{n/2}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREC2003</td>
<td>cos</td>
<td>N.A.</td>
<td>0.9365</td>
<td>0.9806</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>0.9137</td>
<td>0.9010</td>
<td>0.9058</td>
</tr>
<tr>
<td></td>
<td>Blended</td>
<td>N.A.</td>
<td>0.9332</td>
<td>0.9750</td>
</tr>
<tr>
<td></td>
<td>cos</td>
<td>0.7980</td>
<td>0.9225</td>
<td>0.9815</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>0.8668</td>
<td>0.9122</td>
<td>0.8372</td>
</tr>
<tr>
<td></td>
<td>Blended</td>
<td>N.A.</td>
<td>0.6965</td>
<td>0.8935</td>
</tr>
<tr>
<td>TREC2004</td>
<td>cos</td>
<td>0.7980</td>
<td>0.9225</td>
<td>0.9815</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>0.8668</td>
<td>0.9122</td>
<td>0.8372</td>
</tr>
<tr>
<td></td>
<td>Blended</td>
<td>N.A.</td>
<td>0.6965</td>
<td>0.8935</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Data</th>
<th>Metrics</th>
<th>Number of retrieved sentences</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r_{n/2}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREC2003</td>
<td>cos</td>
<td>200.84</td>
<td>259.35</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>189.53</td>
<td>233.04</td>
</tr>
<tr>
<td></td>
<td>Blended</td>
<td>198.53</td>
<td>258.26</td>
</tr>
<tr>
<td>TREC2004</td>
<td>cos</td>
<td>91.20</td>
<td>129.56</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>117.18</td>
<td>124.17</td>
</tr>
<tr>
<td></td>
<td>Blended</td>
<td>119.67</td>
<td>131.17</td>
</tr>
</tbody>
</table>
Therefore, for the novel sentence mining system which needs to satisfy different users’ requirements, the blended metric can be a better choice.

5. Conclusions

Novel sentence mining retrieves the sentences that contain novel information from a time sequence of relevant sentences for a specific topic, based on a user’s history and preferences. Previous research has shown that some typical novelty metrics such as cosine similarity and new word count performed effectively for detecting the novel sentences, but the preconditions for the good performance are not clear. This paper first analyzed the potential pitfalls for these two typical novelty metrics. Then, by testing these novelty metrics on topics with different ratios of novel sentences, we found each metric had its own “weak zone”, while the weak zones of different metrics did not overlap. The nonoverlapping weak zones provided a basis for creating a blended metric. To our knowledge, there is no previous work focusing on blended metrics for novelty detection.

The experimental results showed that, by blending both cosine similarity and new word ratio metrics, a generally better performance than single component metrics can be obtained on both TREC 2003 and 2004 Novelty Track data. Although the results of the blended metric are only incrementally better, they prove that different types of single metrics can benefit from fusing the information from each other. Moreover, we also found that the blended metric can obtain a wider range of values for both recall and precision. Therefore, for a real-time novel sentence mining system which needs to meet different users’ requirements (such as high recall or high precision), and in which novel rates are unknown beforehand, the blended metric can merge the best properties of both metrics.

References


