# CCNU at TREC 2016 Clinical Decision Support Track

**Yue Zhang[1], Fanghong Jian[2], and Po Hu[1]**

[1] School of Computer Science, [2] National Engineering Research Center for E-Learning,
Central China Normal University, Wuhan 430079, China

zy_930914@126.com, jfhrecoba@mails.ccnu.edu.cn, phu@mail.ccnu.edu.cn

**Abstract**

The task of the clinical decision support track in TREC 2016 requires a system to retrieve and return the most relevant biomedical articles after giving the electronic health report as query. Firstly, we consider two important factors i.e., query length and specific term's occurrence, and adopt two models by the combination of a novel TF-IDF method and the traditional BM25. We name them as MATFB and MATFM respectively. Then, a linear combination of the two models which is denoted as NEWBM is considered. The experimental results show that different kinds of queries might prefer one of the three models, and our system performs better than the median performance on most metrics.

## 1. Introduction

The goal of this track is to find relevant articles for the clinician to improve their decision making in the diagnosis, treatment and test of patients. In 2015, 0.73 million reports was published. In 2016, TREC dataset contains 1.25 million reports. Different from the previous cases where a synthetic query is provided, this year the real notes (i.e., hospital records) are given as queries and are extracted by clinicians, which contain a lot of abbreviations and other language styles. A note is usually longer than a summary and a description. Its average length is more than 300 words. All of this will bring new challenges for traditional retrieval systems.

Most of the existing retrieval models use a single term frequency (TF) normalization mechanism. For example, in the probabilistic models, documents are assumed as independent with each other in the collection, and the distribution of a term does not depend on the other terms in the same document. But frequency of a term in a document is relative to the frequency of the other terms in the same document, which gives important clue that cannot be achieved by the commonly used document length based normalization scheme. Another major limitation of the current models is that most models cannot achieve a good balance when dealing with short and long documents. If the parameter b is set to a smaller value, it will perform better for short queries, and when the parameter value is set larger, longer queries will benefit more than the shorter ones. The situation is suitable for BM25, language models or divergence from randomness based models. A novel TF-IDF term weighting scheme MATF [1] was ever proposed to retrieve related documents.

In our system, we consider two important factors i.e., query length and specific term's occurrence, and adopt two models by the combination of a novel TF-IDF method and the traditional BM25 which are named as MATFB and MATFM respectively. Then, a linear combination of the two models is also considered. The evaluation results show that our system performs better than the median performance on most metrics.

## 2. Our Approach

In this section, we will introduce the models used in our system in detail.

### 2.1 MATFM Model

We first propose an approach named MATFM, which combines TF of MATF with IDF of BM25. In 2013, Jiaul h.Paik proposed MATF model which introduces two aspects of term frequency normalization. One is the relative intra-document TF ( $RITF$ ) involving the average term frequency ( $AvgTF(D)$ ) of a document. The other is the length regularized TF ( $LRTF$ ) involving the average length of a documents ( $ADL(C)$ ). This method can maintain a good balance between short and long document and makes use of the query length information. In our system, we use asymptotically bounded function (similar to Robertson and Walker [2]) to transform the TF factors.

$$TFF(t,D) = \omega \times \frac{\log_2(1+TF(t,D))/\log_2(1+AvgTF(D))}{1+\log_2(1+TF(t,D))/\log_2(1+AvgTF(D))} + (1-\omega) \times \frac{TF(t,D)\times(\log_2(1+ADL(C)/len(D))}{1+TF(t,D)\times(\log_2(1+ADL(C)/len(D))} \quad (1)$$

where $\omega$ is the query length factor. Its value will be reduced with the increase of the query length, We confine the query length factor by the requirements. (1) $QLF\ (Q) = 1$ when $|Q| = 1$, (2) $QLF'(Q) < 0$ , (3) $0 < QLF\ (Q) < 1$. Lots of different functions can meet the conditions and our experimental choices are given as follows:

| Query | $\omega = QLF\ (Q)$ |
|---|---|
| summary | $QLF\ (Q) = \frac{2}{1+\log_2(1+3\lvert Q\rvert)}$ |
| description | $QLF\ (Q) = \frac{2}{1+\log_2(1+3\lvert Q\rvert)}$ |
| note | $QLF\ (Q) = \frac{2}{1+\log_2(1+\sqrt{\lvert Q\rvert})}$ |

Different kinds of queries usually have different $\omega$ values, so that we can adjust length information factor accordingly.

The formula of the traditional BM25 [3] is as follows:

$$rsv_q = \sum_{t\in q} \frac{(k_1+1)\cdot tf_{td}}{k_1\cdot(1-b+b\cdot(\frac{len(D)}{ADL(C)})+tf_{td}} \cdot \log(\frac{N-df_t+0.5}{df_t+0.5}) \cdot \frac{(k_3+1)\cdot qtf_t}{k_3+qtf_t} \quad (2)$$

where N is the number of indexed documents in the collection, $qtf_t$ is the within-query term frequency, b and $k_1$ are two tuning parameters. BM25 does not apply document length information well, so we use TFF in Equation 1 with the IDF of BM25 cross multiply, which makes up a new model named MATFM.

$$MATFM = \sum_{t\in q} TFF(t,D) \times \log(\frac{N}{df_t}) \quad (3)$$

Considering that three kinds of queries are provided in this task where the average length of a summary, a description and a note is 34.2, 123.4, and 309.8 respectively, so MATFM model choose to use the query length to adjust.

## 2.2 MATFB Model

In a given document, term frequency TF is usually normalized and inverse document frequency (IDF) is regarded as a measure to prevent a system to prefer longer document. The standard IDF method [4] is as follows:

$$IDF(t,C) = \log(\frac{N+1}{DF(t,C)}) \tag{4}$$

It considers only the presence or absence of a term in a document and does not take into account the specific term occurrence for a document. If two terms have the same document frequencies, the term discrimination will increase with the increase in average elite set term frequency. In MATF model, the average elite set term frequency (AEF) is defined as $\frac{CTF(t,C)}{DF(t,C)}$, where $CTF(t,C)$ denotes the total occurrence of the term t in the collection. The final term discrimination value can be calculated as follows:

$$TDF(t,C) = IDF(t,C) \times \frac{AEF(t,C)}{1+AEF(t,C)} \tag{5}$$

We used TF of Equation 2 and TDF of Equation 5 cross multiplication, making up MATFB model.

$$MATFB = \sum_{t \in q} TDF(t,C) \times \frac{(k_1+1)\cdot tf_{td}}{k_1 \cdot (1-b+b\cdot(\frac{L_d}{L_{avg}})) + tf_{td}} \tag{6}$$

This method reflects a good result in longer queries, such as description and note.

## 2.3 NEWBM Model

Due to lots of medical special terms appeared in documents, three kinds of queries represent different lengths. We further combine the two factors together (i.e., query length and specific term occurrence) by a linear way as follows:

$$NEWBM = p * \sum_{t \in q} TFF(t,D) \times \log(\frac{N}{df_t}) + (1-P) * \sum_{r \in q} TDF(t,C) \times \frac{(k_1+1)\cdot tf_{td}}{k_1 \cdot (1-b+b\cdot(\frac{L_d}{L_{avg}})) + tf_{td}} \tag{7}$$

# 3. Experimental Results

Our system was developed on the Lucene platform. This year’s theme is from three different types i.e., diagnosis, test, and treatment, and the example is introduced as follows:

```
<topic number="25" type="treatment">
  <note>
A 64 yo F w/PMHx sx for AF, COPD, HTN, hyperlipidemia who i
an open ASD repair c/b sternal wound infection and post-ope
[**11-15**] treated with amiodarone. On [**2-20**], she was
through the ED with SOB and back pain, and was noted to hav
fibrillation with RVR. A CTA demonstrating diffuse LAD and
post-obstructive PNA concerning for malignancy. For her atr
fibrillation, she was started on diltiazem gtt, for which s
transferred to the [**Hospital Unit Name 42**] for monitori
thought to be in the setting of a post-obstructive pneumoni
she was treated with antibiotics. She was then transferred
later that same night on metoprolol 50 mg tid. While on the
had a bronchoscopy performed which showed external compress
left mainstem bronchus, and she had a biopsy/FNA performed,
showed large cell carcinoma. She was then readmitted to the
yesterday with atrial fibrillation with HR 130s, and was st
diltiazem gtt.
.
In the [**Hospital Unit Name 42**], she was started on po c
uptitrated to 60 mg qid. She was called out this morning. T
8:30 pm, she was noted to have HR 160s, w/EKG c/w AF with F
which she received metoprolol 5 mg IV x2, followed by dilti
IV x2 without conversion. She denies chest pain, SOB, tachy
does note some diaphoresis and occasional palpitations.
  </note>
  <description>A 64 yo female with with history of atrial fi
  <summary>An elderly female with history of atrial fibrilla
</topic>
```

We eventually submitted five runs for the clinical decision support track in 2016. It includes a summary (for task A), a description (for task B), and three runs of notes (for task C). Table 1 shows the details of the five runs.

Table 1. Details of our system's settings for each run

| System | Task | Model | Topic Fields |
|---|---|---|---|
| CCNU_SUM_R1 | A | $MATFM$ | summary |
| CCNU_DES_R2 | B | $MATFB_{b=0.5}$ | description |
| CCNU_NOTE_R1 | C | $MATFB_{b=0.5}$ | note |
| CCNU_NOTE_R2 | C | $NEWBM_{p=0.5}$ | note |
| CCNU_ NOTE_R3 | C | $MATFM$ | note |

Table 2 shows the InfAP, InfNDCG, R-prec and P @10 of our runs for task A. The last two lines of the table show the best and median performance per topic averaged over all 30 topics for all of the runs in the track. Our run of summary is slightly better than the median. For some queries such as topics 22 and 28, our system performed very well.

Table 2. InfAP, InfNDCG, R-prec and P @10 of our runs for task A over 30 Topics

| System | InfAP | InfNDCG | R-prec | P@10 |
|---|---|---|---|---|
| CCNU_SUM_R1 | 0.0316 | 0.2179 | 0.1502 | 0.2933 |
| top system | 0.0869 | 0.4376 | 0.2553 | 0.6300 |
| median | 0.0196 | 0.1858 | 0.1219 | 0.2633 |

Table 3 shows the InfAP, InfNDCG, R-prec and P@10 of our runs for task B. Our run of description is slightly better than the median. For some queries such as topic 27, our system performed very well.

Table3. InfAP, InfNDCG, R-prec and P @10 of our runs for task B over 30 Topics

| System | InfAP | InfNDCG | R-prec | P@10 |
|---|---|---|---|---|
| CCNU_SUM_R1 | 0.0104 | 0.1047 | 0.0752 | 0.1533 |
| Top system | 0.0397 | 0.2750 | 0.1859 | 0.4600 |
| median | 0.0064 | 0.1005 | 0.0648 | 0.1533 |

Table 4 shows the InfAP, InfNDCG, R-prec and P@10 of the our runs for task C. We provided three runs for task C, and CCNU_NOTE_R1 is the best one of three runs, the performance of CCNU_NOTE_R2 and CCNU_NOTE_R3 are close to the median.

Table4. InfAP, InfNDCG, R-prec and P @10 of our runs for task C over 30 Topics

| System | InfAP | InfNDCG | R-prec | P@10 |
|---|---|---|---|---|
| CCNU_NOTE_R1 | 0.0140 | 0.1284 | 0.0894 | 0.1867 |
| CCNU_NOTE_R2 | 0.0084 | 0.0955 | 0.0651 | 0.1167 |
| CCNU_NOTE_R3 | 0.0098 | 0.1091 | 0.0789 | 0.1833 |
| Top system | 0.0599 | 0.3302 | 0.1993 | 0.51 |
| median | 0.0098 | 0.1227 | 0.0791 | 0.1833 |

## 4. Conclusion

In this paper, we present our system for clinical decision support track in TREC 2016. Since this is our first time to participate TREC, many further investigations and experiments are needed. In general, the result of summary is slightly better than that of description and note. Since each query often includes different levels of medical terms, more advanced query expansion technology will be further investigated in the future.


## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61402191), the Specific Funding for Education Science Research by Self-determined Research Funds of CCNU from the Colleges' Basic Research and Operation of MOE (No. CCNU16JYKX15), and the Thirteen Five-year Research Planning Project of National Language Committee (No. WT135-11).


## References


[1] Jiaul H. Paik. A Novel TF-IDF Weighting Scheme for Effective Ranking. SIGIR'13.

[2] Robertson, S.E., S. Walker. Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. SIGIR'94.

[3] Robertson S.E., S. Walker, S. Jones, M.M. Gatford. Okapi at TREC-3. TREC'94.

[4]Andrew T., Antti P., Blake B. Improvements to BM25 and Language Models Examined. ADCS'14.