SNUMedinfo at TREC CDS track 2014:

Medical case-based retrieval task

Sungbin Choi, Jinwook Choi

Medical Informatics Laboratory, Seoul National University, Seoul, Republic of Korea

wakeup06@empas.com, jinchoi@snu.ac.kr

Abstract. This paper describes the participation of the SNUMedinfo team at the TREC Clinical Decision Support track 2014. This task is about medical case-based retrieval. Case description is used as query text. Per each query, one of three categories (Diagnosis, Test and Treatment) is designated as target information need. Firstly, we used external tagged knowledge-based query expansion method for the relevance ranking. Secondly, machine learning classifier based text categorization method is used for the task-specific ranking. Finally, we combined relevance ranking and task-specific ranking with Borda-fuse method. Our method showed significant performance improvements.

Keywords: Case-based retrieval, Query expansion, Text categorization, Information retrieval

1. Introduction

In this paper, we describe the methods in participation of the SNUMedinfo team at the TREC Clinical Decision Support (CDS) track 2014. The task is about medical case-based retrieval task. Case description is used as query text. Per each query, one of three category (Diagnosis, Test and Treatment) is designated as target information need. For detailed task introduction, please see the overview paper of this track.

2. Methods

Our method can be summarized as following three steps (Section 2.1 to 2.3)

2.1 External tagged knowledge-based query expansion

We used external medical literature corpus (MEDLINE®) as a tagged knowledge source to acquire useful query expansion terms. We leased the 2014 MEDLINE® /PubMed® Journal Citations from the U.S. National Library of Medicine. There are approximately 22 million MEDLINE citations. Article title, abstract text, MeSH descriptor fields are indexed.

We used the unigram query likelihood (QL) model [1] with Dirichlet prior smoothing [2] as our baseline retrieval model. The Indri search engine [3] was used in the experiment. The queries are stopped at the query time using the standard 418 INQUERY stopword list, case-folded, and stemmed using Porter stemmer.

Per each original case query, we retrieved relevant documents from external corpus (MEDLINE) using query likelihood model. We extracted MeSH MajorTopic descriptors from top-k ranked documents. The original case query is expanded with these MeSH MajorTopic terms. Using this expanded query, we retrieved 1,000 documents per each query from target corpus (TREC CDS track). The Indri query is described as follows.

```
#weight ( (1-w) #combine (original query terms)
          w #combine (expansion query terms) )
```

Similar method showed effective performance in our previous study[1] [4] (ImageCLEF case-based retrieval task 2013' [5]).

2.2 Task-specific ranking

Per each query, one of three category (Diagnosis, Test and Treatment) is designated as target information need. We trained task classifiers on the Clinical Hedges database [6] and applied them on the top 1,000 documents from Section 2.1 to have task-specific ranking.

In Clinical Hedges database, documents are manually classified by purpose category (e.g., therapy, diagnosis, prognosis). We trained two task classifiers; CHD_TR_Classifier is trained to classify 'therapy' versus non-'therapy' documents. CHD_DX_Classifier is trained to classify 'diagnosis' versus non-'diagnosis' documents. SVM-perf [7] is used for the classification task. Both classifiers are trained to optimize AUC (area under the ROC curve).

Trained classifiers are applied on the top 1,000 documents from Section 2.1. Then, documents are sorted by classification score.

2.3 Combining relevance ranking with task-specific ranking

We combined relevance ranking and task-specific ranking with Borda-fuse method [8]. When different aspects need to be considered together for the document ranking, Borda-fuse method showed effective performance in our previous experiment [9].

---

[1] Compared to our method used in the ImageCLEF 2013' case-based retrieval task, this time we didn't apply limitation on the publication type of pseudo-relevant documents. We found out that it is not helpful to improve performance in our additional experiments on the ImageCLEF 2013 test set.

2.4 Submitted runs

Details of our submitted runs can be summarized as following table.

Table 1. Submitted runs

RunID	Query Version	Details per Query type
SNUMedinfo1	Summary	**Diagnosis** : Borda-fuse (Relevance ranking + CHD_DX_Classifier) **Test** : Borda-fuse (Relevance ranking + CHD_DX_Classifier) **Treatment** : Borda-fuse (Relevance ranking + CHD_TR_Classifier)
SNUMedinfo2	Summary	**Diagnosis** : Borda-fuse (Relevance ranking + rank_min[2](CHD_DX_Classifier, CHD_TR_Classifier)) **Test** : Borda-fuse (Relevance ranking + CHD_DX_Classifier) **Treatment** : Borda-fuse (Relevance ranking + CHD_TR_Classifier)
SNUMedinfo3	Summary	**Diagnosis** : Relevance ranking only **Test** : Borda-fuse (Relevance ranking + CHD_DX_Classifier) **Treatment** : Borda-fuse (Relevance ranking + CHD_TR_Classifier)

[2] For example, if Document A is ranked 10th by CHD_DX_Classifier, and ranked 800th by CHD_TR_Classifier, then output of rank_min for Document A is 10. If Document B is ranked 900th by CHD_DX_Classifier, and ranked 100th by CHD_TR_Classifier, then output of rank_min for Document B is 100.

SNUMedinfo4	Description	**Diagnosis** : Borda-fuse (Relevance ranking + CHD_DX_Classifier) **Test** : Borda-fuse (Relevance ranking + CHD_DX_Classifier) **Treatment** : Borda-fuse (Relevance ranking + CHD_TR_Classifier)
SNUMedinfo5 (No submit)	Description	**Diagnosis** : Borda-fuse (Relevance ranking + rank_min(CHD_DX_Classifier, CHD_TR_Classifier)) **Test** : Borda-fuse (Relevance ranking + CHD_DX_Classifier) **Treatment** : Borda-fuse (Relevance ranking + CHD_TR_Classifier)
SNUMedinfo6	Description	**Diagnosis** : Relevance ranking only **Test** : Borda-fuse (Relevance ranking + CHD_DX_Classifier) **Treatment** : Borda-fuse (Relevance ranking + CHD_TR_Classifier)

Query type *Treatment* is considered to be match with CHD_TR_Classifier. Query type *Test* is considered to be match with CHD_DX_Classifier.

With regard to the query type *Diagnosis*, by definition it is considered equivalent to the query types used in ImageCLEF case-based retrieval task [5], and that's why we applied only relevance ranking in SNUMedinfo3, SNUMedinfo6. But on the other hand, we thought that also it could be helpful to combine other task-specific ranking with relevance ranking, because *Test* or *Treatment* tasks are closely related to the *Diagnosis* task.

3. Results

Table 2. Evaluation results (query version: Summary)

RunID	infNDCG	infAP	P@10
Baseline (QL)	0.1921	0.0501	0.3400
ExternalQE	0.2224	**0.0589**	0.3200
SNUMedinfo1	0.2188	0.0463	0.3367
SNUMedinfo2	0.2173	0.0458	0.3333
SNUMedinfo3	**0.2406**	0.0582	**0.3467**

QL : Query likelihood model with original query
ExternalQE : External tagged knowledge based query expansion
Best result per column is marked in boldface

Table 3. Evaluation results (query version: Description)

RunID	infNDCG	infAP	P@10
Baseline (QL)	0.1877	0.0436	0.2933
ExternalQE	0.2199	0.0511	0.3200
SNUMedinfo4	0.2502	0.0545	0.3300
SNUMedinfo5	0.2505	0.0556	0.3267
SNUMedinfo6	**0.2674**	**0.0659**	**0.3633**

QL : Query likelihood model with original query
ExternalQE : External tagged knowledge based query expansion
Best result per column is marked in boldface

In Table 2, SNUMedinfo3 showed significant performance improvement over baseline. In Table 3, SNUMedinfo6 showed significant performance improvement over baseline. Both SNUMedinfo3 and SNUMedinfo6 used relevance ranking only for the query type Diagnosis, while Borda-fuse of relevance ranking and task-specific ranking is used for the Test and Treatment query type.

4. Discussion

In Table 4 and Table 5, we compared evaluation results of different methods per query type.

Table 4. Comparison of results per query type
(query version: Summary, evaluation metric : infNDCG)

	Diagnosis	Test	Treatment	Total
Baseline	0.2263	0.1515	0.1984	0.1921
ExternalQE	0.2945	0.1546	0.2182	0.2224
SNUMedinfo3	0.2945	0.1831	0.2443	0.2406

ExternalQE : External tagged knowledge based query expansion

Table 5. Comparison of results per query type (query version: Description, evaluation metric : infNDCG)

	Diagnosis	Test	Treatment	Total
Baseline	0.2270	0.1558	0.1804	0.1877
ExternalQE	0.2977	0.1346	0.2273	0.2199
SNUMedinfo6	0.2977	0.2029	0.3016	0.2674

ExternalQE : External tagged knowledge based query expansion

5. Conclusion

TREC CDS 2014 was a medical case-based retrieval task, and each query had different target task among diagnosis, test or treatment. As a first step, we used external tagged knowledge based query expansion method to retrieve relevant documents. As a second step, we trained machine learning document classifier to compute task-specific ranking of documents. Finally, we combined relevance ranking and task-specific ranking with Borda-fuse method. Our method showed significant improvement over baseline method.

6. Acknowledgements

Use of the Clinical Hedges database was made possible through a collaboration agreement with R B Haynes and N L Wilczynski at McMaster University, Hamilton, Ontario Canada. This study was supported by a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea. (No. HI11C1947)

7. References

1. Ponte, J.M. and W.B. Croft, *A language modeling approach to information retrieval*, in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. 1998, ACM: Melbourne, Australia. p. 275-281.
2. Zhai, C. and J. Lafferty, *A study of smoothing methods for language models applied to Ad Hoc information retrieval*, in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. 2001, ACM: New Orleans, Louisiana, USA. p. 334-342.
3. Strohman, T., et al. *Indri: A language model-based search engine for complex queries*. in *Proceedings of the International Conference on Intelligent Analysis*. 2005. McLean, VA.

4. Choi, S., J. Lee, and J. Choi. *SNUMedinfo at ImageCLEF 2013: Medical retrieval task*. in *CLEF (Online Working Notes/Labs/Workshop)*. 2013.
5. Garcıa Seco de Herrera, A., et al. *Overview of the ImageCLEF 2013 medical tasks*. Working Notes for CLEF 2013 Conference, 2013. **1179**.
6. Haynes, R.B., et al., *Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey*. Vol. 330. 2005. 1179.
7. Joachims, T., *Training linear SVMs in linear time*, in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2006, ACM: Philadelphia, PA, USA. p. 217-226.
8. Aslam, J.A. and M. Montague, *Models for metasearch*, in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. 2001, ACM: New Orleans, Louisiana, USA. p. 276-284.
9. Choi, S., et al., *Combining relevancy and methodological quality into a single ranking for evidence-based medicine.* Information Sciences, 2012. **214**(0): p. 76-90.