NCBI at the 2015 BioASQ challenge task: Baseline results from MeSH Now

Yuqing Mao[1], Zhiyong Lu[1]

[1]National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM)
8600 Rockville Pike, Bethesda, MD 20894, USA

{yuqing.mao, zhiyong.lu}@nih.gov

Abstract. During the 2015 BioASQ challenge, we contributed our method—MeSH Now—as a baseline system by making its prediction results immediately available to all participating teams throughout the task. By doing so, we make it possible for others to build on our award-winning system for further advancement in biomedical literature Indexing. First developed in 2014, MeSH Now is a state-of-the-art system that systematically integrates different indexing approaches via its automatic learning-to-rank framework. To serve as a baseline and maximize its potential in the challenge, we provided MeSH Now results in two separate settings: one favors high F-score and the other Recall. Experimental results show that MeSH Now compares favorably to the other baseline approaches by achieving consistently over 0.60 in F-score and 0.85 in Recall, respectively. Furthermore, MeSH Now is implemented on computer clusters so that it can provide real-time results for the challenge. To conclude, MeSH Now is a competitive and scalable system for indexing biomedical literature.
Availability: http://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/MeSHNow/

Keywords: MeSH; Literature Indexing; Text Categorization; Learning-to-rank

1 Introduction

In recent years, there has been a rapid growth of scholarly publications in biomedicine. Thus finding relevant information is becoming increasingly difficult, even for specialists in this area [1]. To facilitate literature search in PubMed, articles are manually indexed with a set of relevant and controlled keywords known as Medical Subject Headings (MeSH) terms. MeSH indexing is the task of assigning relevant MeSH terms based on a manual reading of scholarly publications by human indexers. This task is highly important for improving literature retrieval and many other scientific investigations in biomedical research [2]. However, given its manual nature, the process of MeSH indexing is extremely time-consuming and costly. It is reported that on average, it costs $9.40 and takes 2 to 3 months for a new article to be indexed upon entering PubMed [3]. To improve productivity and assist human indexers [4], automated MeSH indexing been proposed but several key issues remain including both reliability and scalability [5-12] (see [13] for a brief survey on the past work).

BioASQ[1] [8, 14] is one of the recent community-wide challenge events in BioNLP research area [15]. BioASQ 2015 is their third year focusing on the tasks of large-scale literature indexing (3a) and question answering (3b). We participated in Task 3a this year. In this task, participating teams were provided with a set of newly published articles in PubMed, and were asked to automatically predict the most relevant MeSH terms for each article. During evaluation, text-mined results were compared with the human indexed MeSH terms (known as gold standard).

A brief description of our method for task 3a is presented in Section 2. In Section 3 we show the results of our method on the official BioASQ test datasets, followed by a discussion of the results and our conclusion remarks for the 2015 challenge.

2 Methods

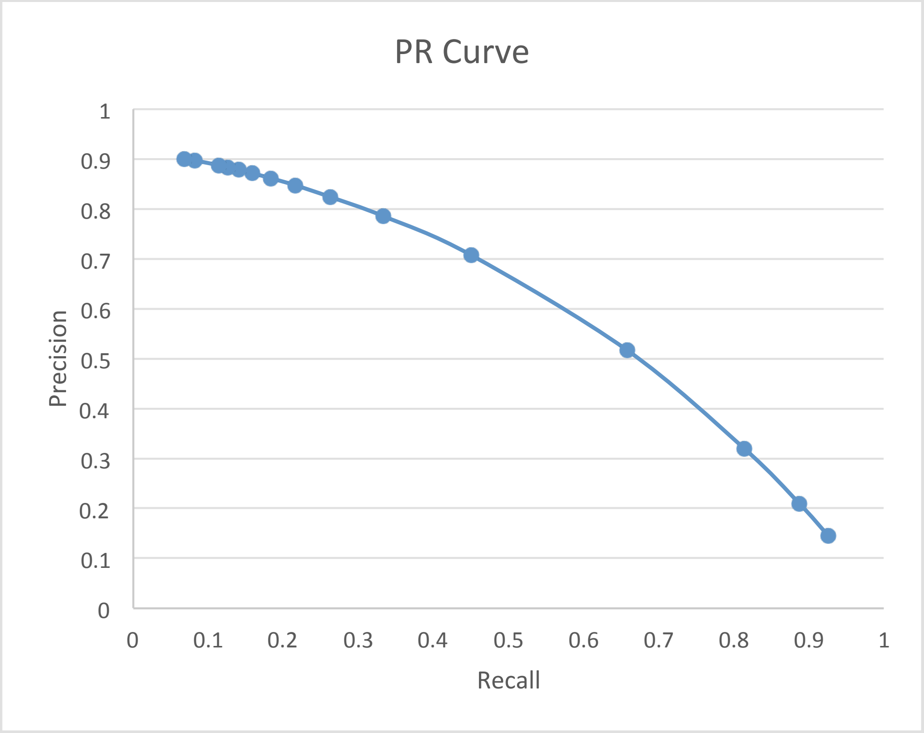
For the literature-indexing task in BioASQ 2015 (Task 3a), we used MeSH Now, an award-winning system we first built when we participated in the same task in 2014[2] [8]. Given a target article, MeSH Now operates in three main steps. First, it obtains a list of candidate MeSH terms from multiple sources/approaches (e.g. previously indexed MeSH terms from related articles). Next, it combines these different inputs systematically via a novel machine-learning framework to rank the candidate terms based on their relevance to the target article. Finally, it selects and returns the highest-ranked MeSH terms for the target article. We refer interested readers to [13] for a full description of MeSH Now.

To serve as baselines in the 2015 challenge, we made several additional updates and customizations: First, we updated our lexicon with MeSH 2015. Second, we updated training documents according to the select BioASQ journals and used a newer set of documents for training our machine-learning model. Third, each week we submitted two baseline runs, namely "MeSH Now BF" and "MeSH Now HR" where the former favors high f-score and the latter recall, respectively. In particular, for the recall-favoring run, we always returned top 100 predicted MeSH terms. According to the precision-recall curve in Figure 1, we can expect our recall to be nearly 90% when returning the top 100 predictions.

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

[2] http://www.bioasq.org/participate/second-challenge-winners

Fig. 1. Precison-Recall Curve of MeSH Now on the BioASQ5000 dataset [13], which consists of 5,000 PubMed documents randomly selected from the BioASQ 2014 test sets



Finally, in order to allow task participants to have our predictions results (both runs) at the earliest time possible every week, we used NCBI's computer cluster to run MeSH Now in parallel so that the response time can be greatly improved. As a result, MeSH Now is able to process individual documents instantly. For processing 3,000-5,000 articles (typical size for each batch in the BioASQ challenge), it takes approximately one hour depending on the concurrent jobs on our computer cluster.

3 Results

The 2015 BioASQ Task 3a was organized for three consecutive periods (batches) of 5 weeks each. Each week, the task organizers distributed new PubMed articles and participants were given a limited response time (less than 24 hours) to submit their computer-predicted MeSH terms.

In Task 3a, the performance of the participating systems was assessed based on two primary measures: one is the flat measure "label-based micro F-measure" and the other the hierarchical measure "Lowest Common Ancestor F-measure (LCA-F)". Below we present our results on the BioASQ Task 3a Batch 2 Week 5. This dataset contains 4,059 articles in total, of which 2,649 articles are with human indexing results as of June 17, 2015.

As shown in table 2, the submitted system "MeSH Now BF" outperformed all other baselines in both flat and hierarchical F-measures, while the choice of top 100 MeSH terms in "MeSH Now HR" resulted in the highest performance in recall. We also note that "MeSH Now BF" consistently achieved around 0.60 in F-score, suggesting that MeSH Now is highly robust on different datasets.

Table 2. Official results for our results on Batch 5 Week 2 test set, compared with three other baseline methods. Our best results among all submissions are highlighed in bold.

Systems	MiF	MiP	MiR	LCA-F	LCA-P	LCA-R
MeSH Now BF	**0.6010**	0.6117	0.5907	**0.4978**	0.5241	0.5086
Default MTI[7]	0.5849	0.5879	0.5819	0.4881	0.5139	0.4987
MTI First Line[3]	0.5821	0.6350	0.5373	0.4812	0.5428	0.4637
BIoASQ_Baseline [14]	0.2647	0.2296	0.3125	0.3027	0.5155	0.3356
MeSH Now HR	0.2131	0.1217	**0.8583**	0.2447	0.1533	**0.6698**

4 Discussion & Conclusion

By making MeSH Now as a baseline, we contributed to the BioASQ 2015 challenge in a new supporting role. During this process, MeSH Now was further improved and streamlined to meet the needs of real-time processing. As a robust framework, MeSH Now showed competitive performance during the BioASQ 2015 evaluations. Given its performance and scalability, we hope that other teams found it useful during the challenge. In the future, we plan to integrate MeSH Now as part of our interactive tool PubTator [4, 16] as well as to explore its other applications in practice.

Acknowledgements

We would like to thank the BioASQ task organizers for providing the task and baseline data. This research is supported by the NIH Intramural Research Program, National Library of Medicine.

[3] http://ii.nlm.nih.gov/MTI/MTIFL.shtml

References

1. Islamaj Dogan, R., Murray, G.C., Neveol, A., Lu, Z.: Understanding PubMed user search behavior through log analysis. Database : the journal of biological databases and curation 2009, bap018 (2009)

2. Lu, Z., Kim, W., Wilbur, W.J.: Evaluation of query expansion using MeSH in PubMed. Information retrieval 12, 69-80 (2009)

3. Huang, M., Névéol, A., Lu, Z.: Recommending MeSH terms for annotating biomedical articles. Journal of the American Medical Informatics Association 18, 660-667 (2011)

4. Wei, C.H., Harris, B.R., Li, D., Berardini, T.Z., Huala, E., Kao, H.Y., Lu, Z.: Accelerating literature curation with text-mining tools: a case study of using PubTator to curate genes in PubMed abstracts. Database : the journal of biological databases and curation 2012, bas041 (2012)

5. Liu, K., Wu, J., Peng, S., Zhai, C., Zhu, S.: The fudan-uiuc participation in the bioasq challenge task 2a: The antinomyra system. CLEF2014 Working Notes 129816, 100 (2014)

6. Mao, Y., Wei, C.-H., Lu, Z.: NCBI at the 2014 BioASQ challenge task: large-scale biomedical semantic indexing and question answering. Proceedings of Question Answering Lab at CLEF (2014)

7. Mork, J.G., Jimeno-Yepes, A., Aronson, A.R.: The NLM Medical Text Indexer System for Indexing Biomedical Literature. CLEF2013 Working Notes (2013)

8. Balikas, G., Partalas, I., Ngomo, A.-C.N., Krithara, A., Gaussier, E., Paliouras, G.: Results of the BioASQ Track of the Question Answering Lab at CLEF 2014. Results of the BioASQ Track of the Question Answering Lab at CLEF 2014, 1181-1193 (2014)

9. Huang, M., Lu, Z.: Learning to annotate scientific publications. Proceedings of the 23rd International Conference on Computational Linguistics: 463-471 (2010)

10. Trieschnigg, D., Pezik, P., Lee, V., de Jong, F., Kraaij, W., Rebholz-Schuhmann, D.: MeSH Up: effective MeSH text classification for improved document retrieval. Bioinformatics 25, 1412-1418 (2009)

11. Zhu, D., Li, D., Carterette, B., Liu, H.: An Incremental Approach for MEDLINE MeSH Indexing. BioASQ@ CLEF (2013)

12. Ruch, P.: Automatic assignment of biomedical categories: toward a generic approach. Bioinformatics 22, 658-664 (2006)

13. Mao, Y., Lu, Z.: MeSH Now: Automatic MeSH Indexing at PubMed Scale via Learning to Rank. Journal of Biomedical Semantics (2015)

14. Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M.R., Weissenborn, D., Krithara, A., Petridis, S., Polychronopoulos, D., Almirantis, Y., Pavlopoulos, J., Baskiotis, N., Gallinari, P., Artieres, T., Ngomo, A.C., Heino, N., Gaussier, E., Barrio-Alvers, L., Schroeder, M., Androutsopoulos, I., Paliouras, G.: An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition. BMC bioinformatics 16, 138 (2015)

15. Huang, C.C., Lu, Z.: Community challenges in biomedical text mining over 10 years: success, failure and the future. Briefings in bioinformatics (2015)

16. Wei, C.H., Kao, H.Y., Lu, Z.: PubTator: a web-based text mining tool for assisting biocuration. Nucleic acids research 41, W518-522 (2013)