An Empirical Study on Utilizing Neural Network for Event Information Retrieval

Minglei Li¹*, Yu Zhu² and Ren Wang¹

¹Institute 706, Second Academy of China Aerospace Science and Industry Corporation, Haidian Beijing 100854, China
²Unit 96901, the Chinese People’s Liberation Army, Haidian Beijing 100094, China
Email: liminglei15@mails.ucas.ac.cn; yzhu@home.ipe.ac.cn; thank816you@163.com

Abstract. Event information retrieval (EIR) is the task of retrieving news articles in response to the event-oriented query, which tailors for dealing with a large scale of news articles on breaking news (e.g., natural disaster). However, the challenges caused by the dynamic and complex natures of the event make EIR less studied than traditional information retrieval (IR). The existing studies usually employed either traditional IR models or machine learning approaches over hand-crafted features to retrieve events, which fails to capture more complex structures of events with event evolution well. To address this issue, we exploit neural models incorporating the characteristics of events to EIR task. Experimental results demonstrate the effectiveness of our method over two datasets, which can regard as the simple but strong baseline for further research in EIR.

Keywords. Event information retrieval; event-oriented; neural network.

1. Introduction

With the explosion of massive news articles generated from various news agencies, it is difficult for people to deal with such a great many news articles. EIR [1] is the task of retrieving news articles in response to an event query, which tailors for dealing with this information overload. Accurately retrieving news articles about a specific event will benefit many downstream applications, such as event monitoring, news recommendation, event tracking, etc.

However, the existing challenges make EIR task less studies than traditional IR. To see its challenge, we analyse the collection of news articles and conclude that the challenge is primarily caused by the dynamic and complex nature of events. Take the Malaysia Airlines Flight 370 disaster (“MH370”) as an example (see figure 1). Given the query “MH370 debris”, whose goal is to find news articles related to the query. By our analysis of the retrieval results, two of the most important ones are pointed out:

- event dynamics: the event consists of several subevents (e.g., “MH370 disappearance”, “Find MH370 debris”) with event evolution.
- structure complexity: a subevent usually associates to multiple new articles and a news article may correspond to multiple subevents in different periods.

A few of existing studies [2-4] dealing with the issue of EIR involve two solutions. One solution is to optimize the traditional retrieval model. For instance, Catena et al. [4] presented a BM25P model. This model computes the score of query results by a linear combination of term statistics in terms of
different portions of articles. Another solution is mining the features of event using machine learning approaches. For example, Setty et al. [3] presented the novel event mining and feature generation methods for ranking news events. These solutions employ either traditional information retrieval models or machine learning approaches over hand-crafted features to retrieval events, which fails to capture more-complex structures of events with event evolution well.

Fortunately, neural models such as neural ranking models (NRMs) [5, 6] and pre-trained neural language models (e.g., BERT [7, 8]) have achieved promising results in ad-hoc retrieval task. However, there is not yet an end-to-end approach for EIR task. Motivated by this, we investigate whether neural models can be exploited to improve the performance of EIR. To our best knowledge, our work is the first contribution to employ neural models to EIR by incorporating the characteristics of events. Experimental results demonstrate the better performance using our method over two real-world datasets, which can regard as the simple but strong baseline for further research.

Figure 1. An example of the Malaysia Airlines Flight 370 for EIR.

2. Approach

2.1. Neural Models

Neural ranking model is the heart of the retrieval system, which determines the quality of the retrieval system definitively. With the advance of deep learning, several studies have leveraged shallow or deep neural networks to the ranking problem effectively in IR, such as [8, 9]. Depending on the different application scenarios, neural models can be categorized into two groups, including models (i.e., DSSM [10], ArcII [11], MatchPyramid [12]) for text matching and models (i.e., DRMM [13], DRMMTKS [14], DUET [15], KNRM [16], Conv-KNRM [5], BERT [8]) for ad-hoc retrieval. However, these models above-mentioned do not consider IR specificities, such as query term importance, exact matching signals for text matching. As such, we only employ models developed for ad-hoc retrieval to EIR task.

2.2. EIR by Neural Models

The input text is usually short and does not involve the complex structure in traditional tasks, such as ad-hoc retrieval, which can be input to neural models directly. By contrast, news articles are usually
long texts and involve more-complex structures. If we truncate a fix-length text from the beginning of the document as input, it will fail to capture the sufficient semantics of events. An alternative is employing query expansion and news summarization to deal with event dynamics and structure complexity, respectively.

Previous studies have shown that external resources are effective for query expansion. Therefore, we use events collected from Wikipedia worldwide current events as external resources to expand the original query. Concretely, this process involves three key steps: (1) crawl the relevant Wikipedia pages for each event with current events portal, (2) obtain candidate terms of each event from these Wikipedia pages with RAKE [17], and (3) select the top three candidates to expand the original query. In addition, we assume that an event can be viewed as an abstract representation of long news articles. So, we use the abstractive news summarization approach [18] to extract the event. Finally, we feed the extensions of query and the refined news article into neural models for EIR.

3. Experiments and Results

We perform the following experiments aim to explore (1) whether neural models are more suitable for EIR task than traditional approaches, and (2) whether our approach using query expansion and news summarization to neural models achieves a better overall ranking quality than without using our approach.

3.1. Setup

We use two publicly available datasets for evaluation in EIR task, including JustEvents [19] and Event-Dataset [20]. The JustEvents dataset consists of 250 events spanned from 18th January 2011 to 7th February 2011, which is collected from more than 1.8 million Wikipedia pages. The Event-Dataset contains 35 past and future events occurred during the years of 1997 to 2022. These events are well reported all over the world, such as “MH370 Disappearance”, “Kashmir Earthquake”. The detailed information of datasets is displayed on table 1.

Table 1. The description of JustEvents and Event-Dataset. Avg #qLen denotes the average length of queries. Avg #dLen denotes the average length of documents.

| Dataset        | #Query | #Doc  | Avg #qLen | Avg #dLen |
|----------------|--------|-------|-----------|-----------|
| JustEvents     | 250    | 6,457 | 45.92     | 5160.86   |
| Event-Dataset  | 35     | 2,932 | 19.54     | 4918.99   |

We evaluate the effectiveness of the approach based on two types of approaches. One is traditional models, including the Okapi BM25 [21] and BM25P. Another is neural models that we have mentioned in Section 2.1, including DRMM, DRMMTKS, DUET, KNRM, Conv-KNRM and BERT.

Considering small datasets, we conduct 10-fold cross validation to minimize the problem of over-fitting. We use MatchZoo toolkit to train these NRMs. We use the cross entropy [22] loss function to rank. In order to evaluate the performance of the proposed approach for EIR in this paper, we use three standard metrics for comparison: MAP [23], nDCG@5 [24], and nDCG@10. In addition, statistical differences between models are computed using the Fisher randomization test [25] (α = 0.05).

3.2. Results Analysis

A summary of performance results of traditional models (i.e., BM25, BM25P) and neural models over two datasets is displayed in table 2. As we can see from the JustEvents dataset, all the neural models perform significantly better than traditional models. For the Event-Dataset dataset, we find that all neural models cannot compete with traditional models. A possible reason is that the Event-Dataset dataset is too small (only 35 queries) to take effect on neural models for improving performance. Considering this issue, we only use the JustEvents dataset to evaluate our approach. The experiment result is reported in table 3. Our approach employing query expansion and news summarization achieves significant performances in all neural models. According to this experiment, we investigate
that neural models incorporating the characteristics of events can be exploited to EIR in terms of the effectiveness under the scenario of a certain scale of data.

**Table 2.** Comparison of traditional models (i.e., BM25, BM25P) with neural models over the JustEvents and Event-Dataset. Statistical significant improvements or degradations w.r.t. BM25 and BM25P are indicated △/▽ and ▲/▼ (p-value <=0.05) respectively.

| Model Name | JustEvents | Event-Dataset |
|------------|-------------|---------------|
|             | MAP  | nDCG@5  | nDCG@10 | MAP  | nDCG@5  | nDCG@10 |
| BM25        | 0.1145 | 0.0008  | 0.0013  | 0.2958 | 0.3073  | 0.3152  |
| BM25P       | 0.1154 | 0.0139  | 0.0149  | 0.3007 | 0.3452  | 0.4355  |
| DRMM        | 0.2068▲ | 0.1470▲ | 0.1987▲ | 0.2805▼ | 0.2272▼ | 0.2631▼ |
| DRMMTKS     | 0.2194▲ | 0.1255▲ | 0.1783▲ | 0.2678▲ | 0.1958▼ | 0.2382▼ |
| DUET        | 0.2316▲ | 0.1655▲ | 0.1792▲ | 0.2752▼ | 0.1944▼ | 0.2281▼ |
| KNRM        | 0.2324▲ | 0.1569▲ | 0.2176▲ | 0.2779▲ | 0.2324▼ | 0.2524▼ |
| Conv-KNRM   | 0.3261▲ | 0.2692▲ | 0.2853▲ | 0.2679▲ | 0.2131▼ | 0.2363▼ |
| BERT        | 0.3678▲ | 0.2893▲ | 0.3014▲ | 0.2944▼ | 0.2309▼ | 0.2909▼ |

**Table 3.** The performance of different models using our method on JustEvents. Statistical significant improvement or degradation w.r.t. corresponding models is indicated (+/-) (p-value <=0.05) respectively.

| Model Name  | DRMM  | DRMMTKS  | DUET | KNRM  | Conv-KNRM | BERT  |
|-------------|-------|----------|------|-------|-----------|-------|
| MAP         | 0.2416⁺ | 0.264⁰   | 0.243¹ | 0.290⁸ | 0.344⁷⁺  | 0.368⁹⁺ |
| nDCG@5      | 0.173⁹⁺ | 0.25²⁰⁺  | 0.21⁴³⁺ | 0.23⁶⁰⁺ | 0.3⁰²⁷⁺  | 0.3⁰³⁴⁺ |
| nDCG@10     | 0.2²³¹⁺ | 0.2⁴⁰⁰⁺  | 0.2⁵³⁹⁺ | 0.3⁰¹²⁺ | 0.3⁴³¹⁺  | 0.3⁷⁰¹⁺ |

**4. Conclusions**

In this paper, we employ neural models incorporating the dynamics and complex nature of events to EIR task. First, we analyse the challenge of EIR and conclude that the challenge is primarily caused by the dynamic and complex nature of events. Then we investigate whether neural models can be exploited to EIR by incorporating these characteristics of events. Finally, the experimental results with our approach outperform other the state-of-the-art approaches over real-world datasets in terms of effectiveness, which can regard as the simple but strong baseline for further research in EIR. As future work we plan to design a neural model tailored for EIR, which can avoid the pipeline way of processing query and news articles.

**References**

[1] Sanekpally R 2019 Event information retrieval from text *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA) p 1447.

[2] Metzler D, Cai C and Hovy E H 2012 Structured event retrieval over microblog archives *HLT-NAACL*.

[3] Setty V, Anand A, Mishra A and Anand A 2017 Modeling event importance for ranking daily news events *WSDM 2017 - Proceedings of the 10th ACM International Conference on Web Search and Data Mining* pp 231-240.

[4] Catena M, Frieder O, Muntean C I, Nardini F M, Perego R and Tonellotto N 2019 Enhanced news retrieval: Passages LEAD THE WAY! *SIGIR’19*. 

4
[5] Dai Z, Xiong C, Callan J and Liu Z 2018 Convolutional neural networks for soft-matching n-grams in ad-hoc search *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* pp 126-134.

[6] Hui K, Yates A, Berberich K and De Melo G 2018 Co-PACRR: A context-aware neural IR model for ad-hoc retrieval *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* pp 279-287.

[7] Devlin J, Chang M-W, Lee K and Toutanova K 2019 BERT: Pre-training of deep bidirectional transformers for language understanding *NAACL-HLT*.

[8] Dai Z and Callan J 2019 Deeper text understanding for IR with contextual neural language Modeling *SIGIR ’19*.

[9] Guo J, et al. 2019 deep look into neural ranking models for information retrieval *arXiv preprint arXiv:1903.06902*.

[10] Huang P-S, He X, Gao J, Deng L, Acero A and Heck L 2013 Learning deep structured semantic models for web search using clickthrough data *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* pp 2333-2338.

[11] Hu B, Lu Z, Li H and Chen Q 2014 Convolutional neural network architectures for matching natural language sentences *Advances in neural information processing systems* pp 2042-2050.

[12] Pang L, Lan Y, Guo J, Xu J, Wan S and Cheng X 2016 Text matching as image recognition *Thirtyith AAAI Conference on Artificial Intelligence*.

[13] Guo J, Fan Y, Ai Q and Croft W B 2016 A deep relevance matching model for ad-hoc retrieval *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* pp 55-64.

[14] Yang Z, et al. 2018 Deep top-K relevance matching model for ad-hoc retrieval *CCIR*.

[15] Mitra B, Diaz F and Craswell N 2016 Learning to match using local and distributed representations of text for web search *WWW*.

[16] Xiong C, Dai Z, Callan J P, Liu Z and Power R 2017 End-to-End neural ad-hoc ranking.

[17] Rose S J, Engel D, Cramer N and Cowley W 2010 Automatic keyword extraction from individual documents.

[18] Li W, He L and Zhuge H 2016 Abstractive news summarization based on event semantic link network *COLING*.

[19] Ceroni A, Gadiraju U and Fisichella M 2017 JustEvents: A crowdsourced corpus for event validation with strict temporal constraints *ECIR*.

[20] Khan S U R and Islam M A 2019 Event-Dataset: Temporal information retrieval and text classification dataset *Data in Brief*.

[21] Robertson S E and Walker S 1994 Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval *SIGIR ’94*.

[22] Thorsten J 2002 Optimizing search engines using clickthrough data *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pp 133-142.

[23] Robertson S E 2008 A new interpretation of average precision *SIGIR*.

[24] Kalervo J 2002 Cumulated gain-based evaluation of IR techniques.

[25] Smucker M D, Allan J and Carterette B 2007 A comparison of statistical significance tests for information retrieval evaluation *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management* pp 623-632.