Multi-level Memory Network with CRFs for Keyphrase Extraction

Tao Zhou · Yuxiang Zhang · Haoxiang Zhu

Civil Aviation University of China, Tianjin, China

The 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2017. Singapore.

中国民航大学
Civil Aviation University of China

Outline

- I. Task description
- II. Current methods
- III. Idea of our work
- IV. Our proposed model
- V. Experimental evaluation
- VI. Conclusion
- VII. Q&A

- Automatic **keyphrase extraction** extracts a set of ***representative phrases*** that are related to the ***main topics*** discussed in a document.

Example:

"**title**":" Enhanced **max margin** learning on **multimodal data mining** in a multimedia database"

"**text**":" The problem of **multimodal data mining** in a multimedia database can be addressed as a structured prediction problem where we learn the mapping from an input to the structured and interdependent output variables. In this paper, built upon the existing literature on the **max margin** based learning, we develop a new **max margin** learning approach called Enhanced **Max Margin** Learning -LRB- EMML -RRB- framework. In addition, we apply EMML framework to developing an effective and efficient solution to the **multimodal data mining** problem in a multimedia database…… ."

"**keyword**": "image annotation; image retrieval; **max margin**; **multimodal data mining**"

II. Current methods

- **Unsupervised approaches** directly treat keyphrase extraction as a ***ranking problem***, scoring each candidate using different kinds of techniques such as language modeling, clustering or graph-based ranking. In particular, the **graph-based algorithms** are widely used in the unsupervised scenario.
  - TextRank(2004), TopicalPageRank(2010), CiteTextRank(2014), PositionRank(2017), Salience Rank(2017)

[1] Mihalcea R, et al. TextRank: bringing order into texts. EMNLP'2004.
[2] Liu ZY, et al. Automatic keyphrase extraction via topic decomposition. EMNLP'2010
[3] Gollapalli SD, et al. Extracting keyphrases from research papers using citation networks. AAAI'2014.
[4] Sterckx L, et al. Topical word importance for fast keyphrase extraction. WWW'2015.
[5] Florescu C, et al. A position-biased pagerank algorithm for keyphrase extraction. AAAI'2017.
[6] Nedelina T, et al. Salience Rank: Efficient Keyphrase Extraction with Topic Modeling. ACL'2017.

- The **supervised approaches** generally treat the keyphrase extraction as a binary classification task, in which a learning model is trained on the features of labelled keyphrases to determine whether a candidate phrase is a keyphrase.
  - Classifiers: Naïve Bayes, Genetic algorithm, Decision tree, Neural networks, Conditional random field, et al.
  - Different types of features: Statistical features, position features, semantic features, external resource-based features.

[1] Caragea C, et al. Citation-enhanced keyphrase extraction from research papers: a supervised approach. EMNLP'2014.
[2] Zhang Q, et al. Keyphrase extraction using deep recurrent neural networks on twitter. EMNLP'2016.
[3] Sujatha D G, et al. Incorporating Expert Knowledge into Keyphrase Extraction. AAAI'2017.
[4] Wang Q, et al. Keyphrase extraction with sequential pattern mining. AAAI'2017.
[5] Meng R, et al. Deep keyphrase generation. ACL'2017.
[6] Alzaidy, et al. Bi-LSTM-CRF sequence labeling for keyphrase extraction from scholarly documents. WWW' 2019.

III. Ideas of our work

- Each article has several different subtopics.
  - Some of them **run through the whole article**.
  - Some are **distributed in different positions of the article**. We need to combine long-distance text information to better understand the text content.
- Whether a word is a keyword or part of a keyword is not only related to the word itself, but also related to **the words around the current word**, and the **long-distance information**.

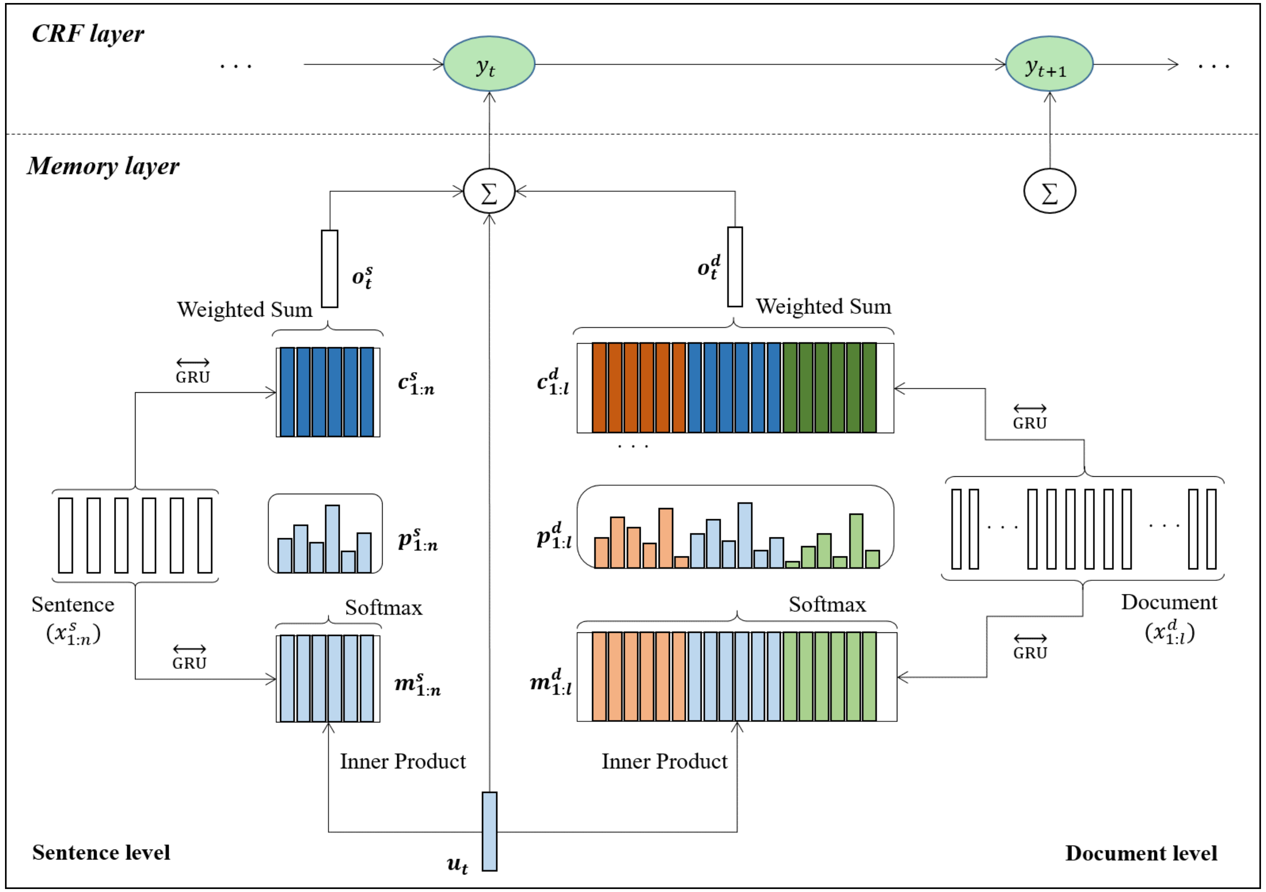
- For capture the **long-range contextual information** hidden in the text sequence, we use the ***memory network*** to enhance the long-term memory capability of deep network.

- We also use the ***CRF*** to capture **local dependencies** between adjacent words in text sequence.

In our model, we formulate keyphrase extraction as a task of sequence labeling. For the keyphrase: "**multimodal data mining**", the corresponding label is "**B-CK M-CK E-CK**".

Enhanced **max margin** learning on **multimodal data mining** in a multimedia database.
O B-CK E-CK O O B-CK M-CK E-CK O O O O

- **MLM-CRF**(Multi-level Memory Network with CRFs) firstly uses the memory network to enhance the text representation, then integrate the multi-level memory network with the CRFs, which has an advantage in modeling the local contextual information.



$$\overrightarrow{\boldsymbol{m}}_t = \overrightarrow{\mathbf{GRU}}(\boldsymbol{x}_t, \overrightarrow{\boldsymbol{m}}_{t-1}) \quad (1)$$

$$\overleftarrow{\boldsymbol{m}}_t = \overleftarrow{\mathbf{GRU}}(\boldsymbol{x}_t, \overleftarrow{\boldsymbol{m}}_{t+1}) \quad (2)$$

$$\boldsymbol{m}_t = \tanh(\overrightarrow{\boldsymbol{W}}_m \overrightarrow{\boldsymbol{m}}_t + \overleftarrow{\boldsymbol{W}}_m \overleftarrow{\boldsymbol{m}}_t + \boldsymbol{b}_m) \quad (3)$$

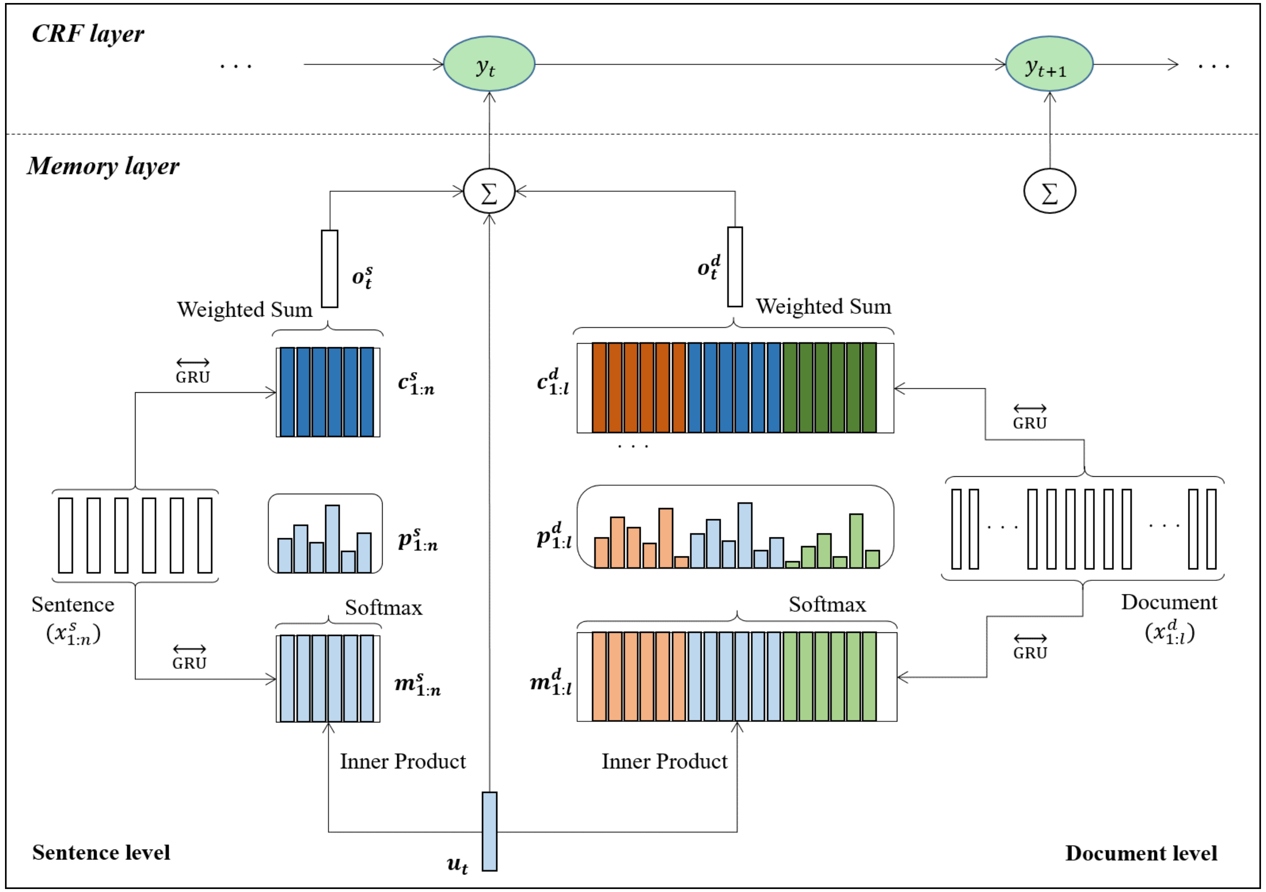
$$\boldsymbol{u}_t = \boldsymbol{m}_t. \quad (4)$$

$$p^s_{t,i} = \mathtt{softmax}(\boldsymbol{u}_t^T \boldsymbol{m}_i^s) \quad (5)$$

$$p^d_{t,i} = \mathtt{softmax}(\boldsymbol{u}_t^T \boldsymbol{m}_i^d) \quad (6)$$

$$\boldsymbol{o}_t = \lambda \sum_i p^d_{t,i} \boldsymbol{c}^d_i + (1-\lambda) \sum_i p^s_{t,i} \boldsymbol{c}^s_i \quad (7)$$

- **MLM-CRF**(Multi-level Memory Network with CRFs) secondly uses the CRF model to capture the dependencies between adjacent words in text sequence.



$$s(\mathbf{x},\mathbf{y}) = \sum_{t=0}^{l} A_{y_t, y_{t+1}} + \sum_{t=1}^{l} P_{t,y_t} \quad (8)$$

$$p(\mathbf{y}|\mathbf{x}) = \frac{\exp(s(\mathbf{x},\mathbf{y}))}{\sum_{\tilde{\mathbf{y}}\in \mathbf{Y_x}} \exp(s(\mathbf{x},\tilde{\mathbf{y}}))} \quad (9)$$

$$\mathcal{L} = \log(p(\mathbf{y}|\mathbf{x})) = s(\mathbf{x},\mathbf{y}) - \log\left(\sum_{\tilde{\mathbf{y}}\in \mathbf{Y_x}} \exp(s(\mathbf{x},\tilde{\mathbf{y}}))\right). \quad (10)$$

$$y^* = \arg\max_{\tilde{\mathbf{y}}\in \mathbf{Y_x}} s(\mathbf{x},\tilde{\mathbf{y}}). \quad (11)$$

Datasets:

We evaluate our models using the research paper datasets collected by recent works on keyphrase extraction.

Table 1: Statistics of the two benchmark datasets.

Dataset	#Abs/ #Kps(All)	#Abs/ #Kps(Loc.)	Missing KPs(%)	#AvgKPs	#unigrams	#bigrams	#trigrams	#> trigrams
KDD	365/1471	315/719	51.12	4.03	363	853	189	66
WWW	425/2073	388/904	56.39	4.87	680	1036	247	110

Evaluation metrics:

We use precision (**P**), recall (**R**) and F1-score (**F1**) to evaluate the results.

Compareative Methods:

- **KEA** employs a supervised *Naïve Bayes classifier* to extract keyphrases using only two features: *TF-IDF* of a phrase and *the distance of a phrase from the beginning of a document*.

- **CeKE** uses a *Naïve Bayes classifier* for extracting keyphrases from research papers embedded in *citation networks*.

- **EK-CRF** uses the *CRF algorithm* based on sequence labeling to extract keyphrases from research papers. This method incorporates the *expert-knowledge* and domain-specific hints.

- **CopyRNN** is the first to employ *sequence-to-sequence* (Seq2Seq) framework with attention and *copy mechanisms* to generate keyphrases. This method is able to predict absent keyphrases that do not appear in the target document.

Table 2: Comparison of the proposed models with other approaches

Method	KDD			WWW		
	Precision	Recall	F1-score	Precision	Recall	F1-score
KEA [24]	0.1551	0.3278	0.2105	0.1549	0.3182	0.2084
CeKE [3]	0.2174	0.3905	0.2793	0.2251	0.2519	0.2377
EK-CRF [9]	0.4068	0.2162	0.2823	0.3689	0.194	0.2547
CopyRNN@5 [16]	0.2221	0.4926	0.3062	0.1907	0.3993	0.2581
M-CRF^s	0.3438	0.2354	0.2794	0.3276	0.2059	0.2528
M-CRF^d	0.3597	0.2542	0.2979	0.3551	0.2129	0.2662
MLM-CRF	**0.3787**	**0.2771**	**0.32**	**0.3251**	**0.2417**	**0.2773**

- **MLM-CRF** *VS* **M-CRF^d** *VS* **M-CRF^s** ***P*** ***R*** ***F***
- **MLM-CRF** *VS* **Others** ***F***

- We divide keyphrases into two categories: **simple keyphrase** (SK) and **complicated keyphrase** (CK).

- Each word in dataset is labeled with non-keyphrase (**O**), simple keyphrase (**SK**) or complicated keyphrase (**CK**).
  - For the complicated keyphrase, **B-CK, M-CK** and **E-CK** correspond to the beginning, middle and end word of complicated keyphrase, respectively.

Text sequence:	Enhanced **max margin** learning on **multimodal data mining** in a multimedia database.
Label sequence:	O B-CK E-CK O O B-CK M-CK E-CK O O O O

Text sequence:	**Boosting** with structure information in the functional space : an application to **graph classification**.
Label sequence:	SK O O O O O O O O O O B-CK E-CK

Table 3: Comparison of our different models in different type keyphrases.

Methods	KDD						WWW					
	SK			CK			SK			CK		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
M-CRF^s	0.3679	0.1235	0.1849	0.3382	0.3059	0.3212	0.3534	0.1868	0.2444	0.3092	0.2247	0.2603
M-CRF^d	0.3180	0.1444	0.1986	0.3734	0.3235	0.3466	0.3659	0.2009	0.2594	0.3460	0.2247	0.2725
MLM-CRF	0.4589	0.1843	0.2630	0.3720	0.3520	0.3617	0.3453	0.2283	0.2749	0.3090	0.2548	0.2793

- **SK:** **M-CRF^s** VS **M-CRF^d** *R* ↑ *F* ↑

  MLM-CRF VS **M-CRF^s** or **M-CRF^d** *R* ↑ *F* ↑ *P* ↑

- **CK:** **M-CRF^s** VS **M-CRF^d** *R* ↑ *F* ↑ *P* ↑

  MLM-CRF VS **M-CRF^s** or **M-CRF^d** *R* ↑ *F* ↑

- **CK** VS **SK:** *R* ↑ *F* ↑ *P* ↑

VI. Conclusion

- We proposed a multi-level memory network with CRFs named **MLM-CRF** for extracting keyphrases from scientific research papers.

- In particular, we first extended the input memory of the **memory network** with two different levels to capture the long-range contextual information hidden in text data. We then employed the **CRF** model to capture the structural dependencies between adjacent words in text sequence and determine whether a candidate phrase is a keyphrase.

Q&A