Title-Aware Neural News Topic Prediction

Chuhan Wu[1], Fangzhao Wu[2], Tao Qi[1], Yongfeng Huang[1], and Xing Xie[2]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
{wu-ch19,qit16,yfhuang}@mails.tsinghua.edu.cn
[2] Microsoft Research Asia, Beijing 100080, China
{fangzwu,xing.xie}@microsoft.com

Abstract. Online news platforms have gained huge popularity for online news reading. The topic categories of news are very important for these platforms to target user interests and make personalized recommendations. However, massive news articles are generated everyday, and it too expensive and time-consuming to manually categorize all news. The news bodies usually convey the detailed information of news, and the news titles usually contain summarized and complementary information of news. However, existing news topic prediction methods usually simply aggregate news titles and bodies together and ignore the differences of their characteristics. In this paper, we propose a title-aware neural news topic prediction approach to classify the topic categories of online news articles. In our approach, we propose a multi-view learning framework to incorporate news titles and bodies as different views of news to learn unified news representations. In the title view, we learn title representations from words via a long-short term memory (LSTM) network, and use attention mechanism to select important words according to their contextual representations. In the body view, we propose to use a hierarchical LSTM network to first learn sentence representations from words, and then learn body representations from sentences. In addition, we apply attention networks at both word and sentence levels to recognize important words and sentences. Besides, we use the representation vector of news title to initialize the hidden states of the LSTM networks for news body to capture the summarized news information condensed by news titles. Extensive experiments on a real-world dataset validate that our approach can achieve good performance in news topic prediction and consistently outperform many baseline methods.

Keywords: News Topic Prediction · Multi-view Learning · Attention Mechanism.

1 Introduction

Online news platforms such as Bing News and Google News have gained huge popularity and attracted millions of users to read digital news online [5]. On many news platforms, news articles are classified into different topic categories such as sports and finance to target user interests and make news recommendations [20]. However, hundreds of thousands of news articles emerge everyday,

Title	James Harden's incredible heroics lift Rockets over Warriors in overtime	The 5 best movies screening around Orange this week
Body	James Harden was blanketed with nowhere to go and no time to get there. It did not matter. The hottest player in the NBA would not be stopped, not by a 20-point deficit, not by all the defensive clamps the tag team of Klay Thompson and Draymond Green could wrap around him...	Spider-Man: Into the Spider-Verse Miles Morales is juggling his life between being a high school student and being Spider-Man. However, when Wilson "Kingpin" Fisk uses a super collider, another Spider-Man from another dimension, Peter Parker, accidentally winds up in Miles' dimension...
Category	sports	movies

Fig. 1. Two illustrative example news with different topics. Several important words in news title and body are highlighted.

and it is too expensive and time-consuming to manually categorize all news articles [6]. Thus, automatically predicting the topic categories of news is very important for online news platforms to provide personalized news services [23].

Learning accurate representations of news is critical for news topic prediction. Many of existing news topic prediction methods build news representations via manual feature engineering [6, 25, 28]. For example, Dilrukshi et al. [6] proposed to use support vector machine (SVM) to classify Twitter news, and use word unigram features to represent news. However, the bag-of-word features used in these methods cannot capture the contexts and orders of words, both of which are important for the prediction of news topics. In recent years, several deep learning based text classification methods are applied to news topic prediction. For example, Zhang et al. [29] proposed to use a character-level convolutional neural network (CNN) to learn news representations from original characters. Lai et al. [13] proposed to use a convolutional recurrent neural network to learn news representations from news bodies. Conneau et al. [4] proposed to apply a deep CNN network with shortcut connections to learn hidden news representations. However, these methods usually simply formulate the news topic prediction task as a document classification problem, i.e., aggregating the title and body of news as a single document, while the differences in characteristics between news title and news body are not taken into consideration.

Our work is motivated by the observation that both news titles and news bodies are useful for learning news representations. The bodies usually convey the detailed information of the news, and the titles usually convey summarized and supplementary information of news. For example, in Fig. 1 the title of the first news shows that this news is about an NBA event, and the body introduces its details. In addition, in the second news, the news body only introduces the details of movies, but the title clearly summarizes the news topic. Thus, incorporating the information of both news titles and bodies has the potential to enhance the learning of news representations for topic prediction. In addition, the characteristics of news titles and news bodies usually have some differences, since titles are usually short and concise sentences, while bodies are usually long documents with rich details. Thus, they should be handled differently. Besides, different words in the same news title or body usually have different informativeness for learning news representations. For example, in Fig. 1 the word "NBA" is more informative than "would". Moreover, different sentences in the same news body may also have different importance. For instance, the first sentence in the body of the first news in Fig. 1 is more informative than the second sentence in inferring the topic of this news.

In this paper, we propose a title-aware neural news topic prediction approach which can utilize both news title and news body information. In our approach, we propose to use a multi-view learning framework to incorporate both titles and bodies as different views of news for learning unified news representations. In the title view, we learn title representations from words via an LSTM network, and use attention mechanism to recognize important words. In the body view, we use a hierarchical LSTM network to first learn sentence representations from words, and then learn body representations from sentences. In addition, we apply attention mechanism to select important words and sentences for learning informative news representations. Moreover, we propose to use the representations vector of news title to initialized the hidden states of the LSTM networks in the body view. Thus, our approach can learn more accurate representations of news bodies with the help of summarized information provided by news titles. Extensive experiments are conducted on a real-world dataset, and the results show that our approach can effectively improve the performance of news topic prediction and consistently outperform many baseline methods.

2 Related Work

News topic prediction is an important task in the natural language processing field and has been extensively explored over years [23]. News topic prediction is usually formulated as a text classification problem, and learning accurate news representations is a core step. Many of existing methods build news representations via manual feature engineering [24–26, 2, 1, 6, 10, 28, 3, 19, 15]. For example, Yang et al. [26] proposed to use logistic regression (LR) to classify news topics and they used the TF-IDF features to build representations of news documents. Dilrukshi et al. [6] proposed to use support vector machine (SVM) to classify Twitter news, and they used word unigrams to represent news. Antonellis et al. [1] proposed to build news representations via the addition of the sentence representations learn by the cosine similarities between news representations and category term representations. Joulin et al. [9] proposed to use a bag of n-grams as the features to represent news documents. In addition, they used a hashing strategy to reduce the memory cost. However, these methods cannot effectively utilize contextual information and word orders. In addition, they cannot distinguish informative contexts from uninformative ones.

In recent years, several deep learning based methods have been proposed to automatically learn news representations from their content [29, 13, 31, 18, 4, 21, 16]. For example, Zhang et al. [29] proposed to apply a CNN network at character-level to learn representations of news from their original characters. Lai et al. [13] proposed a convolutional recurrent neural network to learn representations of news from their bodies by capturing both local and global contexts. Zhu

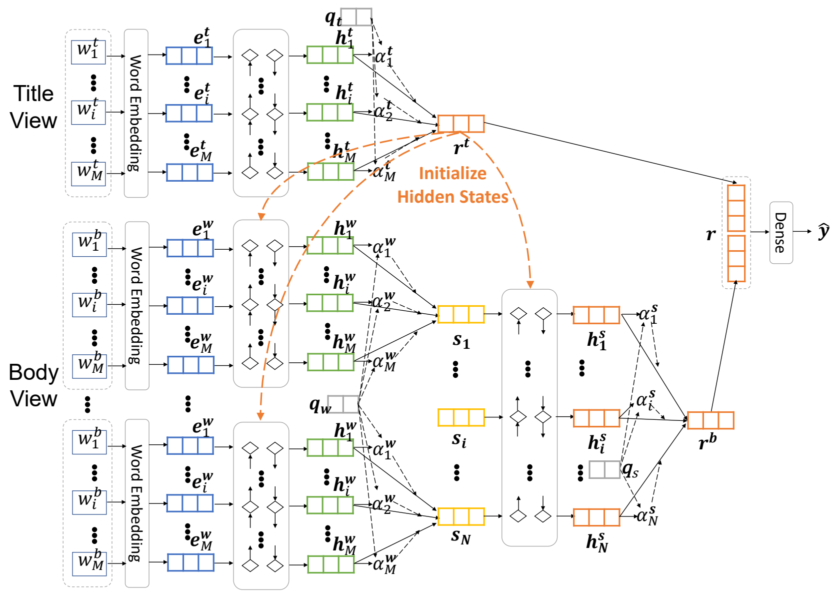

Fig. 2. The framework of our *TAP* approach.

et al. [31] proposed a voting method to learn representations of news from their titles based on an ensemble of a CNN network, a gated recurrent units (GRU) network, and an SVM trained using TF-IDF features. Lu et al. [18] proposed to use a Bi-LSTM network to learn news representations from news headlines, and they used attention mechanism to select important words. However, these methods usually only consider either the title or the body of news, which is usually insufficient for representing news accurately. Different from these methods, our approach can learn title-aware news representations by utilizing the information in both news titles and news bodies.

3 Our Approach

In this section, we will introduce our Title-Aware neural news topic Prediction approach (TAP). There are two main components in our model, i.e., a *title-aware news representation* module to learn news representations from both news titles and bodies, and a *topic classification* module to classify news into different topic categories. The overall framework of our approach is shown in Fig. 2. We will introduce the details of our *TAP* approach as follows.

3.1 Title-Aware News Representation

The *title-aware news representation* aims to learn news representations by incorporating the information of both news title and news body. In our approach, we propose to use a multi-view learning framework to learn unified news representations by regarding titles and bodies as different views of news.

Title Representation. The first view is *title representation*. It contains three layers. The first one is word embedding. It aims to convert a news title with M words from a word sequence into a low-dimensional vector sequence. Denote the input news title as $[w_1^t, w_2^t, ..., w_M^t]$, the output of this layer is an embedding sequence $[\mathbf{e}_1^t, \mathbf{e}_2^t, ..., \mathbf{e}_M^t]$.

The second layer is a bi-directional long-short term memory network (Bi-LSTM) [11]. Global contexts in a news title are usually necessary for understanding this news. For example, in the title of the first news of Fig. 1, the contexts of "Rockets" such as "lift" and "James Harden's" are very useful for representing this news. In addition, the contexts in the future are also useful for learning informative word representations. For example, the word "Warriors" is also important for understanding "Rockets". Thus, we use a Bi-LSTM network to learn word representations by summarizing the past and future contextual information in both directions. It scans the input embedding sequence forward and backward, and outputs a hidden word representation sequence $[\mathbf{h}_1^t, \mathbf{h}_2^t, ..., \mathbf{h}_M^t]$.

The third layer is a word-level attention network. Usually, different words have different contributions to the representation learning of news title. For example, in the second news title of Fig. 1, the word "movies" is very informative for inferring news topic, while the word "around" is uninformative. Thus, we apply an attention network to select important words to learn informative title representations for topic prediction. Denote the attention weight of the i-th word in the same news title as α_i^t, which is computed as:

$$a_i^t = \mathbf{q}_t \times \tanh(\mathbf{W}_t \times \mathbf{h}_i^t + \mathbf{w}_t), \quad \alpha_i^t = \frac{\exp(a_i^t)}{\sum_{j=1}^{M} \exp(a_j^t)}, \tag{1}$$

where $\mathbf{W}_t$ and $\mathbf{w}_t$ are projection parameters, $\mathbf{q}_t$ is the word attention query. The contextual representation of the news title is the summation of the hidden word representations weighted by their attention weights, i.e., $\mathbf{r}^t = \sum_{j=1}^{M} \alpha_j^t \mathbf{h}_j^t$.

Body Representation. The second view is *body representation*, which is used to learn news representations from news bodies. It contains two major modules. The first module is *sentence representation*. Similar with the title view, there are also three layers in this module.

The first one is a shared word embedding layer. Denote a sentence in a news body as $[w_1^b, w_2^b, ..., w_M^b]$, it is converted from a word sequence into a vector sequence $[\mathbf{e}_1^b, \mathbf{e}_2^b, ..., \mathbf{e}_M^b]$.

The second one is a title-aware Bi-LSTM network. Global contexts within the same sentence in a news body are also important for learning body representations. For example, in the first news of Fig. 1, the contexts of the word "player" such as "hottest" and "NBA" are important for inferring the topic of this news. However, it may be difficult to capture important topic information with limited contexts. For example, the word "20-point" may not be topic sensitive in many news, but it is very useful in inferring the topic of this news. Fortunately, the summarized information provided by news titles has the potential to help learn more topic discriminative body representations. Thus, we propose to use a title-aware Bi-LSTM network by using the title representation vector to initialize the hidden states of the Bi-LSTM. In addition, we need to apply a linear transformation to the title representation vector to align its dimension with the size of LSTM hidden states. After initialization, the Bi-LSTM scans the word embedding sequence $[\mathbf{e}_1^w, \mathbf{e}_2^w, ..., \mathbf{e}_M^w]$ in both directions, and outputs the hidden word representation sequence $[\mathbf{h}_1^w, \mathbf{h}_2^w, ..., \mathbf{h}_M^w]$.

The third one is a word attention network. Different words in the same sentence usually have different importance in learning sentence representations. For example, in the first news of Fig. 1, the word "NBA" is much more informative than "would" in representing its third sentence "The hottest...". Thus, we use an attention network to select important words according to their hidden representations. Denote the attention weight of the i-th word in a sentence as α_i^w, which is computed as:

$$a_i^w = \mathbf{q}_w \times \tanh(\mathbf{W}_w \times \mathbf{h}_i^w + \mathbf{w}_w), \quad \alpha_i^w = \frac{\exp(a_i^w)}{\sum_{j=1}^{M} \exp(a_j^w)}, \tag{2}$$

where $\mathbf{W}_w$, $\mathbf{w}_w$ and $\mathbf{q}_w$ are projection parameters. The contextual sentence representation is the summation of the word representations weighted by their attention weight, i.e., $\mathbf{s} = \sum_{i=1}^{M} \alpha_i^w \mathbf{h}_i^w$. The representations of each sentence in the news body are computed in a similar way. We denote the sentence representation sequence as $[\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N]$, where N is the number of sentences.

The second module is *document representation*. It contains two layers. Global contexts of sentences are also important for learning body representations. For example, in the first news of Fig. 1, the relatedness between the first and the third sentence is important for representing the entire document. In addition, since news titles can provide the global topic information, incorporating the summarized information condensed by news titles may help improve the sentence representation learning. Thus, we use a title-aware Bi-LSTM network at sentence-level to learn contextual sentence representations. It takes the sentence representation sequence $[\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N]$ as input, and outputs the contextual sentence representation sequence $[\mathbf{h}_1^s, \mathbf{h}_2^s, ..., \mathbf{h}_N^s]$.

The second one is a sentence-level attention network. Different sentences in the news body may also have different informativeness in learning body representations. For example, in the first news of Fig. 1, the first sentence is more informative than the second sentence for learning body representations. Thus, we propose to use a sentence attention network to select important sentences. Denote the attention weight of the i-th sentence in the news body as α_i^s, which is computed as:

$$a_i^s = \mathbf{q}_s \times \tanh(\mathbf{W}_s \times \mathbf{h}_i^s + \mathbf{w}_s), \quad \alpha_i^s = \frac{\exp(a_i^s)}{\sum_{j=1}^{N} \exp(a_j^s)}, \tag{3}$$

where $\mathbf{W}_s$, $\mathbf{w}_s$ and $\mathbf{q}_s$ are attention parameters. The contextual body representation is computed as: $\mathbf{r}^b = \sum_{i=1}^{N} \alpha_i^s \mathbf{h}_i^s$. The final unified news representation $\mathbf{r}$ is the concatenation of the new representations learned from the title view and the body view, i.e., $\mathbf{r} = [\mathbf{r}^t; \mathbf{r}^b]$.

3.2 Topic Classification

The *topic classification* module is used to classify the topic category of a news by predicting the probabilities $\hat{\mathbf{y}}$ of a news in different topic category as follows:

$$\hat{\mathbf{y}} = softmax(\mathbf{W}_y \times \mathbf{r} + \mathbf{w}_y), \tag{4}$$

where $\mathbf{W}_y$ and $\mathbf{w}_y$ are parameters. In the model training stage, cross entropy is used as the loss function, which is computed as follows:

$$\mathcal{L} = -\sum_{i=1}^{S}\sum_{t=1}^{T} y_{i,t} \log(\hat{y}_{i,t}), \tag{5}$$

where S is the number of training samples, T is the number of topic categories, $y_{i,k}$ and $\hat{y}_{i,k}$ represent the gold and predicted probability of the i-th news in the k-th topic category, respectively.

4 Experiments

4.1 Datasets and Experimental Settings

Since existing news topic classification datasets such as 20 News [14], AG News[3], and UCI News [17] usually only contain either news title or news body, we constructed a new news topic classification dataset which contains both news title and news body information by crawling news articles from the MSN News[4] platform during 12/13/2018 and 01/12/2019. The final dataset has 31,908 news articles in 14 topic categories. The topic distributions of these news articles are shown in Fig. 3. The average number of words per title is 11.16, and the number for news body is 761.86. We use 80% of the news for training, 10% for validation, and the remaining 10% for test.

In our experiments, the word embeddings were 300-dimensional and we used the pre-trained Glove embeddings [22] to initialize them. The dimension of the hidden states in the Bi-LSTM networks was 2×200. The attention query vectors were also 200-dimensional. We used the Adam [12] algorithm to optimize the loss function. To mitigate overfitting, we added 30% dropout to each layer. The size of a training minibatch was 64. These hyperparameters were tuned on the validation set. Each experiment was repeated 10 times and we reported the average accuracy and macro F1score.

4.2 Performance Evaluation

First, we want to evaluate the performance of our approach by comparing it with several baseline methods for news topic prediction, including: (1) *SVM* [6], support vector machine; (2) *LR* [26], logistic regression; (3) *FastText* [9], a famous

[3] http://www.di.unipi.it/~gulli/AG_corpus_of_news_articles.html
[4] https://www.msn.com/en-us/news

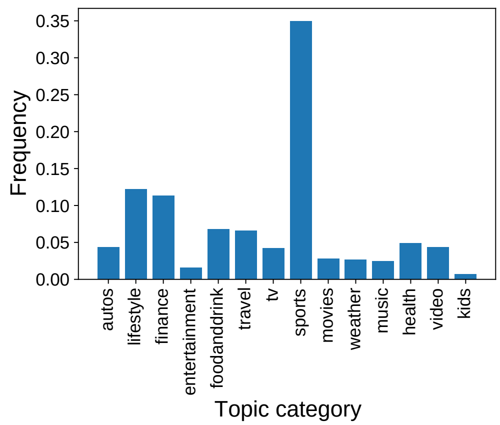

Fig. 3. Distributions of different news topics.

text classification method based on n-grams; (4) *CNN* [11], convolutional neural network. (5) *LSTM* [8], long short-term memory network. (6) *CLSTM* [30, 16], a combination of both CNN and LSTM network. (7) *CNN-Att* [7], CNN with attention mechanism. (8) *LSTM-Att* [18], LSTM with attention mechanism. (9) *HAN* [27], a hierarchical attention network for document classification. (10) *TAP*, our title-aware neural news topic prediction approach. In traditional methods including *SVM* and *LR*, we used the TF-IDF features extracted from news titles and bodies as the input. In other baseline methods (3-9), we used the combination of news titles and bodies by simply regarding them as a single document. The performance of these methods under different ratios of training data are summarized in Table 1. We have several observations from the results.

First, the methods which learn news representations from news content (e.g., *CNN*, *LSTM* and *TAP*) outperform the methods which build news representations via feature engineering (*SVM*, *LR* and *Linreg*). This is probably because the latter methods rely on bag-of-words features, while the contextual information and word orders cannot be effectively modeled. Second, the methods with attention mechanism (e.g., *CNN-Att*) outperform the methods without mechanism (e.g., *CNN*). This is probably because different contexts within a news usually have different informativeness in representing this news, and distinguishing important contexts from unimportant ones can benefit news representation learning. Third, the methods based on hierarchical neural architectures (*HAN* and *TAP*) outperform the methods based on flatten models (e.g., *CNN-Att* and *LSTM-Att*). This is because news are usually long documents, and it may be more effective to model news in a hierarchical manner to exploit the document structures of news. Fourth, our approach can outperform all other baselines. This is probably because the characteristics of news titles and news bodies are quite different, and they should be processed differently. Thus, the performance of other baselines such as *HAN* is sub-optimal. Different from these methods, our approach employs a multi-view learning framework to learn unified news

Method	25%		50%		100%	
	Accuracy	Fscore	Accuracy	Fscore	Accuracy	Fscore
SVM [6]	73.59±0.65	61.11±0.69	74.45±0.71	62.36±0.75	74.88±0.77	63.11±0.79
LR [26]	73.18±0.68	60.24±0.72	74.47±0.72	61.88±0.76	75.35±0.74	62.44±0.78
FastText [9]	74.08±0.72	61.88±0.78	76.10±0.69	63.22±0.73	78.20±0.62	66.34±0.65
CNN [11]	74.79±0.64	62.40±0.66	76.55±0.53	64.31±0.58	78.44±0.40	67.61±0.43
LSTM [8]	74.55±0.48	61.96±0.52	76.66±0.44	64.48±0.47	78.33±0.36	67.55±0.38
CLSTM [30]	75.04±0.60	62.33±0.63	77.11±0.58	65.00±0.62	78.86±0.43	67.94±0.47
CNN-Att [7]	75.22±0.56	62.59±0.60	77.33±0.51	65.29±0.53	78.98±0.44	68.15±0.47
LSTM-Att [18]	74.89±0.51	62.47±0.54	77.59±0.43	65.54±0.46	79.12±0.39	68.23±0.42
HAN [27]	76.22±0.44	63.15±0.47	78.22±0.35	66.06±0.40	79.96±0.35	68.88±0.39
TAP*	**77.63**±0.38	**64.34**±0.42	**79.50**±0.31	**67.44**±0.34	**81.49**±0.26	**70.22**±0.30

Table 1. The performance of different methods under different ratios of training data. *The advantage of our approach over all baselines is significant at $p < 0.005$.

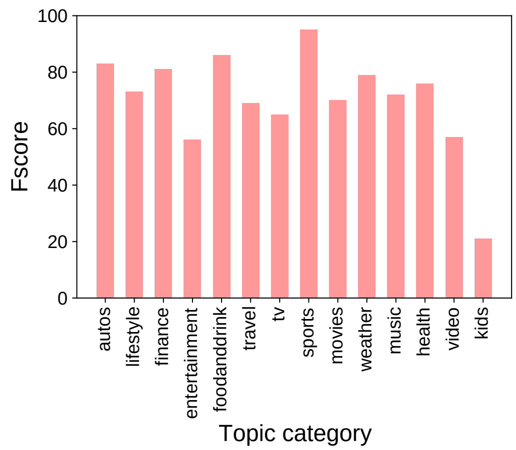

Fig. 4. Performance of the recognition of different topic categories.

representations by regarding news titles and bodies as different views of news, which is useful for learning better news representations.

Then, we want to explore the performance of our approach in the classification of different topic categories. The performance in Fscore of recognizing different topic categories is shown in Fig. 4. From Fig. 4, we find our approach can achieve satisfactory performance in recognizing most topics, even the sports news are dominant in our dataset. However, it is very difficult to classify the news in the category "kids", since the training samples in this category are too scarce. These results show that our approach is effective in news topic prediction.

4.3 Effectiveness of Learning Title-Aware News Representations

In this section, we conducted several experiments to validate the effectiveness of learning title-aware news representations in our approach. We compare the performance of our approach and its three variants, i.e., without title information, incorporating titles via multi-view learning only, and using title-aware

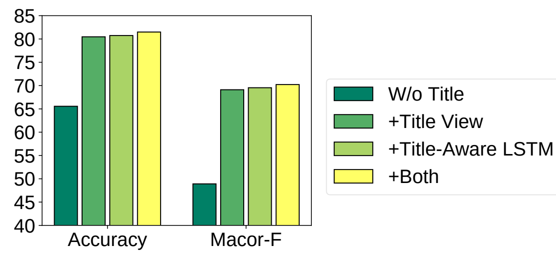

Fig. 5. Effectiveness of learning title-aware news representations.

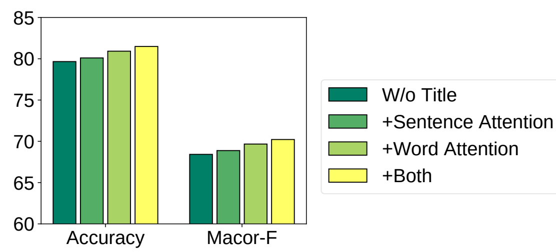

Fig. 6. Effectiveness of different attention networks.

LSTMs only. The results are shown in Fig. 5. According to Fig. 5, we have several observations. First, we find the performance of our approach will decline seriously if title information is not considered. This is probably because news titles are usually synthesis of news topics and highlights, which are critical for topic prediction. Thus, it is necessary to incorporate the information of news titles. Second, incorporating news title and body via multi-view learning can effectively improve the performance of our approach. This may be because news titles can provide useful complementary information to news bodies In addition, using title-aware LSTM networks is also useful. This is probably because news titles can provide summarized information of news, and can help learn more informative body representations. Third, combining both techniques can further improve the performance of our approach. These results validate the effectiveness of learning title-aware news representations in our approach.

4.4 Effectiveness of Attention Mechanism

In this section, we conducted experiments to verify the effectiveness of the word-level and sentence-level attention networks in our approach. The performance of our approach and its variant with different combinations of attention networks is shown in Fig. 6. The results in Fig. 6 lead to several observations. First, the word-level attention networks are very important for our approach. This is probably because different words in the same news title and body usually have different informativeness for learning news representations. Thus, selecting important words in news titles and bodies can help learn more informative news representations. Second, the sentence-level attention network is also useful for our approach. This is probably because different sentences also have different informativeness in representing news bodies, and selecting important sentences is beneficial for body representation learning. Third, combining both kinds of attention networks can further improve our approach. These results validate the effectiveness of the hierarchical attention networks in our approach.

5 Conclusion

In this paper, we propose a title-aware neural news topic prediction approach which can incorporate both news title and news body to learn informative representation of news. In our approach, we propose a multi-view learning framework to exploit the titles and the bodies as different views of news articles. In the title view, we learn the representations of news titles via a Bi-LSTM network, and use an attention network to select important words. In the body view, we learn the representations of news bodies in a hierarchical way. We first learn sentence representations from words, and then learn body representations from sentences. We apply a hierarchical attention network to select important words and sentences. In addition, we propose a title-aware LSTM network by using the representation of news titles to initialize the hidden states of the LSTM networks for body representation learning. Extensive experiments on a real-world dataset validate the effectiveness of our approach.

Acknowledgments

The authors would like to thank Microsoft News for providing technical support and data in the experiments, and Jiun-Hung Chen (Microsoft News) and Ying Qiao (Microsoft News) for their support and discussions. This work was supported by the National Key Research and Development Program of China under Grant number 2018YFC1604002, the National Natural Science Foundation of China under Grant numbers U1836204, U1705261, U1636113, U1536201, and U1536207, and the Tsinghua University Initiative Scientific Research Program.

References

1. Adi, A.O., Çelebi, E.: Classification of 20 news group with naïve bayes classifier. In: SIU. pp. 2150–2153. IEEE (2014)
2. Bracewell, D.B., Yan, J., Ren, F., Kuroiwa, S.: Category classification and topic discovery of japanese and english news articles. ENTCS **225**, 51–65 (2009)
3. Cecchini, D., Na, L.: Chinese news classification. In: BigComp. pp. 681–684 (2018)
4. Conneau, A., Schwenk, H., Barrault, L., Lecun, Y.: Very deep convolutional networks for text classification. In: EACL. pp. 1107–1116 (2017)
5. Das, A.S., Datar, M., Garg, A., Rajaram, S.: Google news personalization: scalable online collaborative filtering. In: WWW. pp. 271–280. ACM (2007)
6. Dilrukshi, I., De Zoysa, K., Caldera, A.: Twitter news classification using svm. In: ICCSE. pp. 287–291. IEEE (2013)
7. Du, J., Gui, L., Xu, R., He, Y.: A convolutional attention model for text classification. In: NLPCC. pp. 183–195. Springer (2017)

8. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
9. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. In: EACL. vol. 2, pp. 427–431 (2017)
10. Kaur, G., Bajaj, K.: News classification using neural networks. Commun. Appl. Electron **5**(1) (2016)
11. Kim, Y.: Convolutional neural networks for sentence classification. In: EMNLP. pp. 1746–1751 (2014)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Lai, S., Xu, L., Liu, K., Zhao, J.: Recurrent convolutional neural networks for text classification. In: AAAI. pp. 2267–2273. AAAI Press (2015)
14. Lang, K.: Newsweeder: Learning to filter netnews. In: Machine Learning Proceedings, pp. 331–339. Elsevier (1995)
15. Lange, L., Alonso, O., Strötgen, J.: The power of temporal features for classifying news articles. In: WWW. pp. 1159–1160. ACM (2019)
16. Li, C., Zhan, G., Li, Z.: News text classification based on improved bi-lstm-cnn. In: ITME. pp. 890–893. IEEE (2018)
17. Lichman, M., et al.: Uci machine learning repository (2013)
18. Lu, Z., Liu, W., Zhou, Y., Hu, X., Wang, B.: An effective approach for chinese news headline classification based on multi-representation mixed model with attention and ensemble learning. In: NLPCC. pp. 339–350. Springer (2017)
19. Majeed, F., Asif, M.W., Hassan, M.A., Abbas, S.A., Lali, M.I.: Social media news classification in healthcare communication. Journal of Medical Imaging and Health Informatics **9**(6), 1215–1223 (2019)
20. Okura, S., Tagami, Y., Ono, S., Tajima, A.: Embedding-based news recommendation for millions of users. In: KDD. pp. 1933–1942. ACM (2017)
21. Peng, H., Li, J., He, Y., Liu, Y., Bao, M., Wang, L., Song, Y., Yang, Q.: Large-scale hierarchical text classification with recursively regularized deep graph-cnn. In: WWW. pp. 1063–1072 (2018)
22. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: EMNLP. pp. 1532–1543 (2014)
23. Qiu, X., Gong, J., Huang, X.: Overview of the nlpcc 2017 shared task: Chinese news headline categorization. In: NLPCC. pp. 948–953. Springer (2017)
24. Sawaf, H., Zaplo, J., Ney, H.: Statistical classification methods for arabic news articles. In: Arabic Natural Language Processing in ACL2001. Citeseer (2001)
25. Tenenboim, L., Shapira, B., Shoval, P.: Ontology-based classification of news in an electronic newspaper. Advanced Research in Artificial Intelligence p. 89 (2008)
26. Yang, B., Sun, J.T., Wang, T., Chen, Z.: Effective multi-label active learning for text classification. In: KDD. pp. 917–926. ACM (2009)
27. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. In: NAACL. pp. 1480–1489 (2016)
28. Yin, Z., Tang, J., Ru, C., Luo, W., Luo, Z., Ma, X.: A semantic representation enhancement method for chinese news headline classification. In: NLPCC. pp. 318–328. Springer (2017)
29. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: NIPS. pp. 649–657 (2015)
30. Zhou, C., Sun, C., Liu, Z., Lau, F.: A c-lstm neural network for text classification. arXiv preprint arXiv:1511.08630 (2015)
31. Zhu, F., Dong, X., Song, R., Hong, Y., Zhu, Q.: A multiple learning model based voting system for news headline classification. In: NLPCC. pp. 797–806 (2017)