

Context-Aware Feature Attention Model for Coreference Resolution on Biomedical Texts

Yufei Li

Xi'an Jiaotong University

Xiangyu Zhou

Xi'an Jiaotong University

Jie Ma

Xi'an Jiaotong University

Xiaoyong Ma

Xi'an Jiaotong University

Chen Li (✉ cli@xjtu.edu.cn)

Xi'an Jiaotong University

Research Article

Keywords: Context Awareness, Semantics, Biomedical Text Mining, Coreference Resolution

Posted Date: January 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-149879/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Context-aware Feature Attention Model for Coreference Resolution on Biomedical Texts

Yufei Li^{1,2,3}, Xiangyu Zhou^{1,2,3}, Jie Ma^{1,2,3}, Xiaoyong Ma^{1,2,3} and Chen Li^{1,2,3*}

Abstract

Background: Bio-entity Coreference resolution is an important task to gain a complete understanding of biomedical texts automatically. Previous neural network-based studies on this topic are domain system based methods which rely on some domain-specific information integration. However, for the identical mentions, this may lead to misleading information, as the model tends to get similar or even identical representations, which further leads to wrong predictions.

Results: we propose a new context-aware Feature Attention model to distinguish identical mentions effectively to better resolve coreference. The new model can represent identical mentions based on different contexts by adaptively exploiting features effectively. The proposed model substantially outperforms the state-of-the-art baselines on the BioNLP dataset with a 64.0% F1 score and further demonstrates superior performance on the differential representation and coreferential link of identical mentions.

Conclusion: The context-aware Feature Attention model adaptively exploit features and represent identical mentions according to different contexts, which significantly makes the system obtain semantic information effectively and make more accurate predictions. Considering that this approach is still limited when context information is insufficient, we expect to utilize such information more fine-grained with the help of the external knowledge base in coreference resolution.

Keywords: Context Awareness; Semantics; Biomedical Text Mining; Coreference Resolution

Background

Context and Motivation

Bio-entity Coreference resolution is a crucial task for artificial intelligence systems to be capable of fully understanding the biomedical texts by improving the performance of several downstream tasks, including information extraction [1], entity linking [2], question answering [3] and so on.

Typical works of Bio-entity coreference resolution use rule-based [4, 5, 6] and hybrid methods [7, 8], which rely on syntactic features and are limited to a specific corpus. Recently, neural network-based methods for automatically identifying coreferences have received widespread attention. Some neural network-based systems have been developed that integrate domain-specific information through pre-trained embeddings and static bio-related features. [6, 9, 10].

However, for similar or identical mentions (formal definition is given in the Task Section), the above-mentioned methods often get similar or even the same representation, which can be misleading and makes the coreferences more puzzling. As shown in Figure 1, there are three identical mentions: "it". While the "it" marked in red and orange refers to two different concepts. As they are likely to get a similar or even identical representation, a false coreferential link between them is often predicted.

Generally, the differential representation of identical mentions based on context is necessary as the above problem accounts for a large proportion of the coreference dataset. Here for protein coreference, we made statistics on the identical mentions of the BioNLP dataset. Table 1 shows the detailed statistics of identical mentions throughout the BioNLP dataset. We can find that the identical mentions account for more than half of the entire dataset, whatever the train or the development. While among these identical mentions, less than 10% have a coreferential link. We further record the frequency ranges and the Pos (part-of-speech) tags of these identical mentions. The results illustrate that mentions with a frequency greater than 100 occupy

¹School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, P.R. China

Full list of author information is available at the end of the article

*Correspondence: cli@xjtu.edu.cn

more than half. Moreover, among them, Wh-qualifiers, prepositions, conjunctions, and noun phrase take up more. In Table 2, we count the identical mentions in each document. We display the percentage of the documents with or without identical mentions and the percentage of documents containing different numbers of the identical mentions. The results show that documents containing identical mentions account for about half of the total documents, of which two identical mentions account for the largest proportion.

Inspired by the recent success of attention mechanism [11, 12], this paper explores how to distinguish similar or identical text units effectively based on contexts and develops a general context-aware Feature Attention mechanism to enhance the neural biomedical coreference system. Based on a general neural coreference system [13], we first run the model with integrating some domain-specific information (embeddings pre-trained on large scale corpus, some biomedical and grammatical features). Then we use Feature Attention mechanism to learn the importance or weight of each feature based on contexts such that the span representation can be improved. In this way, the model can adaptively exploit features and differentiate identical mentions based on different contexts.

The proposed model is evaluated on the BioNLP Protein Coreference dataset [14]. The results show that our work achieves state-of-the-art results with a 64.0% F1 score (+2.0 F1). Besides, experimental analysis shows the Feature Attention mechanism does help in distinguishing the identical mentions based on different contexts, which further helps in coreferential links. Moreover, additional experiments on the coreferential linking performance of identical mentions further illustrate the effectiveness of our model.

Related Work

In bio-entity coreference resolution tasks, words referring to each other are called mentions, while a mention can either be a common noun, a proper noun, or a pronoun. Taking the above example in Figure 1, a coreference system partitions the mentions into two coreference chains: ("Protein phosphorylation", "it", "it"), and ("TPCK", "it").

In recent years, several supervised approaches have been proposed for bio-entity coreference resolution. The work can be categorized into three classes. 1) rule and feature-based models [4, 5, 6], which heavily rely on syntactic parsers to extract manually crafted features and rules. 2) hybrid models [7, 8], which combine rule-based and machine learning-based methods for biomedical coreference resolution 3) neural network-based models [6, 9, 10], which use deep learning or neural networks to solve the problem automatically with

domain-specific information integration, typically by pre-trained embeddings and some biomedical features.

Generally, the above work is summarized in Table 3. Our work is most closely related to the work of [9], while we focus on the problem that identical mentions tend to get similar or even identical representations, which further mislead to make coreferential mistakes.

Methods

Task

In an end-to-end coreference resolution system, the input is a document D with T words, and the output is a set of mention clusters. Let N be the number of possible text spans in D . We consider all possible spans up to a predefined maximum width. $START(i)$ and $END(i)$ are the start and end indices of a span i in D respectively. For each span i the system needs to assign an antecedent $a_i \in \{\epsilon, 1, \dots, i-1\}$ from all preceding spans or a dummy antecedent ϵ . The dummy antecedent represents two cases: (1) the span i is not a mention, or (2) the span i is a mention but not coreferential with any previous span. Finally, all spans that are connected by a set of antecedent predictions are grouped.

The formal definition of identical mention is as follows. Suppose the N mentions in a document D are denoted as $M = \{m_1, m_2, \dots, m_N\}$. The identical mentions are defined by: $M' = \{m_i | \exists m_j = m_i \text{ and } m_j \in M \text{ and } j \neq i\}$. For each identical mention m_i , we define its frequency as the the number of times that this mention appears in the document.

Baseline Model

In this section, we briefly describe the baseline model: [13] which we will later augment with Feature Attention mechanism.

Span Representation

Assuming vector representation of a sentence with L words as $\{x_1, x_2, \dots, x_L\}$, while x_t denotes the concatenation of fixed pre-trained word embeddings and CNN character embeddings [15] for t -th word. The Bidirectional LSTMs [16] are used to encode each x_t .

Then, the model uses the attention mechanism [17] over words in each span to learn a task-specific notion of headedness, and the final representation g_i of span i is produced by:

$$g_i = [x_{START(i)}^*, x_{END(i)}^*, \hat{x}_i, \varphi(i)]$$

where x^* is the output of Bi-LSTM and \hat{x}_i is the head embedding encoded by the head attention mechanism. $\varphi(i)$ is the feature vector encoding the width of the span.

1 Scoring

2 The scoring functions: mention score s_m and an-
3 tecedent score s_a take the span representations as in-
4 put:

$$s_m(i) = w_m \cdot FFNN_m(g_i) \quad (1)$$

$$s_a(i, j) = w_a \cdot FFNN_a([g_i, g_j, g_i \circ g_j, \varphi^*(i, j)]) \quad (2)$$

5 where w_m and w_a are the weight matrix, \circ denotes
6 element-wise multiplication, FFNN is the feed-forward
7 neural network, and $\varphi(i, j)$ is the pair-wise features
8 encoding the distance between the two spans.

9 Feature Attention

10 Domain-specific Information Integration

11 Similar as [9], to integrate domain-specific information,
12 We make the following variations: 1) For word em-
13 bedding and head word embedding, we use BioBERT
14 [18] representations pre-trained by the language model
15 BERT [19] as the input to the LSTMs. 2) We integrate
16 one more feature, mentions' grammatical numbers, to
17 span representation, i.e., whether each mention is sin-
18 gular, plural, or unknown. 3) According to the UMLS
19 semantic types, we employ MetaMapLite [1] to identify
20 the entity's MetaMap type and assign it to all tokens
21 in the entity and encode it as the span's MetaMap
22 feature.

23 Model Structure

24 To avoid the similar representations of identical men-
25 tions, we propose a context-aware attention mecha-
26 nism called Feature Attention to adaptively exploit
27 features based on context.

28 As shown in Figure 2, we use a general attention
29 mechanism that learns the importance or weight of
30 each feature based on contexts. Suppose the initial fea-
31 ture vectors is $\varphi = [\varphi_1, \varphi_2, \dots, \varphi_V]$, where φ_j indicates
32 the j -th feature and x_u is the contexts vectors gen-
33 erated by Bi-LSTM for span u (here we use the aver-
34 age of the context representation of each word in the
35 span). Then the model learns the weight of each fea-
36 ture based on the contexts, and generate new feature
37 vectors φ^* :

$$a'_j = w_{a'} \cdot FFNN(\varphi_j f(x_u)) \quad (3)$$

$$a'_{j,u} = \frac{\exp(a'_j)}{\sum_{v=1}^V \exp(a'_v)} \quad (4)$$

$$\varphi^* = \bigoplus_{j=1}^V a'_{j,u} \cdot \varphi_j \quad (5)$$

38 Where \bigoplus is the concatenation operation and $f(x_u)$ is
39 a linear function to map x_u to the same dimension with

the feature vector. $a'_{j,u}$ is the weight of each feature
based on the contexts and φ^* is the new reweighed
feature vectors.

Span Feature Attention

To use features adaptively, we apply the Feature At-
tention mechanism to the span features: span width,
grammatical number, and MetaMap entity tags.

As shown in Figure 3, a new context-aware fea-
ture vector φ^* is generated by the Feature Attention
method and the new span features are applied to up-
date the span representation, where x_i^* is the contexts
vectors generated by Bi-LSTM for span i and FA is
the Feature Attention mechanism:

$$\varphi^*(i) = FA(\varphi(i), x_i^*) \quad (6)$$

$$g_i = [x_{START(i)}^*, x_{END(i)}^*, \hat{x}_i, \varphi^*(i)] \quad (7)$$

Coreference Score

The final coreference score of span i and j shows that
(1) whether span i is a mention, (2) whether span j is
a mention and (3) whether j is an antecedent of i :

$$s(i, j) = \begin{cases} 0, & j = \varepsilon \\ s_m(i) + s_m(j) + s_a(i, j) + s_c(i, j), & j \neq \varepsilon \end{cases}$$

$$s_c(i, j) = g_i^T \omega_c g_j$$

where $s_m(i)$ is the mention score, $s_a(i, j)$ is the an-
tecedent score, $s_c(i, j)$ is a rough sketch of likely an-
tecedents and w_c is a learned weight matrix.

Experiments

Dataset and Baseline

The experiments are performed on the BioNLP Pro-
tein coreference dataset [14]. For evaluation, we em-
ployed the scorer [2] provided by the organizers to make
fair comparisons with previous work. We use the base-
lines below:

[7]: The system proposes a hybrid approach that
combines both rule-based and learning-based method.

[4]: The system develops a rule-based system for
anaphoric coreference resolution in the biomedical do-
main with simple modules derived from available sys-
tems.

[5]: The system designs a general modular frame-
work, which is based on the smorgasbord architecture
and contains multiple types of coreference types, and
allows fine-grained specification of resolution strategies
to resolve coreference.

[1]<https://metamap.nlm.nih.gov/MetaMapLite.shtml>

[2]<http://bionlp-st.dbcls.jp/CO/eval-test/>

[6]: The system presents (1) a rule-based method, which creates a set of syntactic rules or semantic constraints. (2) a neural network-based method using the LSTM network.

[9]: The system apply a state-of-the-art general system with domain-specific features for biomedical text.

Hyperparameters

We follow the same hyperparameters as in the [13]. We use a window size of 10 for the LSTM inputs. The maximum span width is 30 and only considers 50 antecedents. The model is trained up to 70 epochs with early stopping. We consider the following variations of our model:

BioNeu directly runs the [13] with integrating domain-specific information.

SFA (Span Feature Attention) uses Feature Attention mechanism to encode span feature $\varphi(i)$.

Results

Table 4 shows the performance comparison of our model with some baselines on the BioNLP development and test datasets. Overall, our SFA (span Feature Attention) model outperforms all baselines and achieves state-of-the-art performance with the 64.0% F1 score (+2.0 F1). In general, we notice that on the basis of ensuring the recall, our model performs much better precision than competitors, which indicates that after the distinct representation of mentions based on context, the noise brought by similar or identical mentions is reduced. Besides, we also find that the proposed model has a stronger generalization ability than competitors, which allows it has better scalability.

Moreover, there are also some gaps in the performance of different models. First, BioNeu, similar to [9], makes improvement compared with other neural network-based baseline models. This indicates that domain-specific information pre-trained on large-scale corpora is practical in domain-related tasks when well-designed. Second, the proposed context-aware attention model can help the system to effectively represent mentions based on contexts and make more precise predictions, which can be supported by the comparison with other neural models ([6], [9], and bioNeu). To summarize, the results suggest that the distinction of mentions based on contexts is vital for effectively resolving coreference. In this case, context-aware attention models can help achieve this goal and make accurate predictions.

Discussion

Error Analysis

The scorer^[3] provided by the organizer of the BioNLP shared task divides the results into three categories: 1) Gold: gold annotations of the dataset; 2) Predict: annotations predicted with two types: Correct and Spurious. Spurious means false annotations; 3) Missing: missing gold annotations. In order to better understand the performance of the proposed model, we analyzed these two types of errors in details: spurious link errors and missing link errors.

Figure 4 shows the detailed error statistics on the test dataset compared with some baselines ([6]-rule and [6]-neural). First, compared with [6], BioNeu performs much better in the reduction of spurious links, which indicates that the domain-specific information pre-trained on large-scale language models helps the model to predict more precisely. Second, compared with those neural network-based baselines [6]-neural and BioNeu, the proposed SFA model illustrates that the introduction of the Feature Attention mechanism further improves the model, greatly increasing the number of correct predictions and reducing the two errors. This shows that the distinction of the identical or similar mentions based on context can help the system learn more precisely and make more accurate predictions.

Identical Mention Linking Evaluation

Figure 5 and Figure 6 respectively display the performance of the BioNeu model and the SFA model on coreferential and incoreferential identical mentions with different frequencies (the number of times that the identical mention appears in the document). As there are no coreferential identical mentions with a frequency greater than 3, we only show the frequency of 2 and 3. First, The performance of the SFA model on coreferential identical mentions is better than the BioNeu model. This indicates that the distinction of identical mentions helps in the prediction of the links between them to a certain extent. Second, for incoreferential identical mentions, the SFA model performs better than the BioNeu model on all frequencies, and the superiority is significantly greater when the frequency of the identical mention is greater than 3. The superiority illustrates that, for the problem that the higher the frequency, the more difficult it is to predict, the Feature Attention mechanism does help to distinguish the identical mentions based on context, which provides further help in mention linking.

^[3]<http://bionlp-st.dbcls.jp/CO/eval-test/>

1 Mention Detection Subtask

2 To further understand the utility of the Feature Attention
3 mechanism for mention detection subtask, we list
4 the mention detection performance in Table 5. Over-
5 all, compared with [9], the performance of the pro-
6 posed SFA model is significantly increased by 6.3 F1 on
7 the BioNLP development dataset. Moreover, the SFA
8 model indeed performs much better than the BioNeu
9 in the recall scores. This indicates that, in the base-
10 line model, where there is a span not predicated as a
11 mention, the other identical spans will likely not be
12 detected as mentions due to their similar representa-
13 tions. However, in the SFA model, such false-negative
14 errors are decreased, having benefited from the Fea-
15 ture Attention mechanism that reweighs the features
16 to distinguish identical spans with different represen-
17 tations based on different contexts.

18 Remaining Problems and Future Work

19 From the results of mention detection, we can find that
20 the system is limited in the denoise of the context in-
21 formation when Feature Attention is added, which can
22 be seen from the decrease in precision. In addition, the
23 analysis of identical mention evaluation also show that
24 the system gains less improvement when the frequency
25 is less than 3. That means there are several potential
26 improvements to our model as future work, such as
27 how to adjust the setting of the FA so that the noise
28 of the context information is reduced. Moreover, consid-
29 ering that this approach is still limited when context
30 information is insufficient, we expect to utilize such in-
31 formation more fine-grained with the help of external
32 knowledge base.

33 Case Study

34 To gain further insight into how identical terms' repre-
35 sentations can be distinguished by the attention mech-
36 anism, we take the former case in Figure 1 as an ex-
37 ample to investigate the Feature Attention weights.
38 The attention weights of span features (span width,
39 grammatical number, and Metamap entity tags) are
40 shown in Figure 7. For the three "it" in the figure,
41 we can find that the first two "it" that are coreferen-
42 tial gain similar weights for the three features, where
43 the weight of span width is the highest, followed by
44 the grammatical number, and finally Metamap. While,
45 the third one has the opposite span weights to the first
46 two with a higher weight in grammatical number than
47 Metamap, though they are verbatim. This indicates
48 that the identical spans that have the same features
49 will have different span feature attention weights de-
50 pending on their contexts through the span Feature
51 Attention, thus improving the performance.

Conclusion

1 Identical mentions impose difficulties on the current
2 methods of Bio-entity coreference resolution as they
3 tend to get similar or even identical representations.
4 This problem may directly lead to wrong predictions.
5 In the paper, we focus on this issue and develop a
6 context-aware attention model named Feature Attention
7 and apply the attention mechanism in the pro-
8 cess of span representation to adaptively exploit fea-
9 tures and represent identical mentions considering dif-
10 ferent contexts. The results show that our model with
11 the Feature Attention mechanism performs reasonably
12 well in Bio-entity coreference resolution. The perfor-
13 mance is supported on the BioNLP Protein Corefer-
14 ence dataset. Moreover, as our model learns to distin-
15 guish identical mentions more effectively, it achieves
16 superior performance on the identical mention linking.
17

18 Availability of data and materials

19 The datasets used in this study are publicly available at the official website
20 of BioNLP Shared Task 2011.
21 <http://2011.bionlp-st.org/home/protein-gene-coreference-task>

22 Ethics approval and consent to participate

23 Not applicable.

24 Consent for publication

25 Not applicable.

26 Competing interests

27 The authors declare that they have no competing interests.

28 Author's contributions

29 Yufei Li conceived of the study, participated in algorithm development,
30 performed the analyses, programming, and evaluation, and drafted the
31 manuscript. Xiangyu Zhou performed data analysis, participated in the
32 algorithm development and evaluation design, and helped draft the
33 manuscript. Jie Ma and Xiaoyong Ma participated in data analysis, and
34 helped draft the manuscript. Chen li participated in the design of the study,
35 and helped draft the manuscript. All authors read and approved the final
36 manuscript.

37 Acknowledgements

38 Not applicable.

39 Funding

40 This work has been supported by the National Key Research and
41 Development Program of China (2018YFC0910404); National Natural
42 Science Foundation of China (61772409); Innovative Research Group of the
43 National Natural Science Foundation of China(61721002); Innovation
44 Research Team of Ministry of Education (IRT_17R86); Project of China
45 Knowledge Centre for Engineering Science and Technology.

46 Author details

47 ¹School of Computer Science and Technology, Xi'an Jiaotong University,
48 Xi'an, Shaanxi, P.R. China. ²National Engineering Lab for Big Data
49 Analytics, Xi'an Jiaotong University, Xi'an, Shaanxi, P.R. China. ³ Shaanxi
50 Province Key Laboratory of Satellite and Terrestrial Network Technology
51 Research and Development, Xi'an Jiaotong University, Xi'an, Shaanxi, P.R.
52 China.

53 References

- 54 1. Dai, Z., Fei, H., Li, P.: Coreference aware representation learning for
55 neural named entity recognition. In: Proceedings of the International
56 Joint Conference on Artificial Intelligence (IJCAI), Macau, pp.
57 4946–4953 (2019)

1 2. Hajishirzi, H., Zilles, L., Weld, D.S., Zettlemoyer, L.: Joint coreference
 2 resolution and named-entity linking with multi-pass sieves. In:
 3 Proceedings of the 2013 Conference on Empirical Methods in Natural
 4 Language Processing, pp. 289–299 (2013)

5 3. Weston, J., Bordes, A., Chopra, S., Rush, M.A., Merriënboer, v.B.,
 6 Joulin, A.a.: Towards ai-complete question answering: A set of
 7 prerequisite toy tasks. international conference on learning
 8 representations (2015)

9 4. Choi, M., Zobel, J., Verspoor, K.: A categorical analysis of coreference
 10 resolution errors in biomedical texts. Journal of biomedical informatics
 11 **60**, 309–318 (2016)

12 5. Kilicoglu, H., Demner-Fushman, D.: Bio-scores: a smorgasbord
 13 architecture for coreference resolution in biomedical text. PloS one
 14 **11**(3), 0148538 (2016)

15 6. Li, C., Rao, Z., Zheng, Q., Zhang, X.: A set of domain rules and a
 16 deep network for protein coreference resolution. Database **2018** (2018)

17 7. D’Souza, J., Ng, V.: Anaphora resolution in biomedical literature: a
 18 hybrid approach. In: Proceedings of the ACM Conference on
 19 Bioinformatics, Computational Biology and Biomedicine, pp. 113–122
 20 (2012)

21 8. Li, L., Jin, L., Jiang, Z., Zhang, J., Huang, D.: Coreference resolution
 22 in biomedical texts. In: 2014 IEEE International Conference on
 23 Bioinformatics and Biomedicine (BIBM), pp. 12–14 (2014).
 24 doi:10.1109/BIBM.2014.6999392

25 9. Trieu, H.-L., Nguyen, N.T., Miwa, M., Ananiadou, S.: Investigating
 26 domain-specific information for neural coreference resolution on
 27 biomedical texts. In: Proceedings of the BioNLP 2018 Workshop, pp.
 28 183–188 (2018)

29 10. Trieu, H.-L., Nguyen, A.-K.D., Nguyen, N., Miwa, M., Takamura, H.,
 30 Ananiadou, S.: Coreference resolution in full text articles with bert and
 31 syntax-based mention filtering. In: Proceedings of The 5th Workshop
 32 on BioNLP Open Shared Tasks, pp. 196–205 (2019)

33 11. Ji, G., Liu, K., He, S., Zhao, J., et al.: Distant supervision for relation
 34 extraction with sentence-level attention and entity descriptions. In:
 35 AAAI, vol. 3060 (2017)

36 12. Luo, L., Yang, Z., Yang, P., Zhang, Y., Wang, L., Lin, H., Wang, J.:
 37 An attention-based bilstm-crf approach to document-level chemical
 38 named entity recognition. Bioinformatics **34**(8), 1381–1388 (2018)

39 13. Lee, K., He, L., Zettlemoyer, L.: Higher-order coreference resolution
 40 with coarse-to-fine inference. In: Proceedings of the 2018 Conference
 41 of the North American Chapter of the Association for Computational
 42 Linguistics: Human Language Technologies, Volume 2 (Short Papers),
 43 pp. 687–692. Association for Computational Linguistics, New Orleans,
 44 Louisiana (2018). doi:10.18653/v1/N18-2108.
 45 https://www.aclweb.org/anthology/N18-2108

46 14. Nguyen, N., Kim, J.-D., Tsujii, J.: Overview of the protein coreference
 47 task in bionlp shared task 2011. In: Proceedings of the BioNLP Shared
 48 Task 2011 Workshop, pp. 74–82 (2011)

49 15. Santos, C.D., Zadrozny, B.: Learning character-level representations
 50 for part-of-speech tagging. In: Proceedings of the 31st International
 51 Conference on Machine Learning (ICML-14), pp. 1818–1826 (2014)

52 16. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural
 53 computation **9**(8), 1735–1780 (1997)

54 17. Bahdanau, D., Cho, K., Bengio, Y.: Neural Machine Translation by
 55 Jointly Learning to Align and Translate. (2015). 3rd International
 56 Conference on Learning Representations, ICLR 2015 ; Conference date:
 57 07-05-2015 Through 09-05-2015

58 18. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., Kang, J.:
 59 Biobert: a pre-trained biomedical language representation model for
 60 biomedical text mining. Bioinformatics **36**(4), 1234–1240 (2020)

61 19. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training
 62 of deep bidirectional transformers for language understanding. In:
 63 Proceedings of the 2019 Conference of the North American Chapter of
 64 the Association for Computational Linguistics: Human Language
 65 Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186.
 66 Association for Computational Linguistics, Minneapolis, Minnesota
 67 (2019). doi:10.18653/v1/N19-1423.
 68 https://www.aclweb.org/anthology/N19-1423

69 **Tables**
 70 **Figures**

Table 1 Statistics of identical mentions throughout the BioNLP dataset. Each represents its percentage of all mentions.

	Train (%)	Dev (%)
Nonidentical	41.9	48.4
Identical	58.1	51.6
Coreferential	6.9	2.8
Incoreferential	93.1	97.2
Frequency	Train (%)	Dev (%)
<10	22.0	24.3
10–100	13.7	21.6
>100	64.3	54.1
Type	Train (%)	Dev (%)
NP	18.3	13.9
NN/NNS	5.6	2.1
WDT	20.8	25.6
PRP	7.6	7.2
PRP\$	16.6	18.4
IN	27.1	28.6
Others	4.0	4.2
Top-5	that, which, its, their, it	

Table 2 Statistics of identical mentions in per document. Each represents its percentage of the total number of documents

	Train (%)	Dev (%)
Without identical mentions	57.1	57.3
With identical mentions	42.9	42.7
Num	Train (%)	Dev (%)
2	50.1	45.3
3	13.4	18.8
4	17.2	18.7
5	7.9	3.1
>5	11.4	14.1

Table 3 Coreference resolution performance comparison by the average F1 value of three evaluation metrics.

System	Dataset	Dev-F1	Test-F1	Feature-based Rule-based	Hybrid	Neural
[7]	BioNLP	62.4	60.9		✓	
[8]	BioNLP	68.6	/		✓	
[4]	BioNLP	63.9	48.1	✓		
[5]	BioNLP	67.5	/	✓		
[6]	BioNLP	72.2	62.0	✓		✓
[9]	BioNLP, CRAFT	34.9, 36.0	51.2, 59.2			✓
[10]	CRAFT	50.5	65.7			✓

Table 4 The performance of protein coreference resolution with different models on two evaluation datasets of BioNLP.

	dev			test		
	P	R	F1	P	R	F1
[7]	59.9	77.1	67.4	55.6	67.2	60.9
[4]	63.4	64.4	63.9	46.3	50.0	48.1
[5]	72.4	63.2	67.5	/	/	/
[6]-rule	68.8	76.0	72.2	60.2	63.8	62.0
[6]-neural	60.4	61.9	61.2	54.9	58.0	56.4
[9]	71.7	56.7	63.1	55.6	47.5	51.2
BioNeu	78.1	61.8	69.0	62.9	53.9	58.1
SFA	71.9	65.8	68.7	69.0	59.5	64.0

Table 5 The performance of mention detection with different models on the development datasets of BioNLP.

	P	R	F1
[9]	82.0	66.3	73.3
BioNeu	84.1	73.1	78.2
SFA	83.4	76.1	79.6

[width=3.1 in]fig1.png

Figure 1 Example of linking errors of identical mentions affected by similar representations. Avoiding this error requires context-aware Feature Attention mechanism.

[width=1.5 in]fig2.png

Figure 2 The Feature Attention model. The model learns to weigh each feature based on contexts.

[width=3.1 in]fig3.png

Figure 3 The model of computing the span embedding representations.

[width=3.1 in]fig4.png

Figure 4 Detailed error analysis compared with some baselines

[width= 3.1 in]fig5.png

Figure 5 The performance of the two models on coreferential identical mentions with different frequencies within the document.

[width=3.1 in]fig6.png

Figure 6 The performance of the two models on incoreferential identical mentions with different frequencies within the document.

[width=2 in]fig7.png

Figure 7 An example of span Feature Attention weights in different features. Each row shows the attention weights of all the features of the span.

Figures

[Protein phosphorylation] has an important role in the regulation of these two factors: (1) **[it]** induces the transactivating capacity of the AP-1 protein c-Jun; and (2) **[it]** is involved in the release of the cytoplasmic inhibitor, I kappa B, from NF-kappa B, allowing translocation of the latter into the nucleus. ... Whereas **[TPCK]** inhibits phosphorylation of I kappa B and, consequently, activation of NF-kappa B, **[it]** markedly enhances the activity of JNK, the MAP kinase-related kinase that phosphorylates the transactivation domain of c-Jun.

Figure 1

Example of linking errors of identical mentions affected by similar representations. Avoiding this error requires context-aware Feature Attention mechanism.

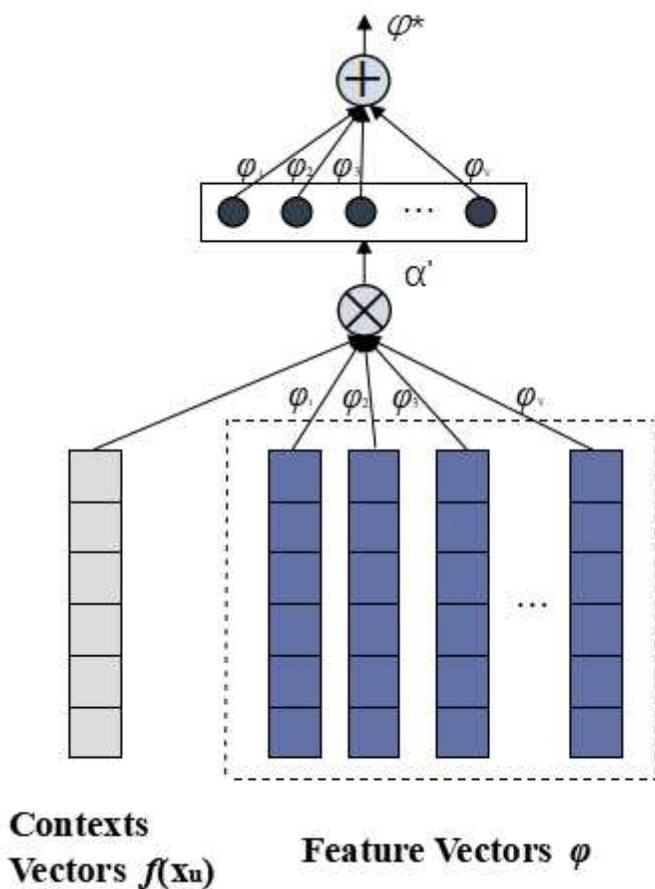


Figure 2

The Feature Attention model. The model learns to weigh each feature based on contexts.

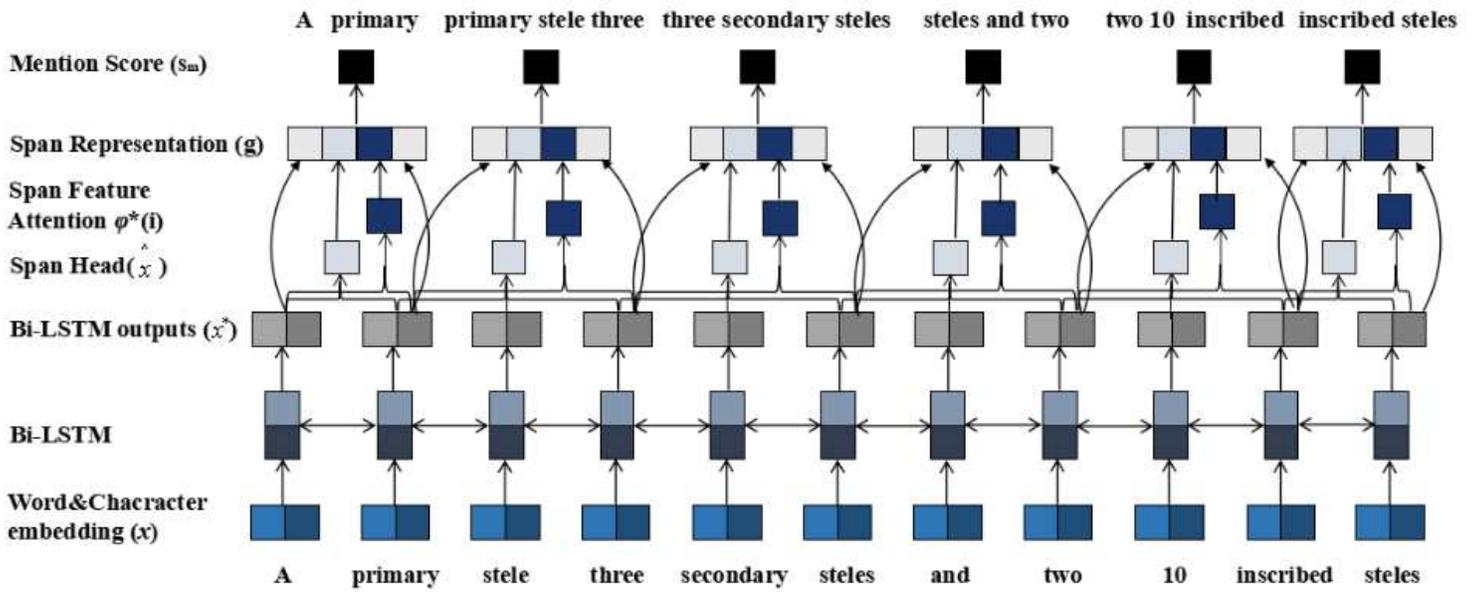


Figure 3

The model of computing the span embedding representations.

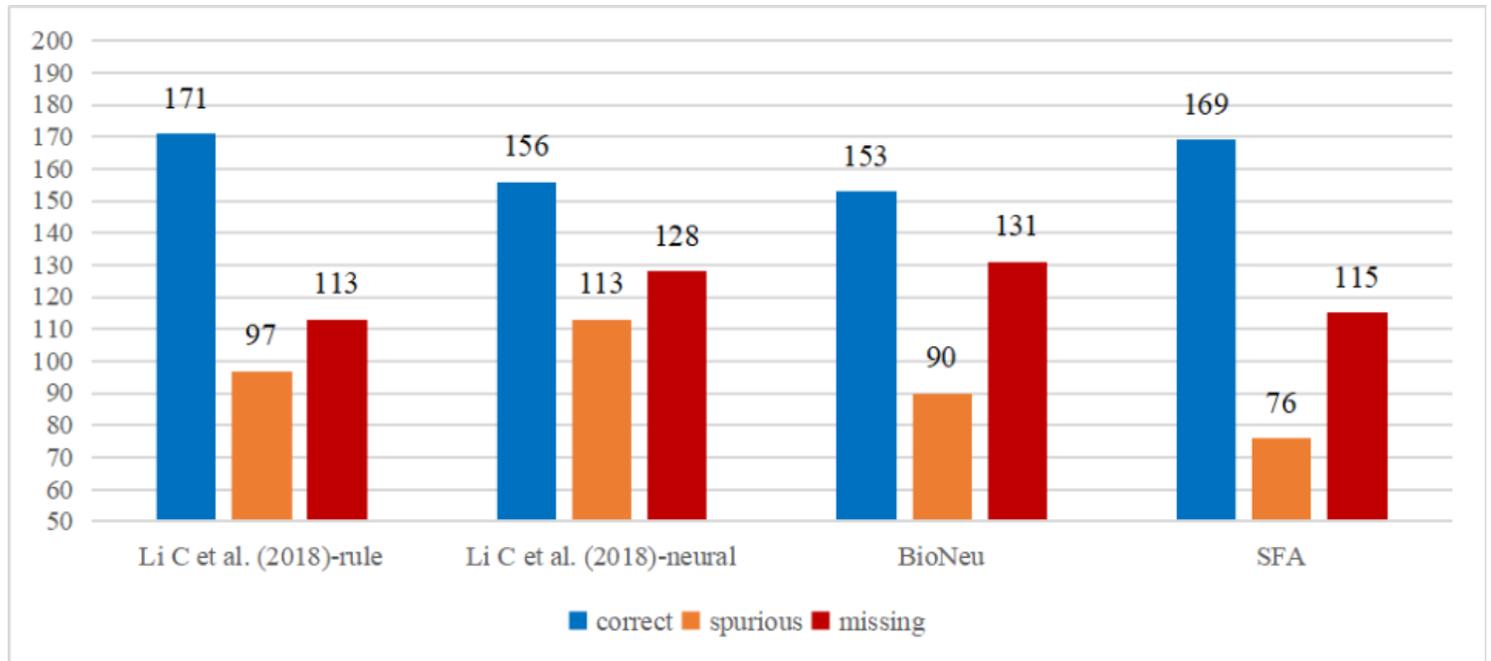


Figure 4

Detailed error analysis compared with some baselines

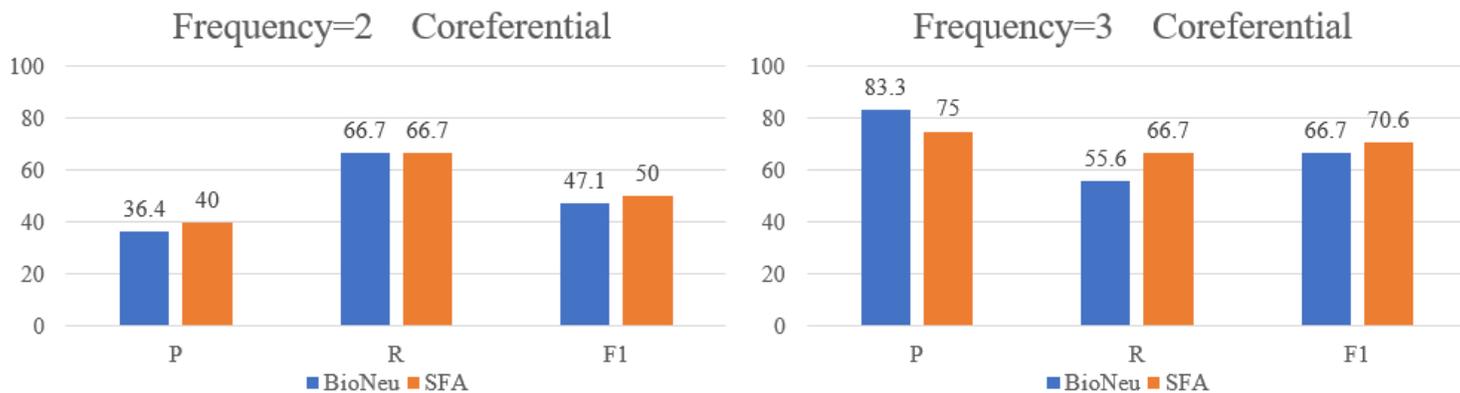


Figure 5

The performance of the two models on coreferential identical mentions with different frequencies within the document.

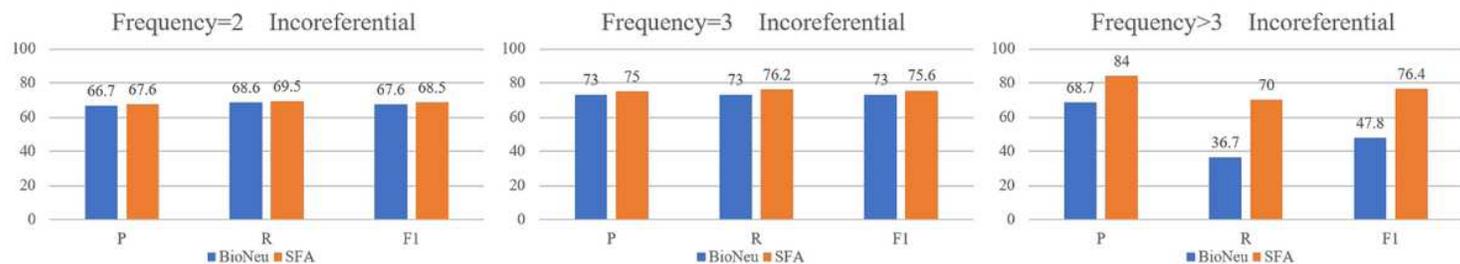


Figure 6

The performance of the two models on incoreferential identical mentions with different frequencies within the document.

Protein phosphorylation has an important role in the regulation of these two factors: (1) **it** induces the transactivating capacity of the AP-1 protein c-Jun; and (2) **it** is involved in the release of the cytoplasmic inhibitor, I kappa B, from NF-kappa B, allowing translocation of the latter into the nucleus. ... Whereas **TPCK** inhibits phosphorylation of I kappa B and, consequently, activation of NF-kappa B, **it** markedly enhances the activity of JNK, the MAP kinase-related kinase that phosphorylates the transactivation domain of c-Jun.



Figure 7

An example of span Feature Attention weights in different features. Each row shows the attention weights of all the features of the span.