Journal of Cheminformatics

RESEARCH ARTICLE

Open Access

LSTMVoter: chemical named entity recognition using a conglomerate of sequence labeling tools

Wahed Hemati* and Alexander Mehler

Abstract

Background: Chemical and biomedical *named entity recognition* (NER) is an essential preprocessing task in *natural language processing*. The identification and extraction of named entities from scientific articles is also attracting increasing interest in many scientific disciplines. Locating chemical named entities in the literature is an essential step in chemical text mining pipelines for identifying chemical mentions, their properties, and relations as discussed in the literature. In this work, we describe an approach to the BioCreative V.5 challenge regarding the recognition and classification of chemical named entities. For this purpose, we transform the task of NER into a sequence labeling problem. We present a series of sequence labeling systems that we used, adapted and optimized in our experiments for solving this task. To this end, we experiment with hyperparameter optimization. Finally, we present `LSTMVoter`, a two-stage application of *recurrent neural networks* that integrates the optimized sequence labelers from our study into a single ensemble classifier.

Results: We introduce `LSTMVoter`, a bidirectional *long short-term memory* (LSTM) tagger that utilizes a conditional random field layer in conjunction with attention-based feature modeling. Our approach explores information about features that is modeled by means of an attention mechanism. `LSTMVoter` outperforms each extractor integrated by it in a series of experiments. On the BioCreative IV chemical compound and drug name recognition (CHEMDNER) corpus, `LSTMVoter` achieves an F1-score of 90.04%; on the BioCreative V.5 chemical entity mention in patents corpus, it achieves an F1-score of 89.01%.

Availability and implementation: Data and code are available at https://github.com/texttechnologylab/LSTMVoter.

Keywords: BioCreative V.5, CEMP, CHEMDNER, BioNLP, Named entity recognition, Deep learning, LSTM, Attention mechanism

Introduction

In order to advance the fields of biological, chemical and biomedical research, it is important to stay on the cutting edge of research. However, given the rapid development of the disciplines involved, this is difficult, as numerous new publications appear daily in biomedical journals. In order to avoid repetition and to contribute at least at the level of current research, researchers rely on published information to inform themselves about the latest research developments. There is therefore a growing interest in improved access to information on biological, chemical and biomedical data described in scientific articles, patents or health agency reports. In this context, improved access to chemical and drug name mentions in document repositories is of particular interest: it is these entity types that are most often searched for in the PubMed [1] database. To achieve this goal, a fundamental preprocessing step is to automatically identify biological and chemical mentions in the underlying documents. Based on this identification, downstream NLP tasks such as the recognition of interactions between drugs and proteins, of side effects of chemical compounds and their

*Correspondence: hemati@em.uni-frankfurt.de
Text Technology Lab, Goethe-University Frankfurt, Robert-Mayer-Straße 10, 60325 Frankfurt am Main, Germany

associations with toxicological endpoints or the investigation of information on metabolic reactions can be carried out.

For these reasons, NLP initiatives have been launched in recent years to address the challenges of identifying biological, chemical and biomedical entities. One of these initiatives is the BioCreative series, which focuses on biomedical text mining. BioCreative is a "Challenge Evaluation", in which the participants are given defined text mining or information extraction tasks in the biomedical and chemical field. These tasks include *GeneMention detection (GM)* [2, 3], *Gene Normalization (GN)* [3–5], *Protein–Protein Interaction (PPI)* [6], *Chemical Compound and Drug Name Recognition (CHEMDNER)* [7, 8] and *Chemical Disease Relation Extraction* [9, 10] tasks.

The current *BioCreative V.5* task consists of two offline tasks, namely *Chemical Entity Mention in Patents (CEMP)* and *Gene and Protein Related Object Recognition (GPRO)*. CEMP requires the detection of chemical named entity mentions. The task requires detecting the start and end indices corresponding to chemical entities. The GPRO task requires identifying mentions of gene and protein related objects in patent titles and abstracts [11]. In this work, we focus on the CEMP task. The CEMP task is an abstraction of the common named entity recognition (NER) tasks, which can be reduced to a sequence labeling problem, where the sentences are represented as sequences of tokens. The task is then to tag chemical entity mentions in these sequences. The settings of the CEMP task are similar to the chemical entity mention recognition (CEM) subtask of CHEMDNER challenge in BioCreative IV [7]. Therefore, we addressed both tasks and their underlying corpora in our experiments. Note that the current article describes an extension of previous work [12].

The article is organized as follows: First we describe our methodical apparatus and resources. This includes the data and corpora used in our experiments. Then, we introduce state-of-the-art tools for NER and explain how we adapted them to perform the CEMP task. Next, we present a novel tool for combining NER tools, that is, the so-called `LSTMVoter`. Finally, we present our results, conclude and discuss further work.

Materials and methods

In this section, we first describe the datasets used in our experiments. Then, the two-stage application of `LSTMVoter` is introduced.

Datasets

In our experiments, two corpora of the BioCreative Challenge were used: the CHEMDNER Corpus [13] and the CEMP Corpus [14].

The CHEMDNER corpus consists of 10,000 abstracts of chemistry-related journals published in 2013. Each abstract was human annotated for chemical mentions. The mentions were assigned to one of seven different subtypes (ABBREVIATION, FAMILY, FORMULA, IDENTIFIER, MULTIPLE, SYSTEMATIC, and TRIVIAL). The BioCreative organizer divided the corpus into training (3500 abstracts), development (3500 abstracts) and test (3000 abstracts) sets.

For CEMP task, the organizers of *BioCreative V.5* provided a corpus of 30,000 patent abstracts from patents published between 2005 and 2014. These abstracts are divided into training (21,000 abstracts) and test (9000 abstracts) sets. The corpus is manually annotated with chemical mentions. For the construction of the CEMP corpus the annotation guidelines of CHEMDNER were used. Therefore, CEMP contains the same seven chemical mention subtypes as CHEMDNER. Table 1 shows the number of instances for both corpora for each of these subtypes.

Both corpora were enriched with additional linguistic features. For this, multiple preprocessing steps were applied on each set including sentence splitting, tokenization, lemmatization and fine-grained morphological tagging by means of Stanford CoreNLP [15] and TextImager [16]. In addition, tokens were split on non-alphanumeric characters, as this variant brought a performance increase. Since the chemical mention detection task can be reduced to a sequence labeling problem, the corpora were converted into a sequence structure. To this end, a sequence of documents with sequences of sentences each containing a sequence of tokens was constructed and transformed according to a TSV format. Each word and its associated features are in one line separated by tabs. Sentences are separated by an empty line. For the labeling of the mentions, the IOB tagging scheme [17] was used (I = *inside of an entity*, O = *outside of an entity*, B

Table 1 Number of instances for each subtype of CEMP and CHEMDNER corpus

Annotation	CEMP	CHEMDNER
Abbreviation	1373	9059
Family	36,238	8313
Formula	6818	8585
Identifier	278	1311
Multiple	418	390
Systematic	28,580	13,472
Trivial	25,927	17,802
No class	0	72
Total count	99,632	59,004

= *beginning of an entity*). IOB allows the annotation of entities that span multiple tokens, where the beginning and the end of the entity is marked. This enables models to learn transition probability. `LSTMVoter` needs four datasets for the training process. Two pairs of training and development sets are required. Each pair is needed in one of the two stages of `LSTMVoter` (see section "System description"). Therefore, we divided the training set of CEMP into two series of training, development and test sets (each half of the original training set was split according to the pattern 60%/20%/20%), where the first series is used for stage one, and the second for stage two. For the CHEMDNER corpus the available training and development sets were joined and split into training and development sets according to the schema 80%/20%—as before, we distinguish two such series. For evaluating our classifiers with respect to CHEMDNER, the test set provided by the organizers of the challenge was used. For the following experiments we used the corpora described as so far.

System description

In this section we describe our system. Our approach implements a two-stage application of long short-term memory (LSTM) using a conglomerate of sequence labelers for the detection of chemical mentions.

In the first stage, we trained and optimized five tools for NER for tackling this task, namely *Stanford Named Entity Recognizer* [18], *MarMoT* [19], *CRF++* [20], *MITIE* [21] and *Glample* [22]. For each of them, we optimized the corresponding hyperparameter settings. Generally speaking, hyperparameter tuning is a challenging task in machine learning. The optimal set of hyperparameters depends on the model, the dataset and the domain [23]. Our experiments focused on optimizing the hyperparameters of each NER system independently, which led to a noticeable increase in F-score compared to the default settings. For each NER, we performed the Tree-structured Parzen Estimator (TPE) [24] with 200 iterations. The results of the best performing model for each of these NER is listed in Table 2.

The NER tools are more or less independent of each other in the sense that one can find a subset of test cases that are correctly processed by one of them, but not by another. Therefore, combining these NERs is a promising candidate for increasing performance. We started with computing combinations of these NERs by means of a simple majority vote [25], where the target label is selected, that is assigned by the majority of classifiers. Our experiments show that a simple majority vote brings no gain in performance compared to the best performing reference systems being examined in our study (see Table 2). Thus, we developed a two-stage model, the so-called `LSTMVoter`, which trains a recurrent neural network (RNN) with attention mechanism to learn the best combination of the underlying sequence labeling tools from stage one.

In the second stage, we combine the sequence labelers of stage one with two bidirectional *long short-term memory* (LSTM) networks with attention mechanism and a conditional random field (CRF) network to form `LSTMVoter`. The architecture of `LSTMVoter` is illustrated in Fig. 1. The core of `LSTMVoter` is based on [22].

LSTM networks are a type of RNN [26]. RNN allow the computation of fixed-size vector representations for sequences of arbitrary length. An RNN is, so to speak, a function that reads an input sequence $x_1, \ldots, x_n$ of length n and produces an output vector h_n, which depends on the entire input sequence. Though, in theory, an RNN is capable of capturing long-distance dependencies in the input sequence, in practice, they may fail due to the problem of vanishing gradients [27, 28]. On the other hand, LSTMs include a memory cell, which can maintain information in memory for long periods of time [29, 30]. This enables finding and exploiting long range dependencies in the input sequences to cope with the problem of vanishing gradients. Figure 2 illustrates an LSTM memory cell, which is implemented as follows:

$$
\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)\\
c_t &= f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)\\
h_t &= o_t \tanh(c_t)
\end{aligned}
$$

where x_t is the input vector (e.g. word embedding) at time t. h_t is the hidden state vector, also called output vector, that contains information at time t and all time steps before t. σ is the logistic sigmoid function [31]. Input gate i, forget gate f, output gate o and cell vector c are of the same size as the hidden state vector h. W_{hi}, W_{hf}, W_{hc} and W_{ho} are the weight matrices for the hidden state h_t. W_{xi}, W_{xf}, W_{xc} and W_{xo} denote the weight matrices of different gates for input x_t.

For `LSTMVoter`, we apply an LSTM to sequence tagging. Additionally, as proposed by [32], we utilize bidirectional LSTM networks. Figure 3 illustrates a bidirectionalLong short-term memory (Bi-LSTM) network, where the input sequence (*Treatment with haloperidol or reserpine ...*) and the output sequence (*O, O, B-Trivial, O, B-Trivial, ...*) are fed as a training instance to a Bi-LSTM. In Bi-LSTMs, the input sequence is presented forward and backward to two separate hidden states to capture past and future information. To efficiently make use of past features (via forward states) and future features

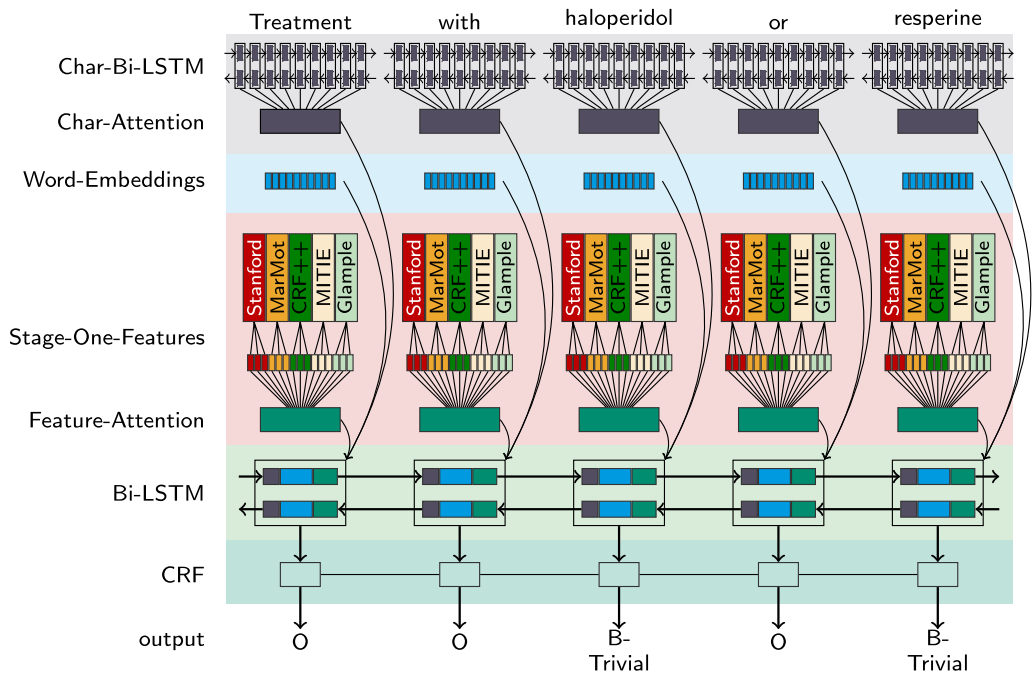

Fig. 1 Architecture of `LSTMVoter`

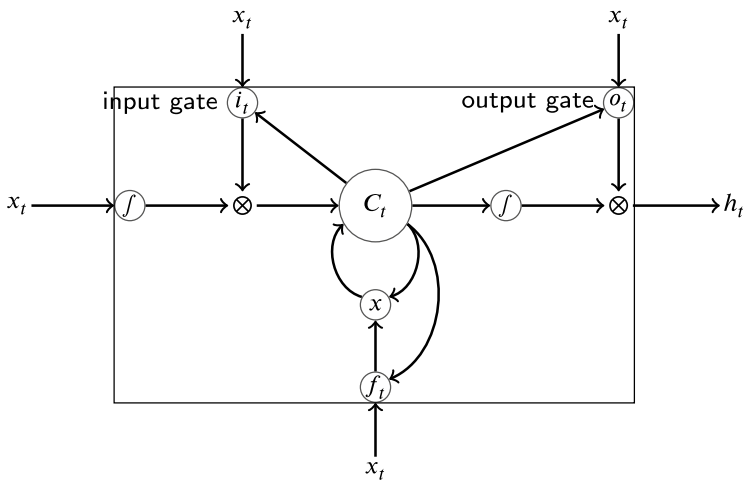

Fig. 2 A long short-term memory cell

(via backward states) for a specific time frame, the two hidden states are concatenated to form the final output. In the final output of a Bi-LSTM, all information of the complete sequence is compressed into a fixed-length hidden state vector, which may result in information loss. To overcome this information loss, an attention mechanism is introduced, which partially fixes the problem.

The method of attention mechanism has recently gained popularity in image caption generation [33], visual question answering [34] and language modeling tasks [35–38]. The attention mechanism plugs a context vector on top of a layer, which enables to take all cells' outputs as input to compute a probability distribution. This enables to capture global information rather then to infer based on one output vector.

For `LSTMVoter`, we utilized Bi-LSTM with attention mechanism to model character-level features (see Fig. 1, *Char-Bi-LSTM*). Character-level features in chemical

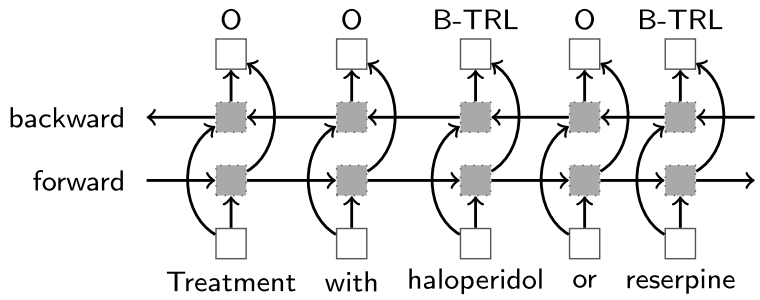

Fig. 3 A bidirectional LSTM network

Table 2 Comparison of annotators trained and tested on CEMP and CHEMDNER corpora measured by precision (P), recall (R), f1-score (F1)

System	CEMP			CHEMDNER		
	P	R	F	P	R	F
Stanford NER	0.85	0.80	0.82	0.82	0.83	0.82
MarMoT	0.87	0.86	0.86	0.85	0.85	0.85
CRF++	0.77	0.73	0.73	0.74	0.71	0.73
MITIE	0.65	0.65	0.65	0.62	0.61	0.62
Glample	0.76	0.79	0.77	0.82	0.84	0.83
Majority vote	0.78	0.79	0.78	0.70	0.76	0.73
`LSTMVoter`	0.90	0.88	**0.89**	0.91	0.90	**0.90**

Bold was intended to compare LSTMVoter to the best reference tool. Bold now shows the system with the highest F-Score, which is LSTMVoter

named entities contain rich structure information, such as prefix, suffix and n-grams. Unlike previous methods [39–41], character-level features do not have to be defined manually, rather they can be learned during training. Unlike [22], who encodes the entire character sequence into a fixed-size vector for each word, we utilize the character-level attention mechanism introduced by [36]. This has the advantage, that by using the attention mechanism, the model is able to dynamically decide how much information and which part of a token to use.

In addition to the character-level features, we implemented word embeddings into our model to capture dependencies between words (see Fig. 1, *Word-Embeddings*). For this, we evaluated various methods, namely GloVe [42], Dependency-Based embeddings [43, 44] trained on the English Wikipedia, and word2vec [45] trained on the English Wikipedia and a biomedical scientific literature corpus containing PubMed abstracts and full texts. In our experiments, the word2vec model trained on biomedical scientific literature gave the best results.

To utilize the results of the NERs from stage one, we encode the respective results of the NERs into one-hot vectors concatenated to a feature vector (see Fig. 1, *Stage-One-Features*). An attention mechanism is placed on the feature vector. By creating a probability distribution through the attention mechanism, `LSTMVoter` learns how to weight each result of the NERs from stage one. With the attention vector it is even possible to determine for each element of a sequence how important the individual partial results from stage one were. This has the advantage that the model is no longer a black box, but can be interpreted as to how important the individual results from stage one were.

All previous elements of `LSTMVoter` encode word-based information. Another Bi-LSTM is used to learn relationships between these word-based information (see Fig. 1, *Bi-LSTM*).

To deal with the independent label output problem, we utilize the output vector as elements. For this we combine the Bi-LSTM layer with a linear-chain CRF (see Fig. 1, *CRF*). Linear-chain CRFs define the conditional probability of a state sequence to be:

$$P(y|x) = \frac{1}{Z_x} exp\left(\sum_{j=1}^{n}\sum_{m=1}^{l}\lambda_m f_m(y_{j-1}, y_j, x, j)\right)$$

where Z_x is the normalization factor that makes the probability of all state sequences sum to one; $f_m(y_{j-1}, y_j, x, j)$ is

a feature function, and λ_m is a learned weight associated with feature f_m. Feature functions measure the aspect of a state transition, $y_{j-1}, y_j \rightarrow y_t$, and the entire observation sequence, x, centered at the current time step, j. Large positive values for λ_m indicate a preference for such an event, whereas large negative values make the event unlikely.

Finally, to optimize the hyperparameters, the Tree Structure Parzen estimator was used.

Results

This section presents the results of our experiments for the chemical named entity recognition on CEMP and CHEMDNER corpus. For evaluation the BioCreative Team has specified standard evaluation statistics, namely precision (P), recall (R) and F1-score (F) [14]. For each sequence labeling tool, the hyperparameters were optimized using Tree Structure Parzen Estimators, which led to a noticeable increase of performance. For example, in the optimization process of CRF++, the difference between the worst to the best performer is 65%. The results show the need for machine learning algorithms to perform hyperparameter optimization.

Table 2 shows the comparison of annotators trained on CEMP and CHEMDNER corpus. The results listed are those obtained after the hyperparameter optimization described in the methods section, which were trained, optimized and tested on the corpora described in this section. Each sequence labeling system classifies a different subset correctly. The combination of sequence labelling systems in a majority vote did not improve performance and is even below the best sequence labelling systems. In contrast, `LSTMVoter` increases the performance and performs best in our experiments.

Conclusions

In this work, we compared a set of sequence labeling systems. We trained and optimized every sequence labeling system to detect chemical entity mention by means the TPE. We showed that optimizing hyperparameter can be crucial. One sequence labeling system in our experiments gained an improvement of more than 65 %. We showed that a naive majority vote does not bring any improvement. For this reason, we introduced and evaluated `LSTMVoter`, a two-stage tool for combining underlying sequence modeling tools (as given by the NER of our comparative study). `LSTMVoter` achieved an improvement of up to 5 % compared to the best reference systems examined in our study. This two-level classifier appears to be capable of being further developed and improved by feeding it with the output of additional sequence labelling systems. In any event, our results and those of the other participants of BioCreative V.5 Task show that the task of NER of chemical entities has not been sufficiently solved yet. For a better recognition, a larger corpus should be generated so that today's popular deep learning algorithms can work on this data. A kind of human-in-the-loop architecture for automatic annotation and intellectual rework would also be helpful at this point in order to successively increase and improve the amount of data.

Abbreviations

Bi-LSTM: bidirectional long short-term memory; CEM: chemical entity mention recognition; CEMP: chemical entity mention in patents; CHEMDNER: chemical compound and drug name recognition; CRF: conditional random field; F: F1-score; GM: gene mention detection; GN: gene normalization; GPRO: gene and protein related object recognition; LSTM: long short-term memory; NER: named entity recognition; P: precision; PPI: protein–protein interaction; R: recall; RNN: recurrent neural network; TPE: tree-structured Parzen estimator.

Author's contributions

WH, and AM conceived the study, WH carried out the implementation. WH and AM wrote the manuscript. All contributed to the intellectual evolution of this project. Both authors have read and approved the final manuscript.

Authors' information

Not applicable.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data and code are available at https://github.com/texttechnologylab/LSTMVoter

Funding

This work was funded by the Federal Ministry of Education and Research (BMBF) via the research project CEDIFOR (https://www.cedifor.de/) and by the German Research Foundations (DFG) as part of the BIOfid project (DFG-326061700) (https://www.biofid.de/de/)

Publisher's Note

Received: 27 July 2018 Accepted: 27 December 2018
Published online: 10 January 2019

References

1. PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/. Accessed 19 July (2018)
2. Smith L, Tanabe LK, nee Ando RJ, Kuo C-J, Chung I-F, Hsu C-N, Lin Y-S, Klinger R, Friedrich CM, Ganchev K, Torii M, Liu H, Haddow B, Struble CA, Povinelli RJ, Vlachos A, Baumgartner WA, Hunter L, Carpenter B, Tsai RT-H, Dai H-J, Liu F, Chen Y, Sun C, Katrenko S, Adriaans P, Blaschke C, Torres R, Neves M, Nakov P, Divoli A, Maña-López M, Mata J, Wilbur WJ (2008) Overview of biocreative II gene mention recognition. Genome Biol 9(2):2. https://doi.org/10.1186/gb-2008-9-s2-s2
3. Hirschman L, Yeh A, Blaschke C, Valencia A (2005) Overview of biocreative: critical assessment of information extraction for biology. BMC Bioinform 6(1):1. https://doi.org/10.1186/1471-2105-6-S1-S1
4. Morgan AA, Lu Z, Wang X, Cohen AM, Fluck J, Ruch P, Divoli A, Fundel K, Leaman R, Hakenberg J, Sun C, Liu H-H, Torres R, Krauthammer M, Lau

WW, Liu H, Hsu C-N, Schuemie M, Cohen KB, Hirschman L (2008) Overview of biocreative II gene normalization. Genome Biol 9(2):3. https://doi.org/10.1186/gb-2008-9-s2-s3
5. Lu Z, Kao H-Y, Wei C-H, Huang M, Liu J, Kuo C-J, Hsu C-N, Tsai RT-H, Dai H-J, Okazaki N, Cho H-C, Gerner M, Solt I, Agarwal S, Liu F, Vishnyakova D, Ruch P, Romacker M, Rinaldi F, Bhattacharya S, Srinivasan P, Liu H, Torii M, Matos S, Campos D, Verspoor K, Livingston KM, Wilbur WJ (2011) The gene normalization task in biocreative III. BMC Bioinform 12(8):2. https://doi.org/10.1186/1471-2105-12-S8-S2
6. Krallinger M, Vazquez M, Leitner F, Salgado D, Chatr-aryamontri A, Winter A, Perfetto L, Briganti L, Licata L, Iannuccelli M, Castagnoli L, Cesareni G, Tyers M, Schneider G, Rinaldi F, Leaman R, Gonzalez G, Matos S, Kim S, Wilbur WJ, Rocha L, Shatkay H, Tendulkar AV, Agarwal S, Liu F, Wang X, Rak R, Noto K, Elkan C, Lu Z, Dogan RI, Fontaine J-F, Andrade-Navarro MA, Valencia A (2011) The protein–protein interaction tasks of biocreative III: classification/ranking of articles and linking bio-ontology concepts to full text. BMC Bioinform 12(8):3. https://doi.org/10.1186/1471-2105-12-S8-S3
7. Krallinger M, Leitner F, Rabal O, Vazquez M, Oyarzabal J, Valencia A (2015) CHEMDNER: the drugs and chemical names extraction challenge. J Cheminform 7(1):1
8. Krallinger M, Rabal O, Lourenço A, Perez M, Rodríguez GP, Vázquez M, Leitner F, Oyarzabal J, Valencia A (2015) Overview of the chemdner patents task. In: Proceedings of the 5th BioCreative challenge evaluation workshop
9. Li J, Sun Y, Johnson RJ, Sciaky D, Wei C-H, Leaman R, Davis AP, Mattingly CJ, Wiegers TC, Lu Z (2016) Biocreative V CDR task corpus: a resource for chemical disease relation extraction. J Biol Databases Curation 2016:068. https://doi.org/10.1093/database/baw068
10. Wei C-H, Peng Y, Leaman R, Davis AP, Mattingly CJ, Li J, Wiegers TC, Lu Z (2016) Assessing the state of the art in biomedical relation extraction: overview of the biocreative V chemical-disease relation (CDR) task. Database 2016:032. https://doi.org/10.1093/database/baw032
11. Krallinger M, Pérez-Pérez M, Pérez-Rodríguez G, Blanco-Míguez A, Fdez-Riverola F, CapellaGutierrez S, Lourenço A, Valencia A (2017) The biocreative v.5 evaluation workshop: tasks, organization, sessions and topics. In: Proceedings of the BioCreative V.5 challenge evaluation workshop, 8–10
12. Hemati W, Mehler A, Uslu T (2017) CRFVoter: chemical entity mention, gene and protein related object recognition using a conglomerate of CRF based tools. In: BioCreative V.5. Proceedings
13. Krallinger M, Rabal O, Leitner F, Vazquez M, Salgado D, Lu Z, Leaman R, Lu Y, Ji D, Lowe DM, Sayle RA, Batista-Navarro RT, Rak R, Huber T, Rocktäschel T, Matos S, Campos D, Tang B, Xu H, Munkhdalai T, Ryu KH, Ramanan S, Nathan S, Žitnik S, Bajec M, Weber L, Irmer M, Akhondi SA, Kors JA, Xu S, An X, Sikdar UK, Ekbal A, Yoshioka M, Dieb TM, Choi M, Verspoor K, Khabsa M, Giles CL, Liu H, Ravikumar KE, Lamurias A, Couto FM, Dai HJ, Tsai RTH, Ata C, Can T, Usié A, Alves R, Segura-Bedmar I, Martínez P, Oyarzabal J, Valencia A (2015) The CHEMDNER corpus of chemicals and drugs and its annotation principles. J Cheminform 7(1):2
14. Pérez-Pérez M, Rabal O, Pérez-Rodríguez G, Vazquez M, Fdez-Riverola F, Oyarzabal J, Valencia A, Lourenço A, Krallinger M (2017) Evaluation of chemical and gene/protein entity recognition systems at biocreative v.5: the cemp and gpro patents tracks. In: Proceedings of the BioCreative V.5 challenge evaluation workshop, 11–18
15. Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D (2014) The Stanford CoreNLP natural language processing toolkit. In: Association for computational linguistics (ACL) system demonstrations, pp 55–60
16. Hemati W, Uslu T, Mehler A (2016) TextImager: a distributed UIMA-based system for NLP. In: Proceedings of the COLING 2016 system demonstrations
17. Lance AR, Mitchell PM (1995) Text chunking using transformation-based learning. CoRR cmp-lg/9505040
18. Finkel JR, Grenager T, Manning C (2005) Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd annual meeting on association for computational linguistics. ACL '05. Association for computational linguistics, Stroudsburg, PA, USA, pp 363–370. https://doi.org/10.3115/1219840.1219885
19. Mueller T, Schmid H, Schütze H (2013) Efficient higher-order CRFs for morphological tagging. In: Proceedings of the 2013 conference on empirical methods in natural language processing. EMNLP 2013, pp 322–332. Association for Computational Linguistics, Seattle, Washington, USA
20. Kudo T (2005) CRF++: yet another CRF toolkit. Software available at https://taku910.github.io/crfpp/
21. Geyer K, Greenfield K, Mensch A, Simek O (2016) Named entity recognition in 140 characters or less. In: Microposts
22. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C (2016) Neural architectures for named entity recognition. Computing Research Repository
23. Marc C, Bart DM (2015) Hyperparameter search in machine learning. Computing research repository abs/1502.02127
24. Bergstra J, Bardenet R, Bengio Y, Kégl B (2011) Algorithms for hyper-parameter optimization. In: Proceedings of the 24th international conference on neural information processing systems. NIPS'11, pp 2546–2554. Curran Associates Inc., USA
25. Dietterich TG (2000) Ensemble methods in machine learning. In: Proceedings of the first international workshop on multiple classifier systems. MCS '00, pp 1–15. Springer, London
26. Jeffrey LE (1990) Finding structure in time. Cognit Sci 14(2):179–211
27. Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. Int J Uncertain Fuzziness Knowl-Based Syst 6(2):107–116. https://doi.org/10.1142/S0218488598000094
28. Pascanu R, Mikolov T, Bengio Y (2012) Understanding the exploding gradient problem. CoRR abs/1211.5063
29. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
30. Hammerton J (2003) Named entity recognition with long short-term memory. In: Proceedings of the seventh conference on natural language learning at HLT-NAACL 2003—vol 4. CONLL '03, pp 172–175. Association for Computational Linguistics, Stroudsburg. https://doi.org/10.3115/1119176.1119202
31. Weisstein EW (2002) Sigmoid function
32. Graves A, Mohamed AR, Hinton GE (2013) Speech recognition with deep recurrent neural networks. CoRR abs/1303.5778
33. Xu K, Ba J, Kiros R, Cho K, Courville AC, Salakhutdinov R, Zemel RS, Bengio Y (2015) Show, attend and tell: neural image caption generation with visual attention. CoRR abs/1502.03044
34. Yang Z, He X, Gao J, Deng L, Smola AJ (2015) Stacked attention networks for image question answering. CoRR abs/1511.02274
35. Golub D, He X (2016) Character-level question answering with attention. CoRR abs/1604.00727
36. Rei M, Crichton GKO, Pyysalo S (2016) Attending to characters in neural sequence labeling models. CoRR abs/1611.04361
37. Luong MT, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. CoRR abs/1508.04025
38. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. CoRR abs/1409.0473
39. Lu Y, Ji D, Yao X, Wei X, Liang X (2015) Chemdner system with mixed conditional random fields and multi-scale word clustering. J Cheminform 7(1):4. https://doi.org/10.1186/1758-2946-7-S1-S4
40. Khabsa M, Giles CL (2015) Chemical entity extraction using CRF and an ensemble of extractors. J Cheminform 7(1):12. https://doi.org/10.1186/1758-2946-7-S1-S12
41. Xu S, An X, Zhu L, Zhang Y, Zhang H (2015) A CRF-based system for recognizing chemical entity mentions (CEMs) in biomedical literature. J Cheminform 7(1):11. https://doi.org/10.1186/1758-2946-7-S1-S11
42. Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: EMNLP
43. Levy O, Goldberg, Y (2014) Dependency-based word embeddings. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (vol 2: short papers), vol 2, pp 302–308
44. Komninos A, Manandhar S (2016) Dependency based embeddings for sentence classification tasks. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp 1490–1500
45. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. CoRR abs/1310.4546