Deep Learning in Biomedical Text Mining: Contributions and Challenges

Tanvir Alam and Sebastian Schmeier

Abstract A large number of biomedical texts are published every day in scientific literature. Finding the relevant and useful information from the massive collection of scientific literature is a challenging task that can be compared to finding needles in the haystack. Biomedical text mining is one of the sophisticated methodologies that leverage the extraction of knowledge from existing biomedical texts automatically. Deep learning (DL) based techniques have rejuvenated this field with huge prospects. In this chapter, we highlighted the contribution of DL based techniques in three specific tasks in the field of biomedical text mining: named-entity recognition, relationship extraction, and question answering. We also discussed the DL based models that are proven to be successful in multiple natural language processing tasks and the related challenges we face using such DL based techniques. We believe DL based methods will play a significant role in the coming years for biomedical text mining.

Keywords Deep learning · Natural Language Processing · Named-entity recognition · Relationship extraction · Question answering

1 Introduction

Biomedical texts and literature are the key knowledge distribution channels for novel scientific findings. More than 3000 new articles are being published every day (Lee et al. 2019) leading to an overwhelming amount of new information for researchers in the biomedical domain (Giorgi and Bader 2018). Extracting relevant scientific information and discovering connections among biomedical entities is a daunting manual task (Jensen et al. 2006). Consequently, automated literature mining, including natural language processing (NLP), has become an integral part

T. Alam (✉)
College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
e-mail: talam@hbku.edu.qa

S. Schmeier
School of Natural and Computational Sciences, Massey University, Auckland, New Zealand

M. Househ et al. (eds.), *Multiple Perspectives on Artificial Intelligence in Healthcare*,
Lecture Notes in Bioengineering, https://doi.org/10.1007/978-3-030-67303-1_14

of biomedical discovery that aids in rapidly accessing novel knowledge contained in large volumes of scientific literature. There are many different tasks related to biomedical text mining, but the most fundamental and useful tasks are named-entity recognition (NER), relation extraction (RE) and question answering (QA) (Lee et al. 2019). Historically, different rule-based (Ananiadou 1994; Dagan and Church 1994), dictionary-based (Salhi et al. 2017) and traditional machine-learning based methods have been used for providing solutions for these tasks. But such methods are heavily dependent on hand-curated features, which are often incomplete and very time-consuming to collect.

Recently, deep learning (DL), a branch of machine learning, has rejuvenated the field of biomedical text mining, including biomedical NLP (BioNLP). The major advantage of DL-based methods over existing methods is that DL-based methods require only a minimal level of hand-curated feature engineering and usually provide much better results, compared to traditional methods. Thus DL, a bio inspired neural network, which deploys multiple layers of artificial neurons to learn hierarchical representation of the data (Chen et al. 2018), is now considered the best paradigm for many different recognition tasks in many scientific domains (Bengio et al. 2013), including BioNLP. More recently, a variety of DL based methods and network architectures have been employed in the context of NLP (Young et al. 2018).

In one of the earliest landmark studies, Collobert et al. showed that DL-based methods can outperform traditional methods in most of the NLP related tasks (Collobert et al. 2011). Since then, DL in NLP has developed a strong following and, additionally, due to the emergence of the concept of word embedding (Mikolov et al. 2013a, b) and advancement of different DL methods (Devlin et al. 2018), it is now being used for all major tasks in NLP and biomedical text processing. In this chapter, we will focus on recent advancements in DL-based methods for biomedical text processing. The structure of this chapter is as follows: Sect. 2 lists DL-based techniques that have commonly been used in biomedical text mining. Sections 3, 4 and 5 discuss the contributions of DL in three key areas of biomedical text processing, namely NER, RE and QA systems. In Sect. 6, we highlight some challenges researchers may face when applying DL based techniques in NLP. Finally, we summarized and concluded the chapter in Sect. 7.

2 Deep Learning Architectures and Techniques that Have Been Proven Successful in NLP

In this section, we will first discuss embedding techniques, which are considered the first step in DL-based NLP. Afterwards, we will briefly describe some classical models that have been used for DL-based NLP. Finally, we will briefly describe some state-of-the-art DL techniques that have been published recently and achieved groundbreaking results in NLP.

2.1 Embeddings

Embedding is a set of feature engineering and language modeling techniques for NLP where each unit (e.g. word, sentence etc.) of the language are mapped to a vector of numbers. For any language modelling task, it is essential to learn the joint probability distribution of such units from input text (Young et al. 2018). However, such learning suffers from the curse of dimensionality as the data size is huge. As an alternative, distributed representations of input texts have been proposed in low dimensional space (Bengio et al. 2003). Learning the character-, word- or sentence-representations is a crucial step in biomedical text processing. Previous studies focused on learning word representations in a context independent manner. However, recent studies have focused on context-dependent representation learning (e.g. ELMo (Peters et al. 2018), CoVe (McCann et al. 2017)).

Distributional representation of words (word embedding) is often considered the first step in DL-based text processing. Word embedding captures the similarity between words based on the hypothesis that words with a similar meaning tend to appear together in similar context. In DL-based models, words, phrases and sentences are usually represented by embedding. The most successful and popular word embedding, Word2vec, was proposed by Mikolov et al. (2013a, b). The authors proposed a continuous bag of words (CBOW) and skip-gram model to build the distributed representation model. GloVe, proposed by Pennington et al., is another example of word embedding (Pennington et al. 2014). GloVe is essentially a count based model which considers a word co-occurrence matrix as input and this matrix is then factorized to generate a low dimensional representation of words.

Word embedding is a very useful tool to extract syntactic and semantic information from text, but intra-word morphological information might be useful for some specific tasks like NER and parts of speech (POS) tagging (Young et al. 2018). Moreover, in some languages (e.g. Chinese), sentences are not composed of multiple words but individual characters. For such languages character level embedding is a better approach to avoid word segmentation (Chen et al. 2015). For example, Peng et al. have used character-level embedding for sentiment classification (Peng et al. 2017). Additionally out-of-the-vocabulary words can not account for relevant tokens and misspellings (Giorgi and Bader 2018) and character-based embedding is a viable option to tackle such challenges (Ling et al. 2015).

2.2 Classical DL Based Techniques: CNN, RNN, LSTM, Attention Mechanism

Convolutional neural networks (CNN) belongs to a class of deep neural networks, which is the most commonly applied technique in DL, owing to its outstanding capacity of capturing spatial information from input data. The basic structure of a CNN consists of convolution layers, non-linear (activation) layers and pooling layers

(Fig. 1) (Lawrence et al. 1997). A convolution layer captures the local connectivity from different parts of the input data by using the same weight vector (weight-sharing policy). Based on this weight-sharing policy and local connectivity, a convolution layer captures intrinsic patterns from the data. The non-linear layer adds non-linear properties from the feature maps generated by the convolution layer. A pooling layer takes the average or maximum value form the non-overlapping region of the feature map.

In addition to spatial dependency in the data, the network also needs to capture temporal and order dependencies from text. Recurrent neural networks (RNN) are designed to exploit temporal relationships form input data. The basic structure of an RNN is shown in Fig. 2.

Though RNNs are designed to capture dependencies from input sequence data, it is generally not a good choice for capturing long range dependencies, as it tends to be biased towards the most recent input from the previous time step (Bengio et al.

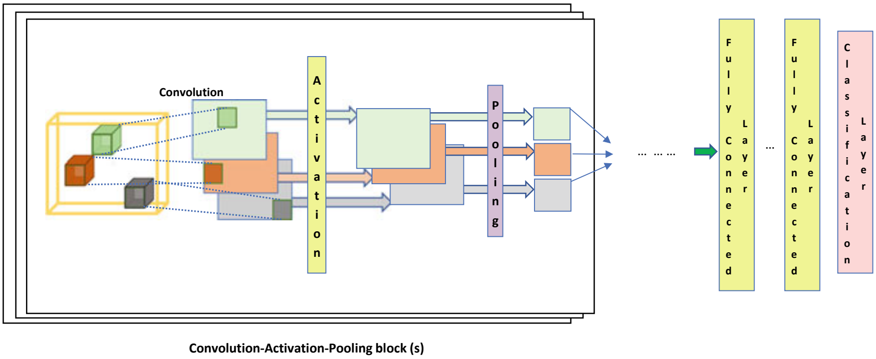

Fig. 1 A simple convolutional neural network (CNN). The major components of a CNN are: convolutional layers, activation (sigmoid/ReLU) layers, pooling (max/min/average) layers. The surrounding black box around these three layers represents the common order that might be used multiple times to increase the depth of the network. Recent CNNs have more computational layers such as Batch Normalization, Dropout, etc.

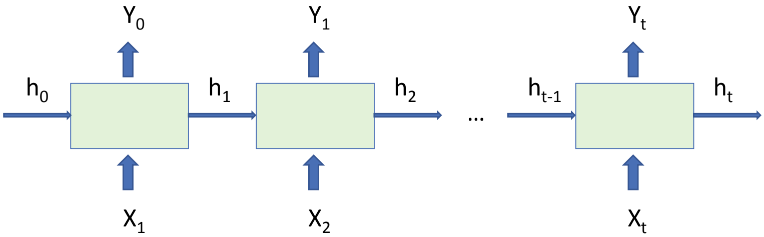

Fig. 2 A high-level diagram of a recurrent neural network (RNN). Computation at each time step t uses the input X_t and the previous time step's hidden-layer vector h_{t-1} to produce an output Y_t for the current time step and a hidden-layer vector h_t for the next time step

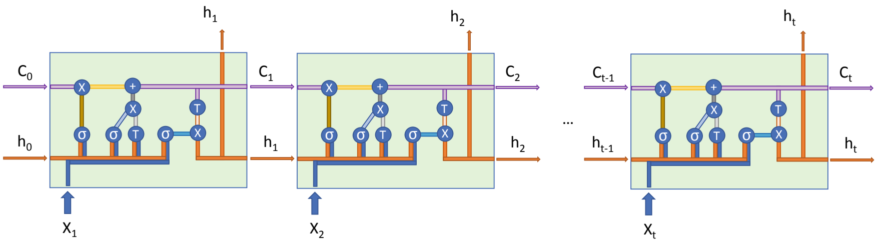

Fig. 3 An improvement to the vanilla RNN—long short-term memory network. It uses a dedicated memory vector, C_t at each time step to remember certain properties of the input (X_t) useful for the task at hand. A combination of the input from the current time-step, the hidden-vector (h_t), and the memory from the previous time-step (C_{t-1}) are used to compute 'gates' that are used in conjunction with these to produce the hidden-vector and memory values for the next time step

1994). A long short-term memory (LSTM) is a specific RNN, which tries to avoid the pitfalls of an RNN by having memory cells, which store summary information from all the preceding elements of input (Hochreiter and Schmidhuber 1997). A LSTM has a "forget" gate over RNN and this gate allows LSTM to back-propagate the error for unlimited time steps. The basic structure of a LSTM is shown in Fig. 3.

In a traditional sequence-to-sequence model for language translation task, the entire input sentence is encoded into a single vector, which is then used by the decoder to produce the output sentence. This model is not accurate in translating long sentences, since long-term dependencies are difficult to be decoded from a single vector representation of the entire input sentence. To alleviate such problems, attention mechanism has been introduced (Vaswani et al. 2017), where each word in the output sentence depends on a locally weighted combination of the words from the input.

2.3 Transfer Learning and Recent DL-Based Architectures that Rejuvenate the NLP Domain

Transfer learning (TL) is the concept to utilize already trained models to perform a similar task on a target dataset (Pan and Yang 2009; Weiss et al. 2016; Day and Khoshgoftaar 2017). TL has been successfully used in many different domains like computer vision (Yosinski et al. 2014; Oquab et al. 2014), speech recognition (Wang and Zheng 2015), etc. Recently, Mou et al. proposed a TL-based method to classify sentences using CNN (Mou et al. 2016). It has been a growing trend in the scientific community of NLP to use embedding with TL (Lee et al. 2019). However for biomedical text mining, the available embedding (e.g. based on wikipedia) needs to be modified to integrate the biomedical vocabulary as there is a huge difference between general corpus text of general corpus (e.g. wikipedia) and a biomedical text corpus (e.g. PubMed, PMC) (Lee et al. 2019).

Generative pre-training (GPT) is a recent model, developed by OpenAI, which achieved state-of-the-art results for many NLP tasks in 2018 (Radford 2018). Instead of using word embedding or character embedding, Radford et al. opted for a subword representation generated by a byte pair encoding (BPE) algorithm (Sennrich et al. 2016). They adopted a semi-supervised approach for language understanding tasks using an unsupervised pre-training approach followed by a supervised fine-tuning approach. In the first stage of model training, a transformer (Vaswani et al. 2017) mechanism was used to learn a universal representation of texts from huge amounts of unlabeled data from a diverse corpus with long stretches of contiguous text. In the second stage of model training, the model was fine tuned using a small amount of labelled data.

Bidirectional encoder representation from transformer (BERT) (Devlin et al. 2018), developed by Google, is a state-of-the-art DL-based word representation model that contextualizes words using bidirectional transfer. BERT proposed that bidirectionally (left-to-right and right-to-left) trained models can have a deeper understanding of the context than single direction language models. BERT uses a masked language model that can predict randomly picked words in a sentence and it showed that the pre-trained representation can reduce the need for a task-specific heavily-engineered DL architecture.

3 Deep Learning for Named-Entity Recognition in Biomedical Texts

Named-entity recognition (NER) is the process to recognize and label entities from a given text. NER is one of the most fundamental tasks in biomedical text mining. In the biomedical domain, the most common entity types are genes, proteins, chemicals and diseases (Yoon et al. 2019). NER methods can be broadly categorized into three groups: rule-based, dictionary-based and machine-learning based approaches. Rule based methods are scalable but specific to a particular task and it requires hand curated features and rules to fit into the model (Fukuda et al. 1998; Proux et al. 1998). In the dictionary-based approach, the entity mentioned in the text is checked against a dictionary of words of interest (Salhi et al. 2017; Hettne et al. 2009; Song et al. 2015a). The main drawback of dictionary-based NER is that these methods can not detect out-of-vocabulary words and it is tedious to build an up-to-date dictionary (Yoon et al. 2019). Until recently, NER tools for the biomedical domain were heavily relying on hand curated domain-specific features (Giorgi and Bader 2018). Conditional random fields (CRF) (Lafferty et al. 2001) are considered as the de-facto method for feature-based NER tasks. The process of feature engineering and dictionary creation is time consuming and depends on expert opinions (Leser and Hakenberg 2005) which leads to a domain-specific NER tool which, ultimately, is not generalizable for usage in other domains.

DL-based NER tasks are gaining popularity nowadays due to the advancements of new DL-based architectures that outperform existing rule-based and dictionary-based methods (Crichton et al. 2017; Wang et al. 2019). Recently Habibi et al. proposed a new DL-based long short-term memory network-conditional random field (LSTM-CRF) model which outperformed the state-of-the-art entity specific NER methods (Habibi et al. 2017). Their method combines word embedding, LSTM and CRF into a model for biomedical NER. TL based methods have achieved great attention in the scientific community as they showed significant improvements in NER performance. Lee et al. focused on TL using a CNN for NER (Lee et al. 2017). However, this was not meant for biomedical texts. To the best of our knowledge the first TL-based approach that was applied to biomedical NER was proposed by Giorgi and Bader (2018). Recently, Weber et al. developed HUNER (2019) which is a TL-based method for NER in the biomedical domain. HUNER extended the model proposed by Giorgi and Bader and outperformed the state-of-the-art tools tmChem (Leaman et al. 2015) and GNormPlus (Wei et al. 2015) in recognizing genes, species and chemical entities.

BioBERT, in a recent study gained a lot of attention from the scientific community for NER recognition in biomedical texts (Lee et al. 2019). In BioBERT, Lee et al. considered BERT (Devlin et al. 2018) as the backbone architecture and integrated biomedical articles from Pubmed and PMC with minimal domain-specific architecture modifications to outperform BERT in recognizing four different biomedical entities, genes, drugs, diseases and species.

4 Deep Learning for Relationship Extraction from Biomedical Texts

After biomedical entities have been identified in the literature, it is essential to discover underlying relationships among different entities (Rebholz-Schuhmann et al. 2012). Relationship extraction (RE) is meant to determine whether there is an association between entities. This task is more challenging than NER as the RE algorithms need to understand the meaning of a sentence (sentence-level RE) or the meaning within the whole document (document-level RE). RE at document-level is more difficult than sentence-level RE and most of the tools consider sentence-level RE without considering the context from the whole document (Wu et al. 2019).

One of the earliest examples for RE was Diseasome (Goh et al. 2007), where the authors provided the association information regarding 22 categories of human disorders and genes. There are many types of biomedical entities and different solutions have been tailored for identifying association among entities (Rebholz-Schuhmann et al. 2012). The most common type of associations, that is of primary interest for biomedical researchers, are gene-disease associations, protein–protein interactions, drug-drug interactions and gene-variants associations. For such RE tasks, different types of computational methods have been proposed: co-occurrence-based methods

(Hakenberg et al. 2012), pattern-/rule-based methods (Song et al. 2015b), as well as machine-learning based methods (Chun et al. 2006).

The simplest approach to identify a relationship between entities is entity co-occurrence (Stapley and Benoit 2000; Jenssen et al. 2001). A relationship can be inferred if two entities are co-occurring within the same sentence, paragraph, section or a document. Based on co-occurrence, Hakenberg et al. proposed an automated method to create a repository, SNPshot, that highlights genetic variants and their associations to different drugs and diseases (Hakenberg et al. 2012). Salhi et al. developed a knowledgebase, DES-ncRNA, based on 19 topic-specific dictionaries, to find associations between non-coding RNAs (micro-RNAs and long non-coding RNAs) and other entities, including diseases, mutations etc. (Salhi et al. 2017). Rule-based methods have been investigated for a long time for RE tasks from biomedical texts. Xie et al. developed miRCancer, based on 75 rules, to identify miRNAs that are involved in cancer based on text mining from biomedical literature (Xie et al. 2013). They built their own dictionary, regular expressions and rules to capture miRNA expressions and find their association to cancer. Song et al. developed a public knowledge discovery tool, called PKDE4J, to identify entities and extract relationships between entities (Song et al. 2015b). PKDE4J extends the Stanford CoreNLP (Manning et al. 2014) for dictionary-based NER and rule-based RE. Interested readers may refer to the publication (Song et al. 2015b) to understand more details about rule-based RE. Traditional machine-learning-based methods provided many sophisticated solutions for different RE tasks (Leach et al. 2009). Examples of such RE tasks include, but are not limited to, protein–protein interactions (Bui et al. 2011), protein subcellular localization prediction (Brady and Shatkay 2008), gene-disease associations (Chun et al. 2006), drug-drug interactions (Bui et al. 2014), etc.

Recently DL-based methods have gained a lot of attention in RE from biomedical texts. For extracting gene-disease associations from biomedical texts, Wu et al. developed RENET (Wu et al. 2019), a DL-based framework that not only captures sentence-level associations between genes and diseases but also models gene-disease associations across sentences in a document. In RENET, sentence-level representations were computed based on Word2Vec embedding (Mikolov et al. 2013a) and then passed through a CNN. Afterwards the sentence-level representations are transformed into document-level representations through an RNN. Finally, the document-based representation is used for gene-disease association prediction. BioBERT, mentioned above, uses a pre-trained model based on BERT to recognize gene-disease association in biomedical literature (Lee et al. 2019). BioBERT outperformed the state-of-the-art model result for GAD (Bravo et al. 2015) and EU-ADR (Mulligen et al. 2012) datasets in multiple evaluation metrics.

Protein–protein interaction (PPI) extraction from text is a challenging task, where DL-based methods have been used extensively. The majority of the DL-based PPI extraction tasks are performed by either a CNN (Quan et al. 2016; Peng and Lu 2017; Choi 2018) or RNN (Hsieh et al. 2017; Ahmed et al. 2019). Hua and Quan used the shortest dependency path (SDP) and a CNN to extract PPI from biomedical texts (Hua and Quan 2016). Recently, Zhang et al. proposed a residual CNN network for the PPI extraction task and their method achieved the best result in five benchmark data set (HPRD50, LLL, IEPA, BioInfer, AIMed) for PPI extraction corpora (Zhang et al. 2019).

DL-based methods have made major contributions in extracting drug-drug interaction (DDI) extraction from literature. Sahu et al. and Huang et al. developed a two-stage LSTM-based model to extract interaction between drugs from literature (Sahu and Anand 2018; Huang et al. 2017). Once a DDI is discovered, the authors categorized their interaction into one of four different groups: advice, effect, mechanism and interaction. Lim used a LSTM based model to extract DDI and their model outperformed other models on DDI Extraction Challenge'13 test data (Lim et al. 2018). Zhao et al. proposed a CNN based model to extract DDI (Zhao et al. 2016). They used a novel syntax embedding approach along with position specific features and POS features to categorize the DDI into five different categories: advice, effect, mechanism, interaction and negative. Liu et al. developed a multilayer bidirectional LSTM with transfer weight matrix (TWM) and a memory network to classify DDI into multiple types (Liu et al. 2019) and their model outperformed the other methods in DDI Extraction 2013 Task (Segura-Bedmar et al. 2014).

5 Deep Learning for Question Answering from Biomedical Texts

Question answering (QA) is the process of extracting answers to a specific question given one or multiple contexts (Wiese et al. 2017). The task of QA has been addressed in both, an open domain setup (Voorhees 2001) or domain-specific setup (Tsatsaronis et al. 2015). Based on the experimental setup different datasets have been proposed for the QA task. Stanford Question Answering Dataset (SQuAD) is the largest collection of QA dataset based on Wikipedia articles. SQuAD v1.0 dataset contains ~108 thousand QA pairs (Rajpurkar et al. 2016). However SQuAD is a generic dataset for QA and not specific to the biomedical domain. BioASQ is the most matured QA dataset in the biomedical domain, which comprises ~900 single answers (factoid) or multiple answers (list) question answering (QA) instances (Tsatsaronis et al. 2015).

Traditional QA systems can be divided into multiple modules: NER, question classification, and correct answer processing (Jurafsky and Martin 2009). Such systems have been applied to biomedical QA with limited success. For example Zi et al. developed the OAQA system, which integrates domain-specific information (Yang et al. 2016). Recently, due to the advancement of neural network-based DL techniques, scientific communities are developing end-to-end QA systems, rather than the traditional approach of subdividing the QA system into multiple discrete steps (Wiese et al. 2017). This end-to-end neural QA system usually starts with an embedding layer. Afterwards, an encoding layer is used to process the token vectors, usually by an RNN. The third layer is usually the interaction layer, which captures interactions between questions and contexts. Finally, an answering layer assigns scores for all the context tokens. A list of such neural QA systems is proposed in Wiese et al. (2017), Xiong et al. (2016), Seo et al. (2016), Weissenborn et al. (2017), Wang and Jiang (2016).

Recently, Du et al. proposed a hierarchical attention-based transfer learning model to build a QA system for the biomedical domain (Du et al. 2019). Authors adopted BERT to enrich the semantic representation and a dot-product based attention mechanism to capture the question interaction clues for passage encoding. Their system achieved state-of-the-art performance and outperformed existing solutions for factoid questions (in 2016) and BioASQ-Task B (in 2017). Weissenborn et al. developed FastQA, an RNN-based neural QA system for extractive QA (Weissenborn et al. 2017). In FastQA, authors proposed that to build a high performance QA system, context/type matching heuristics should be considered, as well as more complex composition functions, instead of simple bag of words models. Wiese et al. employed several transfer learning techniques to develop a neural QA system, which achieved state-of-the-art results on factoid QA and good results on a list questions (Wiese et al. 2017). Recently, Lee et al. developed a QA system, which is a part of BioBERT (Lee et al. 2019), by fine tuning the BERT system. For biomedical QA systems, Lee et al. used BioASQ to adopt the same structure of BERT. On all the BioASQ datasets (4b, 5b, 6b), BioBERT outperformed the existing models, considering the mean reciprocal rank (MRR) evaluation metric. Table 1 summarizes DL-based techniques that have been used for NER, RE and QA tasks in biomedical texts and literature.

6 Challenges and Future Perspectives

No single method is universally applicable in all NLP domains and the choice of how to use DL techniques is still problem-specific and challenging. Traditional approaches for biomedical text processing will definitely remain valid because of their advantage to succeed even with small amounts of data. Also, to assess the statistics of any finding is still difficult in DL-based techniques (Angermueller et al. 2016). Additionally, the training complexity (e.g. hyperparameter tuning, avoiding overfitting etc.) for DL-based models are much higher compared to traditional machine learning based approaches, which is a common pitfall for all DL techniques. QA tasks from biomedical text are still far away from maturity and likely still a long way off before a mature system emerges. In the last few years, we have observed outstanding conversational agents appearing on the market (e.g., Microsoft Cortana, Apple Siri). But these agents can perform a relatively simple task of answering factual questions (Dhingra et al. 2017). Lack of ability to learn from interactions with a user is the bottleneck in the QA task and reinforcement learning (RL) based techniques will play a big part in the improvement of existing QA systems (Dhingra et al. 2017). In the future, we will expect a lot of improvement and application of DL-based techniques in the QA tasks. Such an improved QA system will play a pivotal role in implementing highly accurate and useful chatbots in the healthcare sector as well. But such systems

Table 1 Brief list of recent publications and DL-based techniques that have been used in different BioNLP tasks

Tasks related to BioNLP	Deep learning based techniques	References
NER	Deep neural network	Yoon et al. (2019)
	CNN	Crichton et al. (2017)
	RNN/LSTM	Wang et al. (2019), Habibi et al. (2017), Weber et al. (2019)
	Transfer learning	Lee et al. (2019), Giorgi and Bader (2018)
RE	CNN	Gene-disease (Wu et al. 2019) PPI (Quan et al. 2016; Peng and Lu 2017; Choi 2018; Hua and Quan 2016; Zhang et al. 2019) DDI (Zhao et al. 2016)
	RNN/LSTM	PPI (Hsieh et al. 2017; Ahmed et al. 2019) DDI (Sahu and Anand 2018; Huang et al. 2017; Lim et al. 2018; Liu et al. 2019)
	Transfer learning	Lee et al. (2019)
QA	RNN/LSTM	Wiese et al. (2017), Xiong et al. (2016), Seo et al. (2016), Weissenborn et al. (2017), Wang and Jiang (2016)
	Transfer learning	Lee et al. (2019), Wiese et al. (2017), Du et al. (2019)

need to be significantly enhanced and tested rigorously before applying into real-life clinical setup.

7 Conclusions

Deep learning is a useful technique, which has facilitated manifold improvements in biomedical text processing. In this article, we have provided a brief summary of some of the DL-based techniques and their contributions in three key areas of biomedical text processing: NER, RE and QA. This article does not cover all aspects of DL (e.g. deep reinforcement learning) and all tasks related to NLP. However, we focused on the most relevant DL-based techniques that have been used in BioNLP, recently. We believe this article will aid the research community to have an overview of the contributions of DL in biomedical text processing.

Acknowledgements We would like to thank Mohammad Tariqul Islam and Faisal Ahmed for the discussion on multiple topics of NLP.

References

Ahmed M, Islam J, Samee MR, Mercer RE (2019) Identifying Protein-protein interaction using tree LSTM and structured attention. In: 2019 IEEE 13th international conference on semantic computing (ICSC). 2019. https://doi.org/10.1109/icosc.2019.8665584

Ananiadou S (1994) A methodology for automatic term recognition. In: Proceedings of the 15th conference on computational linguistics. https://doi.org/10.3115/991250.991317

Angermueller C, Pärnamaa T, Parts L, Stegle O (2016) Deep learning for computational biology. Mol Syst Biol 12:878

Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5:157–166

Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A Neural probabilistic language model. J Mach Learn Res 3:1137–1155

Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35:1798–1828

Brady S, Shatkay H (2008) EpiLoc: a (working) text-based system for predicting protein subcellular location. Pac Symp Biocomput 604–615

Bravo À, Piñero J, Queralt N, Rautschka M, Furlong LI (2015) Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. https://doi.org/10.1101/007443

Bui Q-C, Katrenko S, Sloot PMA (2011) A hybrid approach to extract protein-protein interactions. Bioinformatics 27:259–265

Bui Q-C, Sloot PMA, van Mulligen EM, Kors JA (2014) A novel feature-based approach to extract drug-drug interactions from biomedical text. Bioinformatics 30:3365–3371

Chen X, Xu L, Liu Z, Sun M, Luan H (2015) Joint learning of character and word embeddings. In: Twenty-fourth international joint conference on artificial intelligence. Available: https://www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/view/11000

Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 40:834–848

Choi S-P (2018) Extraction of protein–protein interactions (PPIs) from the literature by deep convolutional neural networks with various feature embeddings. J Inf Sci 60–73. https://doi.org/10.1177/0165551516673485

Chun H-W, Tsuruoka Y, Kim J-D, Shiba R, Nagata N, Hishiki T et al (2006) Extraction of gene-disease relations from Medline using domain dictionaries and machine learning. Pac Symp Biocomput 4–15

Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P (2011) Natural language processing (Almost) from scratch. J Mach Learn Res 12:2493–2537

Crichton G, Pyysalo S, Chiu B, Korhonen A (2017) A neural network multi-task learning approach to biomedical named entity recognition. BMC Bioinformatics 18:368

Dagan I, Church K (1994) Termight: Identifying and translating technical terminology. In: Proceedings of the fourth conference on applied natural language processing. https://doi.org/10.3115/974358.974367

Day O, Khoshgoftaar TM (2017) A survey on heterogeneous transfer learning. J Big Data. https://doi.org/10.1186/s40537-017-0089-0

Devlin J, Chang M-W, Lee K, Toutanova K (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. Available: http://arxiv.org/abs/1810.04805

Dhingra B, Li L, Li X, Gao J, Chen Y-N, Ahmed F et al (2017) Towards end-to-end reinforcement learning of dialogue agents for information access. In: Proceedings of the 55th annual meeting of the association for computational linguistics, vol 1. Long Papers. https://doi.org/10.18653/v1/p17-1045

Du Y, Pei B, Zhao X, Ji J (2019) Deep scaled dot-product attention based domain adaptation model for biomedical question answering. Methods. https://doi.org/10.1016/j.ymeth.2019.06.024

Fukuda K, Tamura A, Tsunoda T, Takagi T (1998) Toward information extraction: identifying protein names from biological papers. Pac Symp Biocomput 707–718

Giorgi JM, Bader GD (2018) Transfer learning for biomedical named entity recognition with neural networks. Bioinformatics 34:4087–4094

Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L (2007) The human disease network. Proc Natl Acad Sci U S A 104:8685–8690

Habibi M, Weber L, Neves M, Wiegandt DL, Leser U (2017) Deep learning with word embeddings improves biomedical named entity recognition. Bioinformatics 33:i37–i48

Hakenberg J, Voronov D, Nguyên VH, Liang S, Anwar S, Lumpkin B et al (2012) A SNPshot of PubMed to associate genetic variants with drugs, diseases, and adverse reactions. J Biomed Inform 45:842–850

Hettne KM, Stierum RH, Schuemie MJ, Hendriksen PJM, Schijvenaars BJA, van Mulligen EM et al (2009) A dictionary to identify small molecules and drugs in free text. Bioinformatics 25:2983–2991

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hsieh Y-L, Chang Y-C, Chang N-W, Hsu W-L (2017) Identifying protein-protein interactions in biomedical literature using recurrent neural networks with long short-term memory. In: Proceedings of the Eighth international joint conference on natural language processing, vol 2. Short Papers, 240–245

Hua L, Quan C (2016) A shortest dependency path based convolutional neural network for protein-protein relation extraction. Biomed Res Int 2016:8479587

Huang D, Jiang Z, Zou L, Li L (2017) Drug–drug interaction extraction from biomedical literature using support vector machine and long short term memory networks. Inf Sci 100–109. https://doi.org/10.1016/j.ins.2017.06.021

Jensen LJ, Saric J, Bork P (2006) Literature mining for the biologist: from information retrieval to biological discovery. Nat Rev Genet 7:119–129

Jenssen T-K, Lægreid A, Komorowski J, Hovig E (2001) A literature network of human genes for high-throughput analysis of gene expression. Nat Genet 21–28. https://doi.org/10.1038/ng0501-21

Jurafsky D, Martin JH (2009) Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall

Lafferty JD, McCallum A, Pereira FCN (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the eighteenth international conference on machine learning. Morgan Kaufmann Publishers Inc., pp 282–289

Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw 98–113. https://doi.org/10.1109/72.554195

Leach SM, Tipney H, Feng W, Baumgartner WA, Kasliwal P, Schuyler RP et al (2009) Biomedical discovery acceleration, with applications to craniofacial development. PLoS Comput Biol. 2009;5: e1000215

Leaman R, Wei C-H, Lu Z (2015) tmChem: a high performance approach for chemical named entity recognition and normalization. J Cheminform 7:S3

Lee JY, Dernoncourt F, Szolovits P (2017) Transfer learning for named-entity recognition with neural networks. Available: http://arxiv.org/abs/1705.06273

Lee J, Yoon W, Kim S, Kim D, Kim S, So CH et al (2019) BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz682

Leser U, Hakenberg J (2005) What makes a gene name? Named entity recognition in the biomedical literature. Brief Bioinform 6:357–369

Lim S, Lee K, Kang J (2018) Drug drug interaction extraction from the literature using a recursive neural network. PLoS ONE. e0190926. https://doi.org/10.1371/journal.pone.0190926

Ling W, Dyer C, Black AW, Trancoso I, Fermandez R, Amir S et al (2015) Finding function in form: compositional character models for open vocabulary word representation. In: Proceedings of the 2015 conference on empirical methods in natural language processing. https://doi.org/10.18653/v1/d15-1176

Liu J, Huang Z, Ren F, Hua L (2019) Drug-drug interaction extraction based on transfer weight matrix and memory network. IEEE Access 101260–101268. https://doi.org/10.1109/access.2019.2930641

Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D (2014) The stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations. https://doi.org/10.3115/v1/p14-5010

McCann B, Bradbury J, Xiong C, Socher R (2017) Learned in translation: contextualized word vectors. Adv Neural Inf Proc Syst 6294–6305

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013a) Distributed representations of words and phrases and their compositionality. Adv Neural Inform Proc Syst 3111–3119

Mikolov T, Chen K, Corrado G, Dean J (2013b) Efficient estimation of word representations in vector space. Available: http://arxiv.org/abs/1301.3781

Mou L, Meng Z, Yan R, Li G, Xu Y, Zhang L et al (2016) How transferable are neural networks in NLP Applications? In: Proceedings of the 2016 conference on empirical methods in natural language processing. https://doi.org/10.18653/v1/d16-1046

Oquab M, Bottou L, Laptev I, Sivic J (2014) Learning and transferring mid-level image representations using convolutional neural networks. In: 2014 IEEE conference on computer vision and pattern recognition. https://doi.org/10.1109/cvpr.2014.222

Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE J Mag. [cited 28 Sep 2019]. Available: https://ieeexplore.ieee.org/abstract/document/5288526

Peng Y, Lu Z (2017) Deep learning for extracting protein-protein interactions from biomedical literature. BioNLP 2017. https://doi.org/10.18653/v1/w17-2304

Peng H, Cambria E, Zou X (2017) Radical-based hierarchical embeddings for Chinese sentiment analysis at sentence level. In: The Thirtieth international flairs conference. Available: https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS17/paper/view/15460

Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). https://doi.org/10.3115/v1/d14-1162

Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K et al (2018) Deep contextualized word representations. In: Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: human language technologies, vol 1. (Long Papers). https://doi.org/10.18653/v1/n18-1202

Proux D, Rechenmann F, Julliard L, Pillet VV, Jacq B (1998) Detecting gene symbols and names in biological texts: a first step toward pertinent information extraction. Genome Inform Ser Workshop Genome Inform 9:72–80

Quan C, Hua L, Sun X, Bai W (2016) Multichannel convolutional neural network for biological relation extraction. Biomed Res Int 2016:1850404

Radford A (2018) Improving language understanding by generative pre-training. [cited 28 Sep 2019]. Available: https://pdfs.semanticscholar.org/cd18/800a0fe0b668a1cc19f2ec95b5003d0a5035.pdf

Rajpurkar P, Zhang J, Lopyrev K, Liang P (2016) SQuAD: 100,000 questions for machine comprehension of text. In: Proceedings of the 2016 conference on empirical methods in natural language processing. https://doi.org/10.18653/v1/d16-1264

Rebholz-Schuhmann D, Oellrich A, Hoehndorf R (2012) Text-mining solutions for biomedical research: enabling integrative biology. Nat Rev Genet 13:829–839

Sahu SK, Anand A (2018) Drug-drug interaction extraction from biomedical texts using long short-term memory network. J Biomed Inform 86:15–24

Salhi A, Essack M, Alam T, Bajic VP, Ma L, Radovanovic A et al (2017) DES-ncRNA: A knowledgebase for exploring information about human micro and long noncoding RNAs based on literature-mining. RNA Biol 14:963–971

Segura-Bedmar I, Martínez P, Herrero-Zazo M (2013) Lessons learnt from the DDIExtraction-2013 shared task. J Biomed Inform 152–164. https://doi.org/10.1016/j.jbi.2014.05.007

Sennrich R, Haddow B, Birch A (2016) Neural machine translation of rare words with subword units. In: Proceedings of the 54th annual meeting of the association for computational linguistics, vol 1. Long Papers. https://doi.org/10.18653/v1/p16-1162

Seo M, Kembhavi A, Farhadi A, Hajishirzi H (2016) Bidirectional attention flow for machine comprehension. Available: http://arxiv.org/abs/1611.01603

Song M, Yu H, Han W-S (2015a) Developing a hybrid dictionary-based bio-entity recognition technique. BMC Med Inform Decis Mak 15(Suppl 1):S9

Song M, Kim WC, Lee D, Heo GE, Kang KY (2015b) PKDE4J: entity and relation extraction for public knowledge discovery. J Biomed Inform 57:320–332

Stapley BJ, Benoit G (2000) Biobibliometrics: information retrieval and visualization from co-occurrences of gene names in Medline abstracts. Pac Symp Biocomput 529–540

Tsatsaronis G, Balikas G, Malakasiotis P, Partalas I, Zschunke M, Alvers MR et al (2015) An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition. BMC Bioinf 16:138

van Mulligen EM, Fourrier-Reglat A, Gurwitz D, Molokhia M, Nieto A, Trifiro G et al (2012) The EU-ADR corpus: Annotated drugs, diseases, targets, and their relationships. J Biomed Inf 879–884. https://doi.org/10.1016/j.jbi.2012.04.004

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al (2017) Attention is all you need. Adv Neural Inf Process Syst 5998–6008

Voorhees EM (2001) The TREC question answering track. Nat Lang Eng, 361–378. https://doi.org/10.1017/s1351324901002789

Wang S, Jiang J (2016) Machine comprehension using match-LSTM and answer pointer. Available: http://arxiv.org/abs/1608.07905

Wang D, Zheng TF (2015) Transfer learning for speech and language processing. In: 2015 Asia-Pacific signal and information processing association annual summit and conference (APSIPA). https://doi.org/10.1109/apsipa.2015.7415532

Wang X, Zhang Y, Ren X, Zhang Y, Zitnik M, Shang J et al (2019) Cross-type biomedical named entity recognition with deep multi-task learning. Bioinformatics 35:1745–1752

Weber L, Münchmeyer J, Rocktäschel T, Habibi M, Leser U (2019) HUNER: improving biomedical NER with pretraining. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz528

Wei C-H, Kao H-Y, Lu Z (2015) GNormPlus: an integrative approach for tagging genes, gene families, and protein domains. Biomed Res Int 2015:918710

Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J Big Data. https://doi.org/10.1186/s40537-016-0043-6

Weissenborn D, Wiese G, Seiffe L (2017) Making neural QA as simple as possible but not simpler. In: Proceedings of the 21st conference on computational natural language learning (CoNLL 2017). https://doi.org/10.18653/v1/k17-1028

Wiese G, Weissenborn D, Neves M (2017) Neural domain adaptation for biomedical question answering. In: Proceedings of the 21st conference on computational natural language learning (CoNLL 2017). https://doi.org/10.18653/v1/k17-1029

Wu Y, Luo R, Leung HCM, Ting H-F, Lam T-W (2019) RENET: a deep learning approach for extracting gene-disease associations from literature. Lect Notes Comput Sci 272–284. https://doi.org/10.1007/978-3-030-17083-7_17

Xie B, Ding Q, Han H, Wu D (2013) miRCancer: a microRNA-cancer association database constructed by text mining on literature. Bioinformatics 29:638–644

Xiong C, Zhong V, Socher R (2016) Dynamic coattention networks for question answering. Available: http://arxiv.org/abs/1611.01604

Yang Z, Zhou Y, Nyberg E (2016) Learning to answer biomedical questions: OAQA at BioASQ 4B. Proc Fourth BioASQ Workshop. https://doi.org/10.18653/v1/w16-3104

Yoon W, So CH, Lee J, Kang J (2019) CollaboNet: collaboration of deep neural networks for biomedical named entity recognition. BMC Bioinf 20:249

Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? Adv Neural Inf Process Syst 3320–3328

Young T, Hazarika D, Poria S, Cambria E (2018) Recent trends in deep learning based natural language processing [Review Article]. IEEE Comput Intell Mag 55–75. https://doi.org/10.1109/mci.2018.2840738

Zhang H, Guan R, Zhou F, Liang Y, Zhan Z-H, Huang L et al (2019) Deep residual convolutional neural network for protein-protein interaction extraction. IEEE Access. 89354–89365. https://doi.org/10.1109/access.2019.2927253

Zhao Z, Yang Z, Luo L, Lin H, Wang J (2016) Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. Bioinformatics. p. btw486. https://doi.org/10.1093/bioinformatics/btw486