TSINGHUA PBCSF 清华五道口 | IFR 金融科技研究院

研究报告

（2020 年 第 11 期 总第 64 期） 2020 年 4 月 30 日

PBCSF

Deep Learning for Sentiment Analysis

金融大数据研究中心、百融云创

【摘要】 过去十年随着深度学习的崛起，人工智能在自然语言处理、机器视觉等很多应用领域取得了突破性进展。情绪分析是自然语言处理领域的一个重要研究课题，近几年越来越多的研究工作在积极探索利用深度学习来更好地解决文本情绪分析的问题。两大方面因素促使该研究方向成为热点和重点：一方面网络的全面普及和社会化媒体（例如 Twitter、博客、维基、论坛、社交网络、内容社区等）的快速发展使人类社会积累了史无前例的大量的文本数据，这如同创造了一片全新的文本数据的“海

洋”；另一方面深度学习技术的突破为大规模自动化处理包括文本在内的非结构化数据提供了强大有效的工具，这也使得高效探索文本信息海洋并从中挖掘相关的知识成为可能。

除了传统的结构化数据，经济金融领域也积累了大量相关的文本信息，所以文本分析技术的发展无疑对经济金融学术研究和产业应用都有巨大潜在影响力。在前述两方面因素的共同驱动下，利用深度学习进行文本挖掘成为机器学习和金融科技领域共同的研究热点。

本文旨在汇总分析深度学习在文本分析中的应用和取得的成果。为此，我们首先探讨分析了深度学习架构近几年的最新突破和进展；在此基础上，本文汇总整理了不同深度学习架构在文本挖掘上的应用和成果；最后，我们指出了该领域当前仍面临的挑战和今后的一些研究方向。

（注：本篇报告主要基于外文参考文献展开研究，正文以英文形式呈现。）

PBCSF

目 录

PBCSF

Deep Learning for Sentiment Analysis

金融大数据研究中心、百融云创

1. Overview of Sentiment Analysis

Sentiment analysis has been an ongoing and important research topic in the natural language processing field. It aims to determine the judgement of a writer with respect to a certain topic based on a given text, typically a single sentence or an entire document (Glorot et al., 2011). There are a number of different names for this task, such as opinion extraction, opinion mining, sentiment mining, and subjectivity analysis.

As sentiment analysis is closely related to decision-making and predictions, it has found numerous applications in a wide span of domains, from buying products online from Amazon or Alibaba to making multi-million investment decisions in emerging technology, from predicting US presidential election to extracting investor views from stock reviews. From a high-level perspective, the capability to extract opinions from the flooding digital text enables sentiment analysis to shape the business landscape remarkably and to exert influence on the society profoundly.

In the computer science literature, sentiment analysis is typically framed as a classification or regression problem. Historically, machine learning researchers have adopted a variety of algorithms, both supervised and unsupervised, to tackle the sentiment analysis task, but only achieved limited success. For instance, earlier research

PBCSF

investigates multiple classic machine learning algorithms for sentiment classification, such as naive bayes, support vector machine, and maximum entropy. However, these early approaches are not competent enough to extract fine-grained sentiment from the given text.

Over the past decade, the deep learning evolution has brought many ground-breaking progresses in various subfields from computer vision to natural language processing. In line with this trend, the machine learning community has adopted a variety of deep learning techniques for sentiment analysis and therefore improved the performance on benchmark datasets substantially.

In addition to the methodological breakthroughs primarily driven by deep learning, the exponentially increasing online data across domains further stimulates the interest in sentiment mining. In particular, the rise of social media on the Web such as social networks, micro-blogs, ratings and reviews contributes to the accumulation of a huge opinionated digital data. Indeed, the society as a whole has produced an ocean of online text on an unprecedented scale in human history. Now, the challenge is how to efficiently and effectively explore this ocean of data. "We are drowning in information and starving for knowledge", as Rutherford Roger put it. In light of this, sentiment mining is part of the effort toward this goal of extracting knowledge from the vast text ocean.

In a nutshell, the advancements in machine learning techniques and the exploding online data together make sentiment analysis a hot and crucial topic in both academic

PBCSF

research and industrial practices. On the one hand, we see a surge of research in sentiment classification in the past decade. On the other, we also witness a rapidly growing applications of sentiment analysis methods by corporations, start-ups as well as government agencies. For instance, there are start-ups focusing on providing sentiment analysis of unstructured data. Moreover, global companies are establishing in-house apartments to build such capacity.

This study seeks to provide a survey of recent applications of deep learning approaches in sentiment analysis. In doing so, we begin by reviewing the relevant background including word embeddings and important deep learning architectures. Equipped with these, we then survey representative work on applying these deep learning techniques for sentiment classification. With a comprehensive picture in mind, we summarize certain remaining challenges in sentiment analysis and discuss future research directions.

PBCSF

2. Word Embedding and Deep Learning Architectures

2.1 Word Embedding

The arguably most important common denominator of NLP tasks is figuring out how to represent words as the input to any language models. Words are typically represented as vectors that can effectively capture the similarities, differences as well as relationships between them.

Among various ways to do word embedding, the simplest one to encode words into numeric vectors is one-hot vector, which represent every word as a V-dimensional vector with all 0s but one 1 at the index of the word of that language. V is the size of the vocabulary. While simple, the disadvantages of one-hot vector are obvious: the vector is very high-dimensional and sparse; moreover, modeling each word as an independent vector fails to directly capture any similarities and all semantic connections among words. These motivate researchers to investigate into alternative ways, which gives rise to different classes of word embedding methods, as we will see below.

SVD Based Word Embedding

For SVD-based word embedding, the first step is to loop over a large dataset to count the co-occurrence of words. Such counts then form a matrix. Two specific choices for this matrix are the word-document matrix and the window-based co-occurrence matrix. Secondly, we perform SVD on the matrix and get USV^T. Then the rows of U will be used as the word embeddings. In practice, we typically select a submatrix of U based on desired percentage variance, which gives a representation of all words in the corpus. Of course, this class of methods is not perfect and has its own problems. The matrix is very sparse and high-dimensional, which requires quadratic time to calculate, and changes every time we introduce new words into the corpus, just to name a few.

Iteration Based Word Embedding

Iteration-based methods address some of the problems mentioned above. The main idea is to use the classic backpropagating method. We build a model whose parameters are the word vectors, and design an appropriate objective to train the model. During the

PBCSF

training process, we update the parameters per iteration. In other words, we capture the word co-occurrence one at a time instead of capturing all of them at the same time like the SVD method. In this category of word embeddings, popular methods include word2vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014).

Sentiment-Oriented Word Embedding Through Neural Network

In addition, we can use neural network to learn sentiment-oriented word embedding. To capture both semantic and sentiment information, Maas et al. (2011) mix an unsupervised probabilistic model and a supervised sentiment component to learn word vectors with "semantic term-document information as well as rich sentiment content." Alternatively, Labutov and Lipson (2013) propose to build up on an existing word embedding and re-purpose it to improve sentiment classification performance. To do so, they take the word embedding and label data as input and learn a re-embedding vectors that achieves better improvement on sentiment analysis.

2.2 Attention Mechanism, Transformer and BERT for Sentiment

We refer the readers to a Deep Learning textbook for a review of the classic deep learning architectures including Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long-Short Term Memory (LSTM)[①]. Due to limited space, we choose to focus on the recent breakthroughs in deep learning architectures, which has far-reaching influence on sentiment analysis and the NLP field at large.

① LSTM is one kind of RNNs.

PBCSF

2.2.1 Attention Mechanism

The research on attention has a long history in cognitive neuroscience. Take human vision for example, we do not process all sensory input at any time; instead, attention allows us to prioritize and focus on a small fraction of the input that is most relevant. There are various types of attention in cognitive neuroscience, selective attention, covert attention, and spatial attention (Zhang et al, 2020). These findings inspire the deep learning community to introduce similar attention idea into the neural networks.

In NLP, attention is firstly introduced to tackle the information bottleneck problem of the Seq2Seq model (Bahdanau et al., 2015). Specifically, the attention mechanism provides direct connections between the decoder and encoder (instead of only the encoder output) and enables the decoder to focus on a particular part of the source sequence. While attention is initially used in the Seq2Seq model, it actually has very wide applications beyond this particular architecture. Indeed, the attention mechanism is a general deep learning technique.

PBCSF

We now focus on how to transform the attention into mathematical models. In a general setting, we have a query as input and a set of values in the memory, attention is used to compute a weighed sum of the values, which depends on the query (Figure 1). Let us begin with one attention layer and breakdown the attention calculation process into three steps. First, we calculate an attention score via a score function. Then, we take a softmax of the attention scores, which gives an attention distribution. Finally, we use the attention distribution to compute a weighted sum of the values, which is the attention output.

In the first step, it is worthy to note that there are different choices of the score function, which leading to different attention layers accordingly. A simple and popular one is dot-product attention. Alternative score functions include the multiplicative attention and the additive attention.

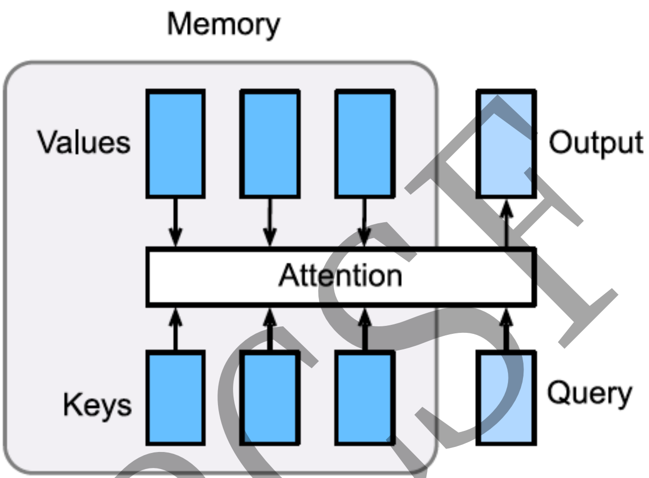


Figure 1: Illustration of attention mechanism (Zhang et al., 2020)

PBCSF

2.2.2 Transformer

A primal motivation for the Transformer is to tackle the fundamental constraint of sequential computation in existing recurrent models. In recent years, the attention mechanism has been incorporated into the conventional recurrent network to draw global dependencies of input and output. Such combined architectures have achieved state-of-the-art performance in multiple NLP tasks. While the attention mechanisms have become an integral part of both sequence modeling and transduction models, the sequential nature of the recurrent network still precludes the parallelization of the model training procedure. This constraint motivates the researchers to search for an architecture without recurrent network.

Transformer is the first network architecture that is solely based on attention mechanisms (Vaswani et al., 2017). It generalizes well to various NLP tasks and achieves superior results, while significantly reduces the training time compared to existing RNN-based architectures. In general, the Transformer has an encoder-decoder structure, as shown in Figure 2. In particular, it uses stacked self-attention and point-wise, fully connected layers for both the encoder and decoder.

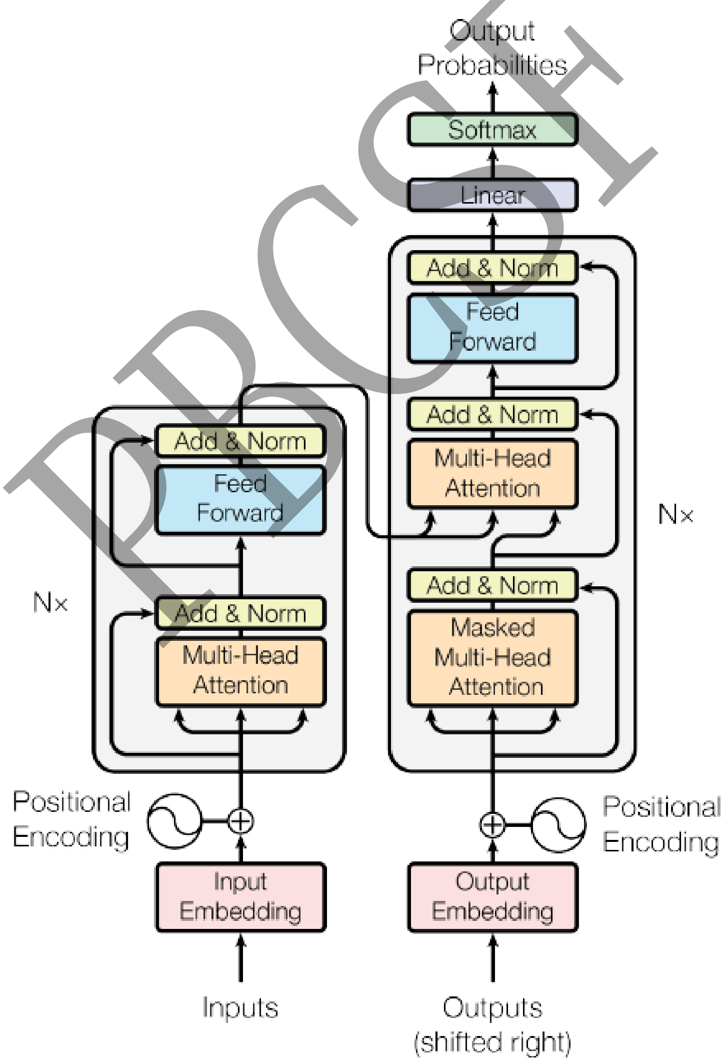


Figure 2: The model architecture of the Transformer (Vaswani et al., 2017)

We now dive into the self-attention block in the Transformer. The input consists of queries, keys, and values. The attention is calculated as a weighted sum of the values, while the weights is the dot product of queries and keys scaled by the square root of the dimension of the keys (Figure 3).

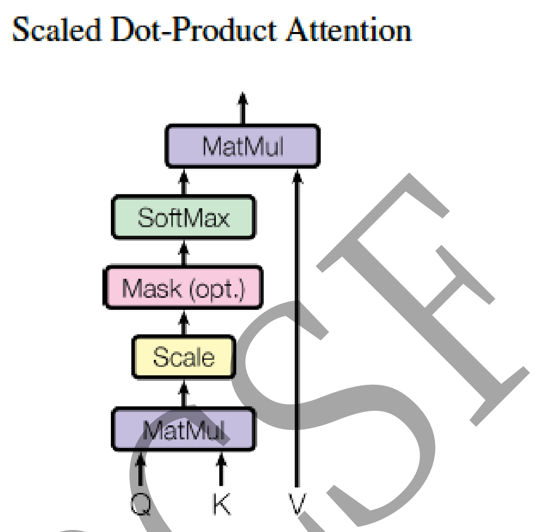


Figure 3: Scaled dot-product attention (Vaswani et al., 2017)

In practice, the queries, keys, and values are firstly projected into a lower dimension h times. Then we conduct the scaled dot-product attention on each of these projected versions and concatenate the results. In this way, we get the so-called multi-head attention. The underlying intuition is that the "multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions" (Vaswani et al., 2017).

PBCSF

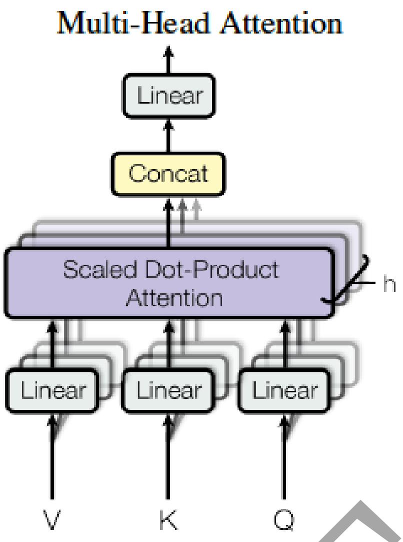


Figure 4: Multi-head attention (Vaswani et al., 2017)

2.2.3 BERT: Pretraining and Fine Tuning

Based on the Transformer, Devlin et al. (2018) introduce a new language representation model called Bidirectional Encoder Representations from Transformers (BERT), which is designed to "pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers". This pre-trained model can then be fine-tuned with one additional output layer to yield SOTA performance on downstream tasks including sentiment analysis.

The overall architecture of BERT is a multi-layer bidirectional Transformer encoder. This model architecture allows BERT to make crucial improvements over existing techniques. Recognizing that a key constraint of standard language models is their unidirectionality, BERT alleviates this limitation by using masked language models to enable pre-trained deep bidirectional representations. The pre-trained representations of BERT greatly reduce the need for heavily-engineered task-specific

PBCSF

architecture. In fact, a distinguished feature of BERT is that it uses a unified architecture across different tasks.

There are two steps in the BERT framework: pre-training and fine-tuning (Figure 5). For pre-training, the model is trained on unlabeled data over different pre-training tasks. For fine-tuning, BERT is firstly initialized by the pre-trained parameters and then all the parameters are fine-tuned using labeled data from the downstream tasks of interest.

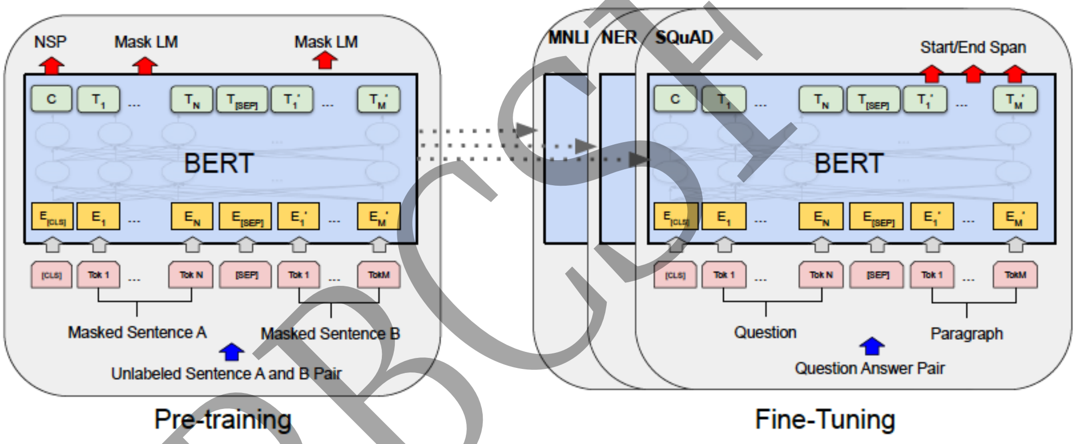


Figure 5: Overall pre-training and fine-tuning for BERT (Devlin et al., 2018)

PBCSF

3. Applications in Sentiment Analysis

In the past decade, the deep learning evolution has made many breakthroughs in the NLP field and achieved state-of-the-art performances in key NLP tasks. As an important and active research topic in the NLP field, we see a burgeoning research work that leverages deep learning methods to investigate sentiment classification. In this section, we provide a survey of deep learning applications in sentiment analysis.

Sentiment analysis has a number of subareas, each of which addresses a different level of research. In this section, we firstly categorize the literature into three groups by the granularity of sentiment analysis tasks; secondly, within each task, we further organize the literature based on the deep learning architecture used.

3.1 Classic Deep Learning Architectures for Sentiment Analysis

Sentiment analysis can be grouped into three categories based on the levels of granularity of sentiment: sentence-level, document-level, and aspect-level (Liu, 2015).

3.1.1 Sentence-level Sentiment Analysis

Existing literature typically frame the sentence-level sentiment as a three-way classification problem: positive, negative, and neutral. Below, we organize the sentence-level sentiment analysis according to the deep learning architecture adopted.

Autoencoder and RNN

Socher and coauthors conduct a series of research on sentiment analysis using autoencoders and various forms of the RNN architecture. Socher et al. (2011) introduce a recursive autoencoder model to predict the sentence-level sentiment label distribution (Figure 6). This model learns a semantic vector space representation of phrases, which can capture more complex linguistic phenomena compared to the bag-of-words method. In this model, they firstly map the word indices into a semantic vector space; then, use an autoencoder network to recursively merge the semantic vectors into a fixed length sentence representation; and finally use the vectors at every node as features in order to predict the distribution over sentiment labels.

PBCSF

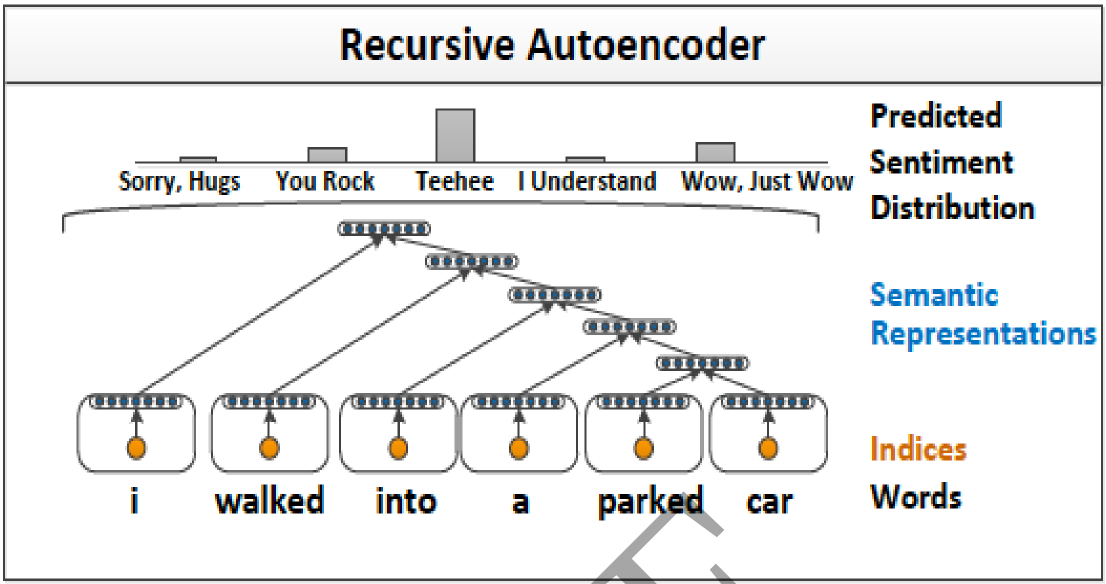


Figure 6: Recursive Autoencoder architecture for predicting sentence-level sentiment distribution (Socher et al., 2011)

To capture the compositional meaning of phrases, Socher et al. (2012) propose a matrix-vector RNN (MV-RNN) model to learn compositional vector representations for both phrases and sentences of arbitrary length. To do so, MV-RNN assigns a vector as well as a matrix to each node in a parse tree: the vector is used to capture the inherent meaning of the constituent, and the matrix aims to capture its impacts on neighboring words and phrases (Figure 7).

PBCSF

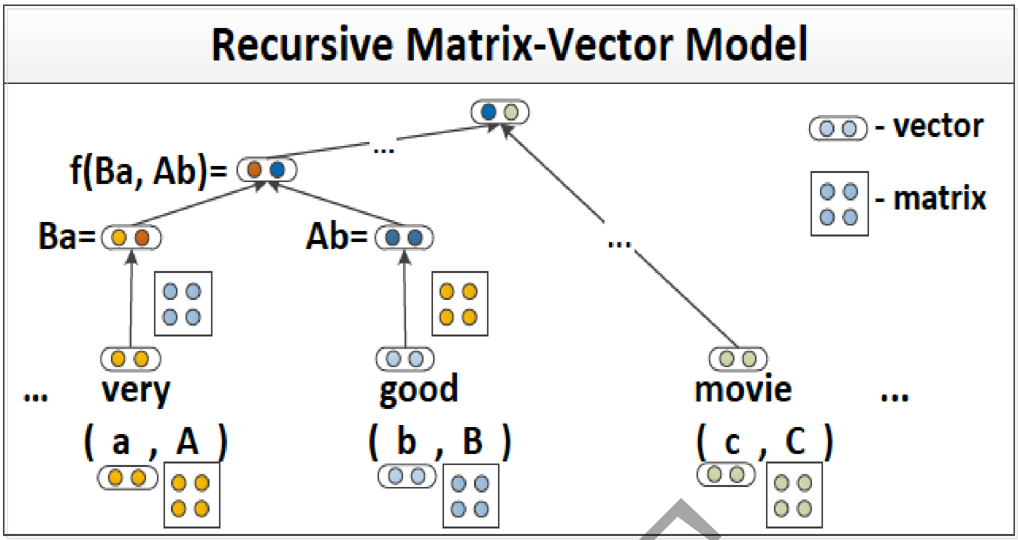


Figure 7: A recursive neural network that learns semantic vector representation of phrases (Socher et al., 2012)

Following this, Socher et al. (2013) introduce the Recursive Neural Tensor Network (RNTN) and achieve the state-of-the-art result in single sentence sentiment classification (negative/positive) with accuracy of 85.4%. Moreover, RNTN is able to capture the effects of negation and its scope for phrases. RNTN is based on the previous MV-RNN model, whose main idea is representing words and phrases in a parse tree as a combination of a vector and a matrix (Socher et al., 2012), and improves it by using the same and tensor-based composition function for all nodes.

PBCSF

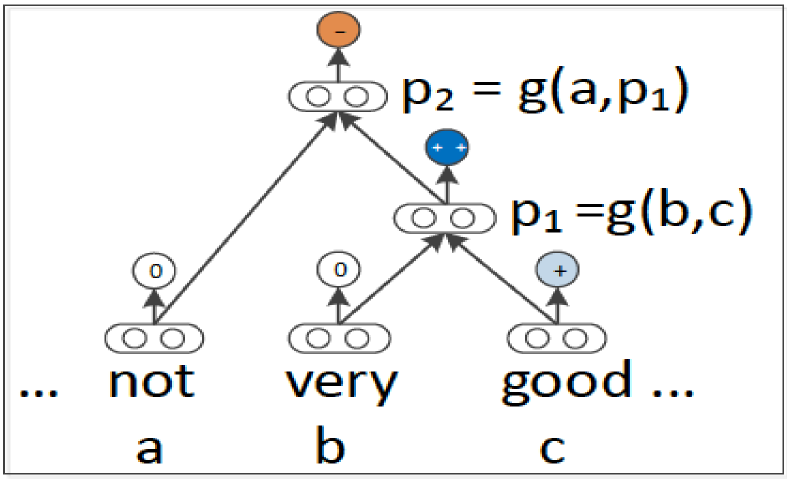


Figure 8: Approach of Recursive Neural Network models for sentiment (Socher et al., 2013)

LSTM

To better capture the interactions between words in the compositional process, Wang et al. (2015) use the Long Short-Term Memory (LSTM) to analyze the twitter sentiment. The intuition is that the structure of the LSTM unit has the potential to provide more flexibility to simulate the compositional results compared to the vanilla RNN.

A bidirectional LSTM (bi-LSTM) structure is applied to address both the semantic compositionality and word sense variations issues by Teng et al. (2016). As shown in Figure 9, the sentiment score of a sentence is calculated as a weighted sum of sentiment words scores and a sentence-level bias score: sentiment words scores are obtained from sentiment lexicons, and weights are context-sensitive. In this way, this structure has the potential to capture context-dependent semantic composition effects on sentence-level.

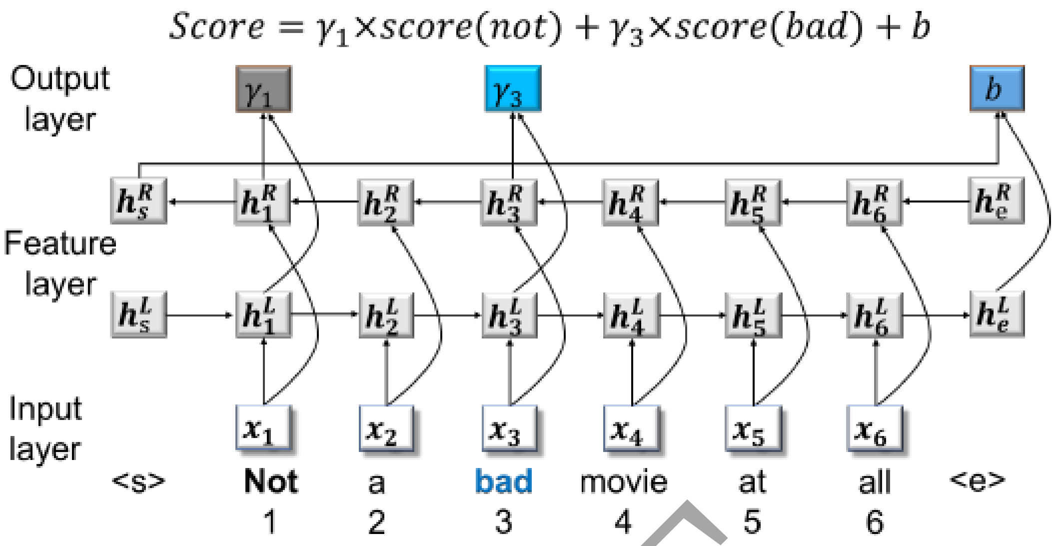


Figure 9: The bi-LSTM model structure for sentence-level sentiment score (Teng et al., 2016)

As the syntactic properties of natural language combines words to phrases, Tai et al. (2015) generalize LSTMs to tree-structured network topologies to improve the semantic representations of sentences. The so-called Tree-LSTM model outperformed other existing systems on sentiment classification then. Specifically, the Tree-LSTM extended the standard LSTM by using a tree topology to incorporate information from multiple child units (Figure 10).

PBCSF

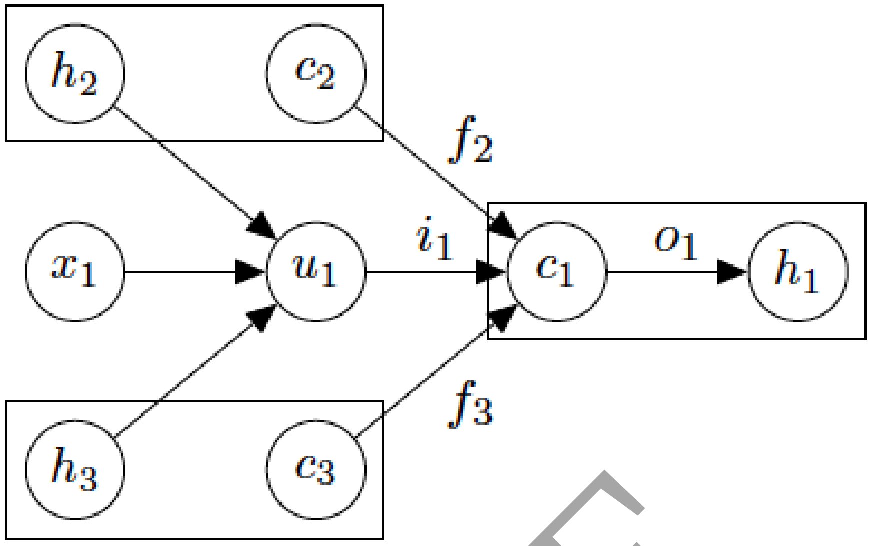


Figure 10: Composing the memory cell and hidden state of a Tree-LSTM unit with two children (Tai et al., 2015)

CNN

While CNN is originally invented for computer vision problems (LeCun et al., 1998), prior work shows that this architecture can also be effective for NLP tasks. Besides the rich literature using RNN structures for sentiment tasks described above, the CNN architecture has also been adopted to tackle the sentiment classification problem. Kim (2014) applies a simple CNN model on pre-trained word vectors to sentence-level sentiment extraction and achieved excellent results on several benchmarks (Figure 11). The word vectors are from Mikolov et al. (2013), which is trained by an unsupervised neural language model. The excellent results on sentence classification benchmarks indicates that the pre-trained word vectors can be viewed as 'universal' feature extrators that can be then used for multiple NLP tasks besides

PBCSF

sentiment analysis.

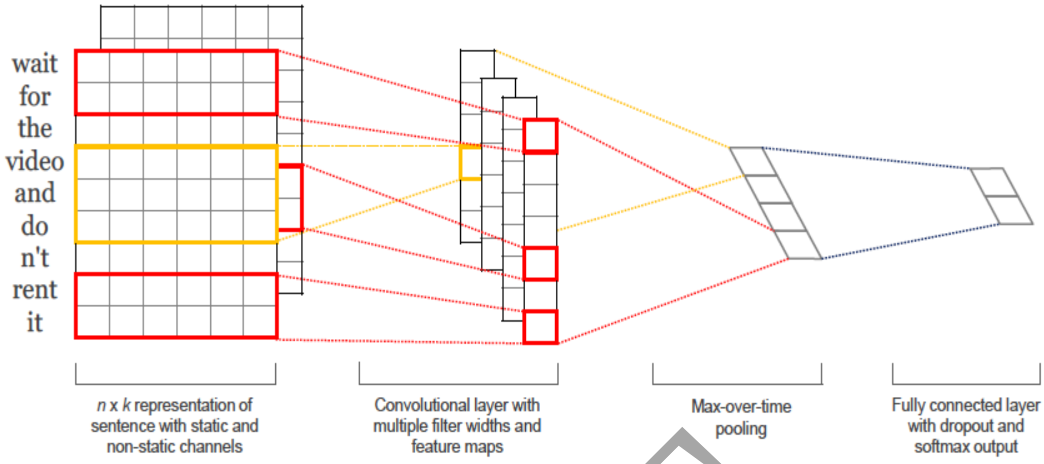


Figure 11: CNN model architecture with two channels for sentence sentiment analysis (Kim, 2014)

Following Kim (2014) using CNN in sentiment analysis, Poria et al. (2015) make an improvement by using the activation values of the hidden layers as features for a more advanced classifier and further pushed the sentiment analysis accuracy.

In the same year, Kalchbrenner et al. (2014) introduce Dynamic Convolutional Neural Network (DCNN), a variant of the standard CNN, for the semantic modeling of sentences. The key feature of DCNN is using dynamic k-max pooling, which is a generalization of max pooling, over linear sequences. In Figure 12, the dynamic k-max pooling layers have values k of 5 and 3.

PBCSF

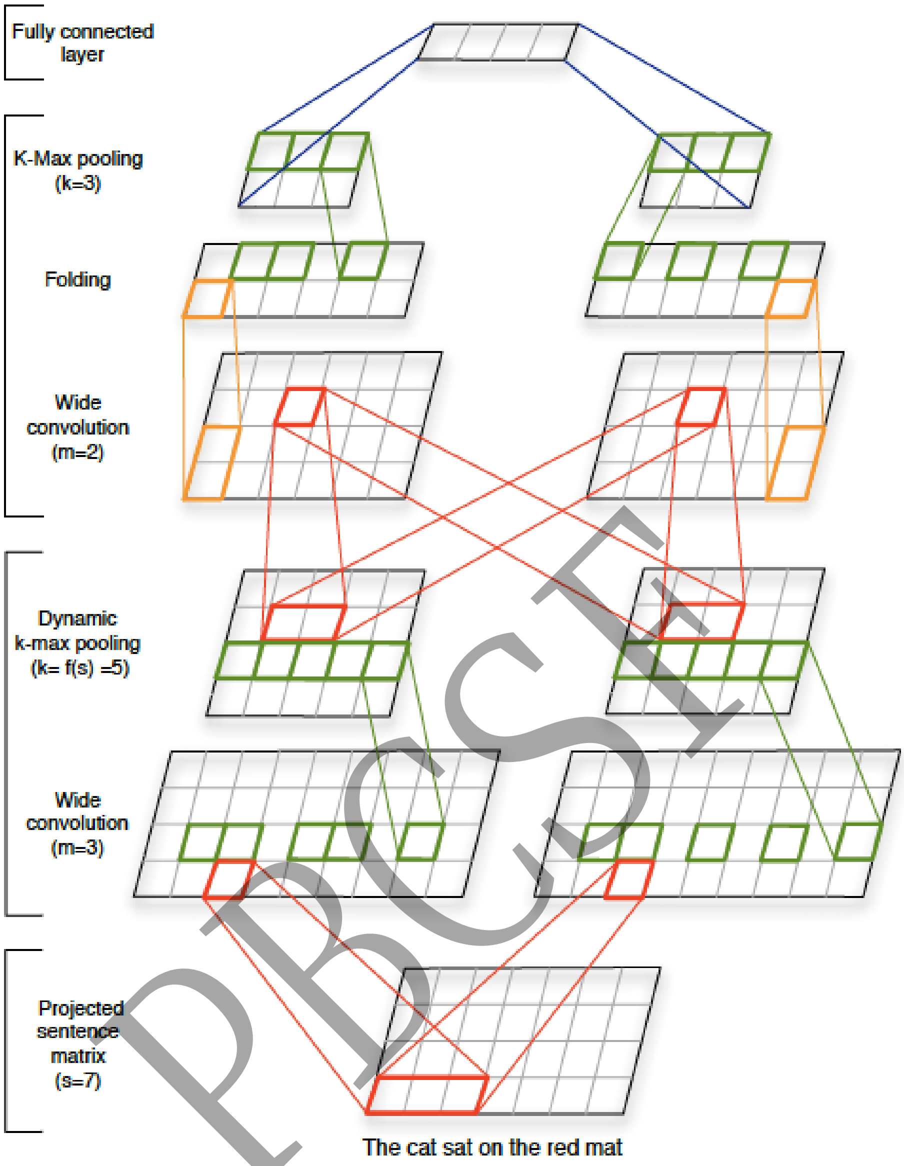


Figure 12: The DCNN architecture for input sentence (Kalchbrenner et al., 2014)

Hybrid or Ensemble Models

Earlier literature, as described above, typically adapts one of the main deep neural networks for sentiment analysis, either RNN(LSTM) or CNN. In recent years, researchers has also investigated various hybrid models with the aim of getting the benefits from both worlds. Among these, Wang et al. (2016) propose a regional CNN-LSTM model to provide fine-grained valence-arousal (VA) ratings of texts (Figure 13).

The regional CNN takes a single sentence as a region and therefore divides an entire text into several regions. This helps to extract useful information from each sentences and weight them based on their contribution to the prediction. The LSTM then sequentially integrates the output of CNN for VA prediction.

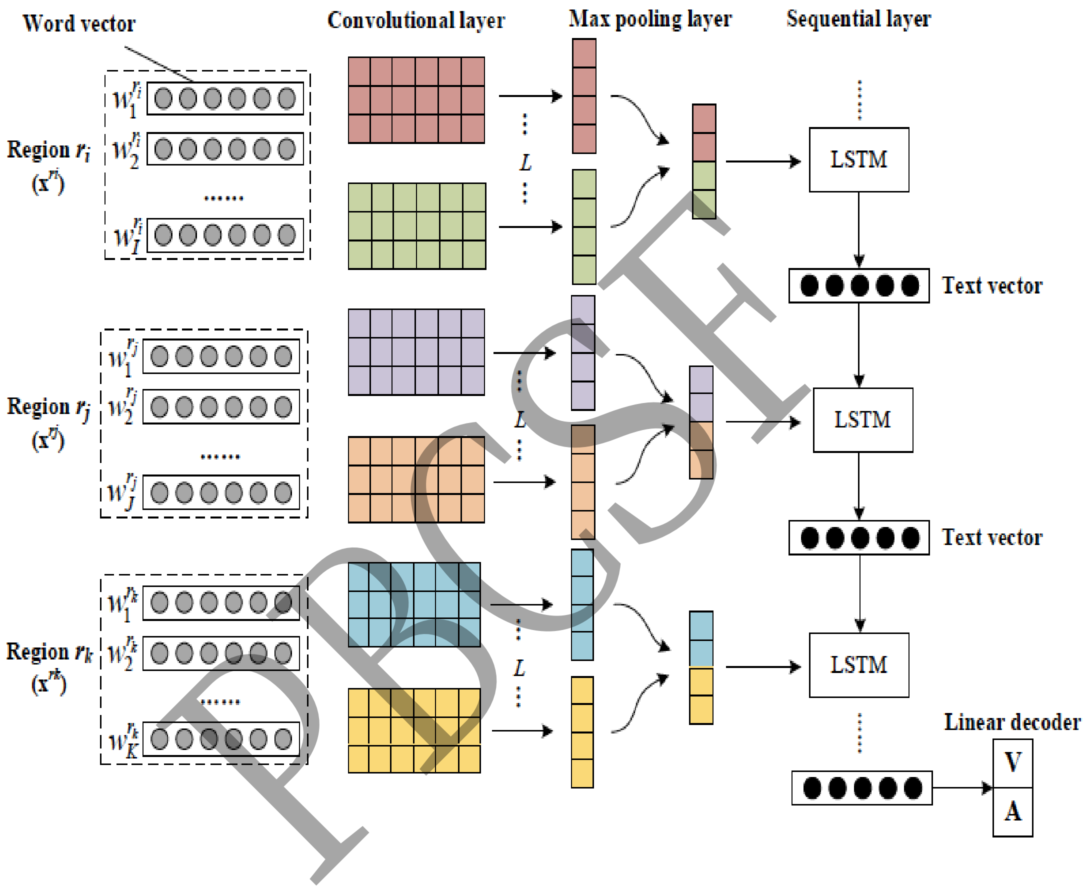
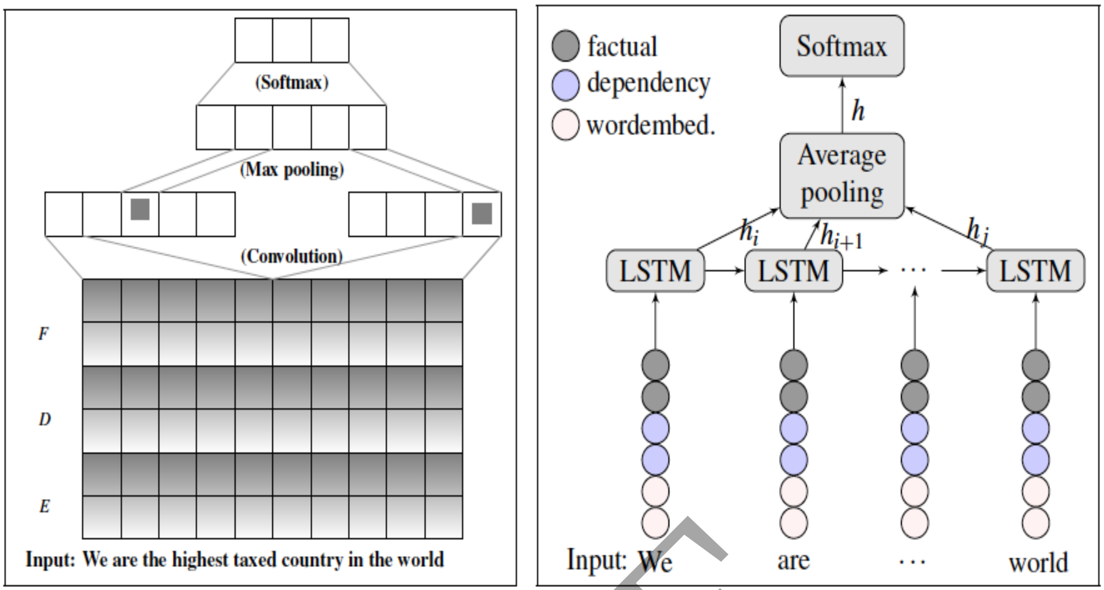


Figure 13: Architecture of regional CNN-LSTM model (Wang et al., 2016)

Guggilla et al. (2016) also describe a system that combined CNN and LSTM for processing arguments in online user interactive discourse (Figure 14).



(a) CNN with multiple word embeddings

(b) LSTM with multiple word embeddings

Figure 14: Illustration of two methods for claim classification (Guggilla et al., 2016)

Combining both deep learning techniques and the classical feature-based models, Akhtar et al. (2017) propose a multilayer perceptron (MLP)-based ensemble architecture for financial sentiment analysis. As the first step, they develop CNN, LSTM, and Gated Recurrent Unit (GRU) models, which are "trained on top of pre-trained, autoencoder-based, financial word embeddings and lexicon features". Then, these models are combined with support vector regression to form an ensemble model (Figure 15). This model achieves impressive results on SemEval-2017 shared task on financial sentiment analysis.

PBCSF

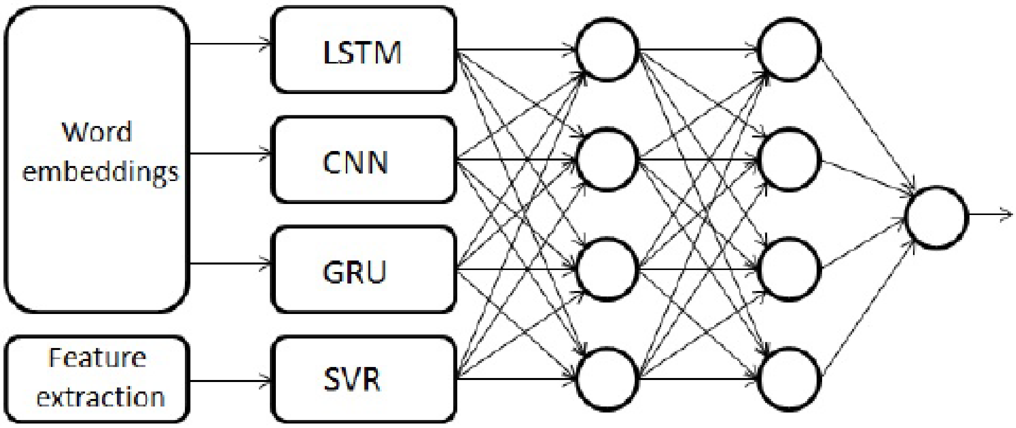


Figure 15: MLP based ensemble architecture (Akhtar et al., 2017)

3.1.2 Document-level Sentiment Analysis

Document-level sentiment refers to the assignment of an overall sentiment orientation or polarity to an entire document. Usually, it is framed as a binary classification or a regression problem (the number of classes varies). It is assumed to be more challenging than the sentence-level sentiment classification, as we have to capture the complex relationships between sentences in a document.

Obviously, it is critical to firstly develop a good documentation representation, which lays out the foundation to extract the information contained in the document's words and sentences. After this critical step, we can then feed the representation into the deep learning architectures to calculate the document-level sentiment. There is a rich literature that focus on developing a proper document representation.

The conventional bag-of-words representation is a common fixed-length feature, but it fails to take the ordering and semantics of words into account. To overcome these weaknesses, Le and Mikolov (2014) propose an unsupervised algorithm Paragraph

PBCSF

Vector to learn a fixed-length feature representation of texts. This algorithm can represent a document by a dense vector that addresses the disadvantages of bag-of-words. They apply Paragraph Vector to several benchmark sentiment analysis datasets and the experiment results outperformed that of bag-of-words.

An interesting work in this line investigates the domain adaptation for sentiment classifiers (Glorot et al., 2011). While the online reviews and recommendations are increasing exponentially, such digital text typically belongs to different domains and it is challenging to get annotated data for each of them. To address this problem, Glorot et al. (2011) extract a high-level feature representation for every review via a deep learning approach, which allows it to perform the domain adaption on a large dataset composed of 22 domains.

Tang et al. (2015a) introduce an ensemble neural network to learn vector-based document representation in two steps (Figure 16). They first use CNN or LSTM to learn sentence representation. Built up on this, this paper then utilizes the gated recurrent neural network to produce a document representation by adaptively encoding semantics of sentences as well as their relationships.

PBCSF

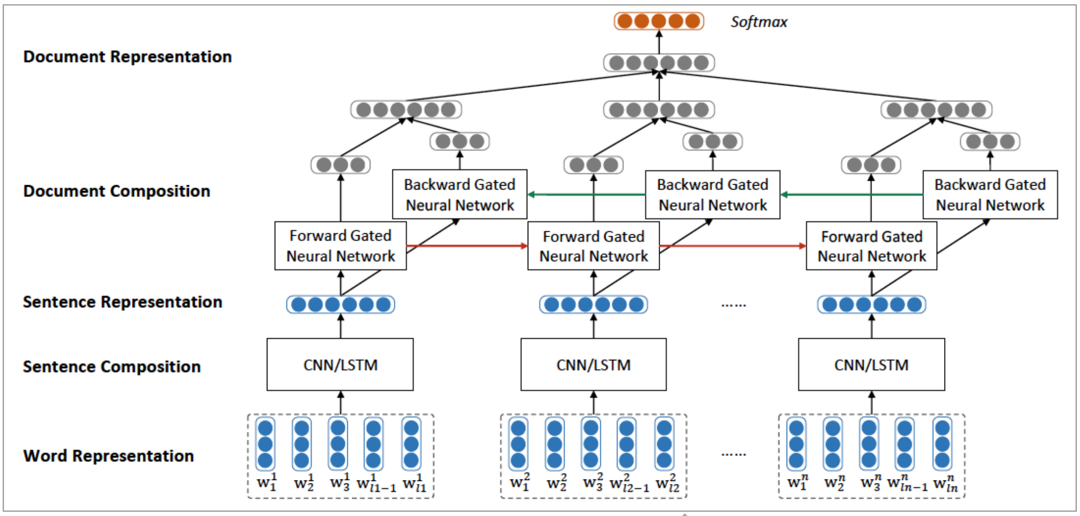


Figure 16: Model for document-level sentiment classification (Tang et al., 2015a)

In a related work, Tang et al. (2015b) incorporate the information on users who expresse the sentiment as well as the evaluated products into the document sentiment analysis (Figure 17). To do so, this paper uses vector space models to take such information into account, which contributes to developing a better document representation.

PBCSF

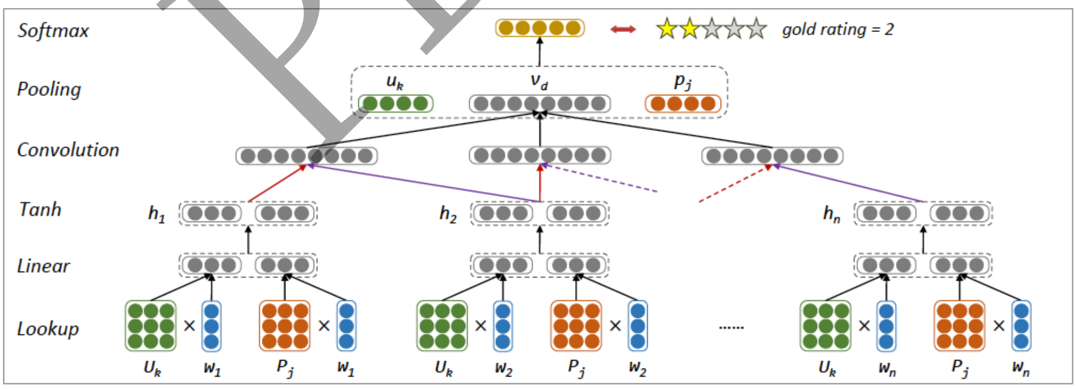


Figure 17: Neural network architecture for sentiment classification (Tang et al., 2015b)

3.1.3 Aspect-level Sentiment Analysis

In contrast to the sentence-level and the document-level sentiment classification described above, the aspect-level sentiment analysis aims to find and aggregate the sentiment on entities mentioned in either sentences or documents (Schouten and Frasincar, 2015). In this regard, the aspect-level sentiment is more challenging. It has to firstly identify the opinion targets (an entity or entity aspect), which is called the aspect detection or the aspect extraction, and then measure the corresponding sentiment. Furthermore, it is difficult to model the relationship between the target and its contextual words in the sentence or document, which also makes the aspect-level sentiment research difficult. As we'll see in the next session, recent works exploit the attention mechanism to tackle such challenges.

To begin with, we review recent works that exploit deep learning approaches for the first-step subtask in the aspect sentiment analysis: the aspect extraction. Katiyar and Cardie (2016) applies deep bidirectional LSTMs to extract both opinion entities and two kinds of relationships (IS-FROM and IS-ABOUT) among them. While standard LSTMs are not competitive, they find that introducing sentence-level and relation-level optimization enables the LSTM to achieve SOTA performance.

Different from the work using RNN, Poria et al., (2016) resort to the CNN architecture and propose to use a 7-layer deep convolutional neural network to tag each work as either aspect or non-aspect in a sentence. To further improve the accuracy, they incorporate a set of linguistic patterns developed for aspect extraction into the CNN architecture, which yield an ensemble classifier.

PBCSF

After the aspect extraction, we now turn to the second subtask in the aspect sentiment analysis, which focuses on measuring the sentiment for entities in the opinionated-text. In this subfield, we see an increasing application of aspect sentiment in customer reviews. A recent paper by Ruder et al. (2016) argues that "knowledge of the review structure and sentential context should inform the classification of each sentence". In light of this, the authors use a hierarchical bidirectional LSTM to model the interdependencies of sentences in a review for aspect sentiment task (Figure 18), which outperform non-hierarchical baselines.

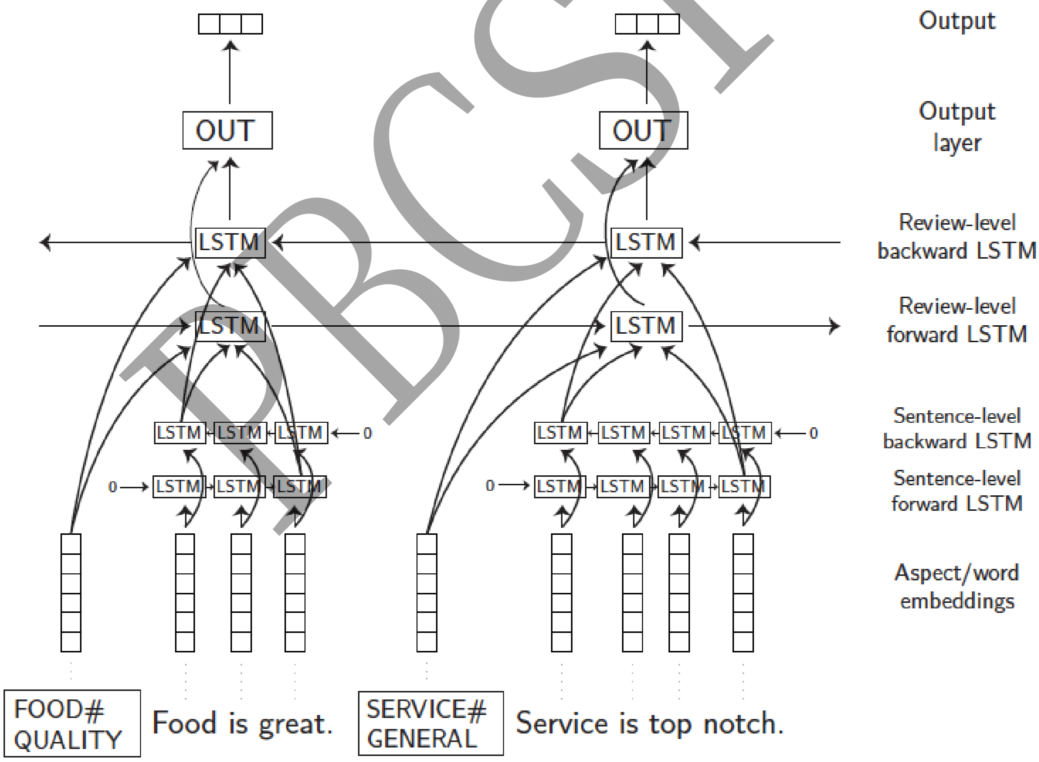


Figure 18: The hierarchical bidirectional LSTM for aspect sentiment analysis (Ruder et al., 2016)

3.2 Applying Attention Mechanism, Transformer and BERT for Sentiment

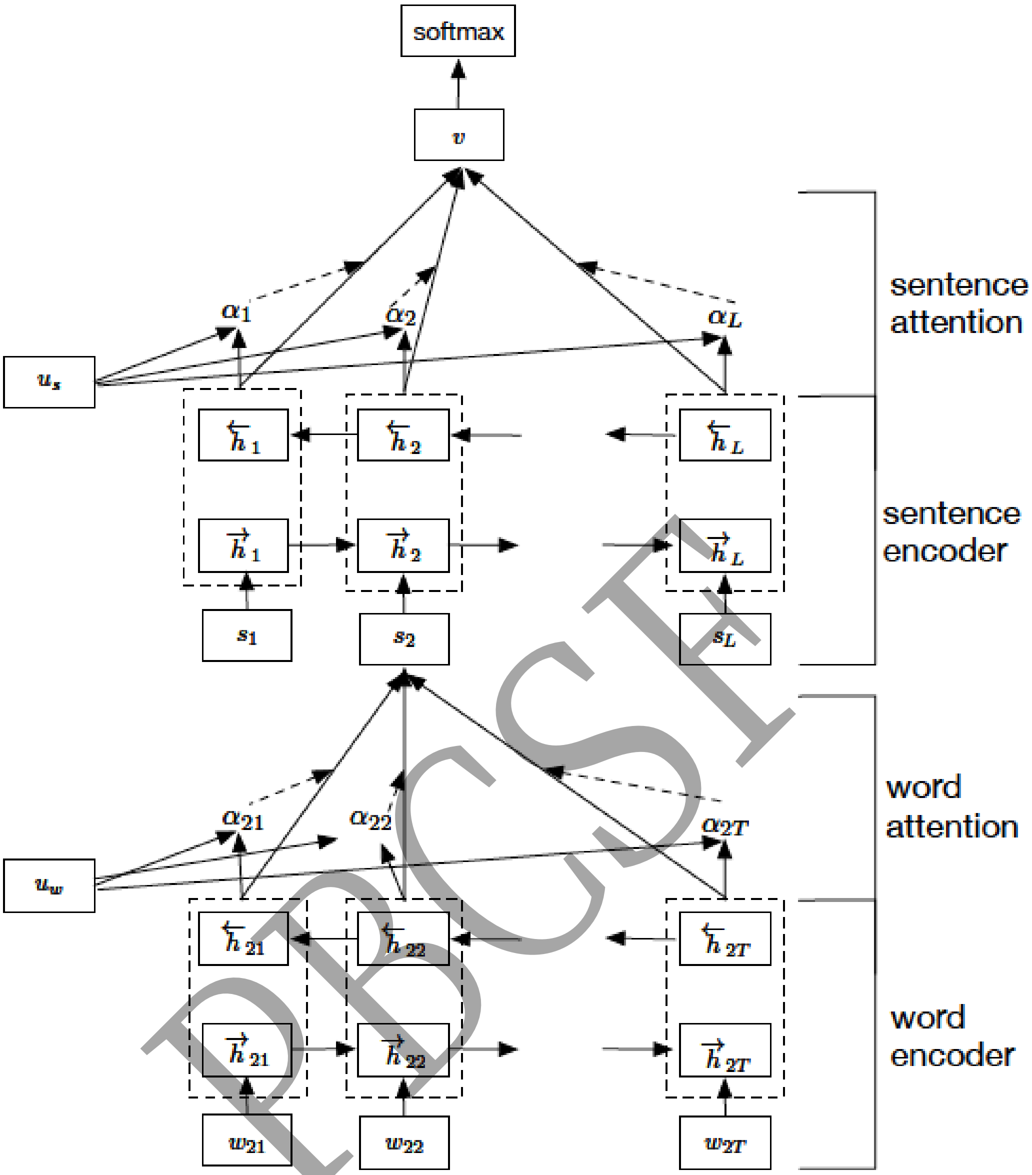
Analysis

In this section, we review the representative literature that utilizes the attention mechanism, Transformer, or BERT for Sentiment analysis. In line with the earlier section, we further group the research based on the sentiment granularity.

Document-level sentiment analysis

Yang et al. (2016) propose a hierarchical attention network for document classification. There are two distinctive features of this model (Figure 19). The first method is to use the hierarchical structures to model the hierarchical structure of the document; the second is to combine both word-level attention and sentence-level attention to pay more or less attention to words and sentences. This model outperforms previous methods on six benchmark datasets.

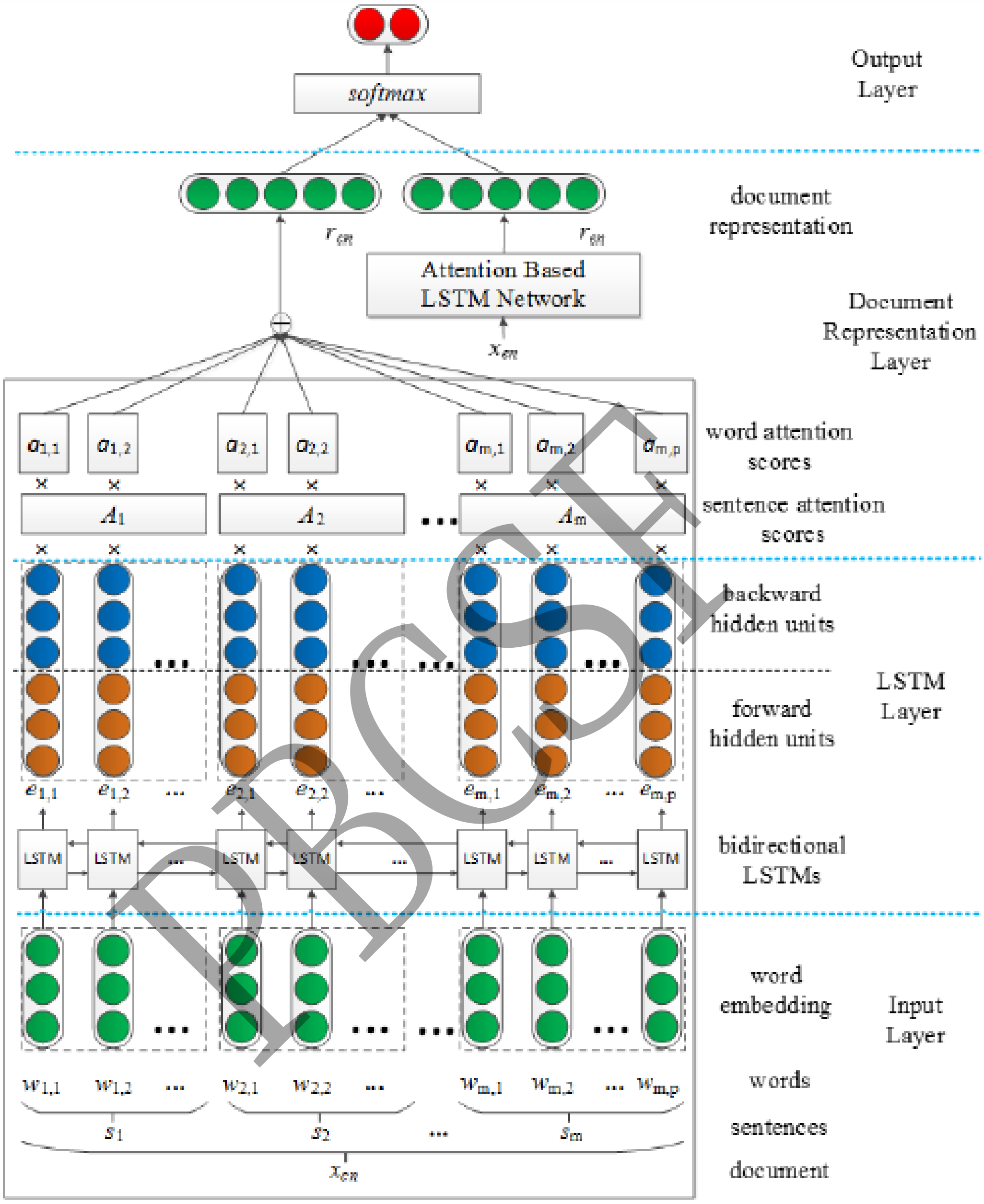
PBCSF



PBCSF

Figure 19: Hierarchical attention network (Yang et al., 2016)

As labeled resources are typically very imbalanced across languages, cross-lingual sentiment classification aims to address this challenge by adapting sentiment resources in a resource-rich language to a resource-poor language. To this end, Zhou et al., (2016) present an attention-based bilingual representation learning model to learn the distributed semantics of the documents in the source and target languages. In addition, they propose a hierarchical attention mechanism for the bilingual LSTM (Figure 20), which achieves good results on a benchmark dataset.



PBCSF

Figure 20: Attention-based LSTM architecture (Zhou et al., 2016)

Domain adaptation is a common challenge in many machine learning tasks. While standard deep learning models can learn a representation shared by different domains, limited interpretability prevents them from identifying pivots. To solve this problem, Li et al. (2017) propose an end-to-end Adversarial Memory Network (AMN) for cross-domain sentiment analysis. This framework is composed of two parameter-shared memory networks for sentiment classification and domain classification respectively, which are trained jointly (Figure 21). AMN has SOTA performance on the Amazon review dataset.

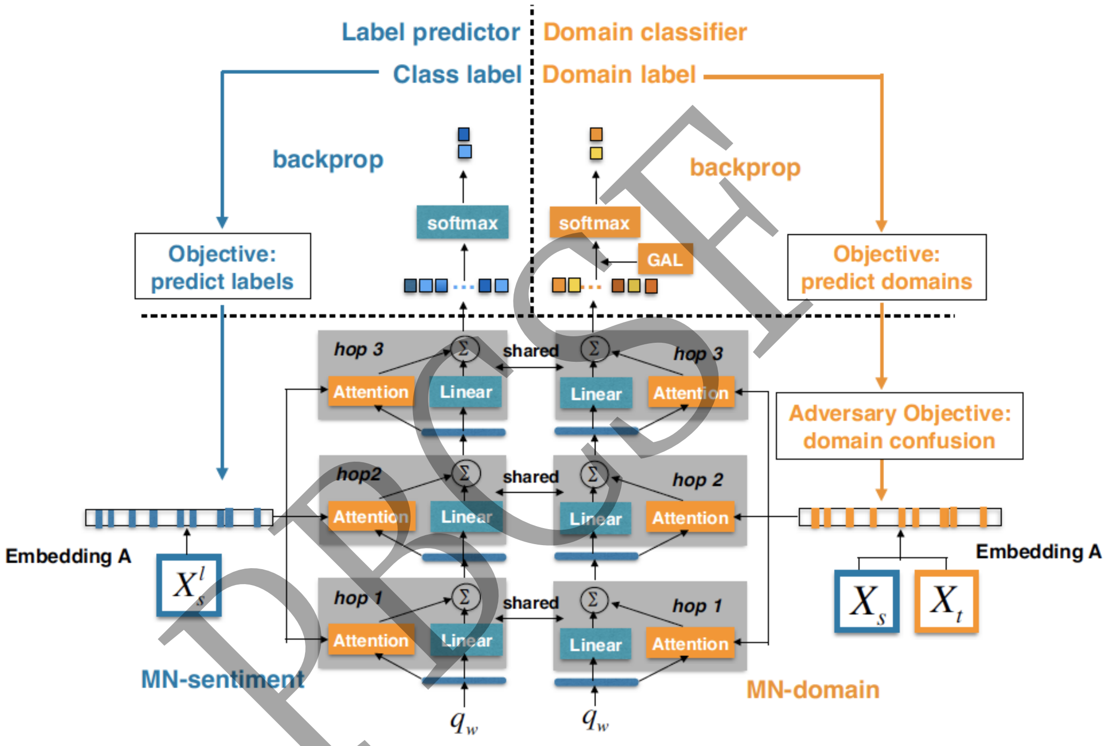


Figure 21: The Adversarial Memory Network architecture (Zhang et al., 2017)

Yin et al. (2017) focus on document-level multi-aspect sentiment classification and modele the task as a machine comprehension problem. They introduce a hierarchical iterative attention model to learn aspect-specific representations through repeated interactions between documents and aspect questions (Figure 22). The hierarchical architecture is designed to capture both word-level and sentence-level information. And the attention modules are applied to “aspect questions and document

PBCSF

alternatively with the multiple hop mechanism".

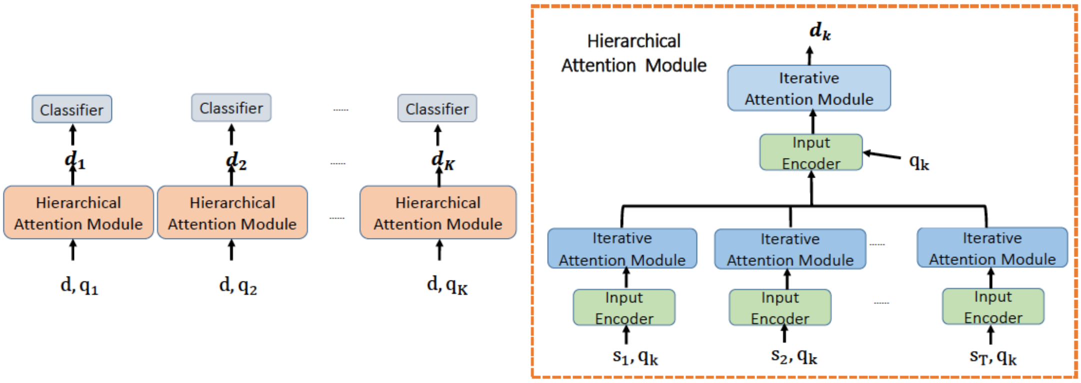


Figure 22: Model architecture (Yin et al., 2017)

<u>*Aspect-level sentiment*</u>

Chen et al. (2017) propose a similar hierarchical LSTM model to generate sentence and document representations, and then incorporate user and product information via the attention mechanism. The over architecture is displayed in the figure below.

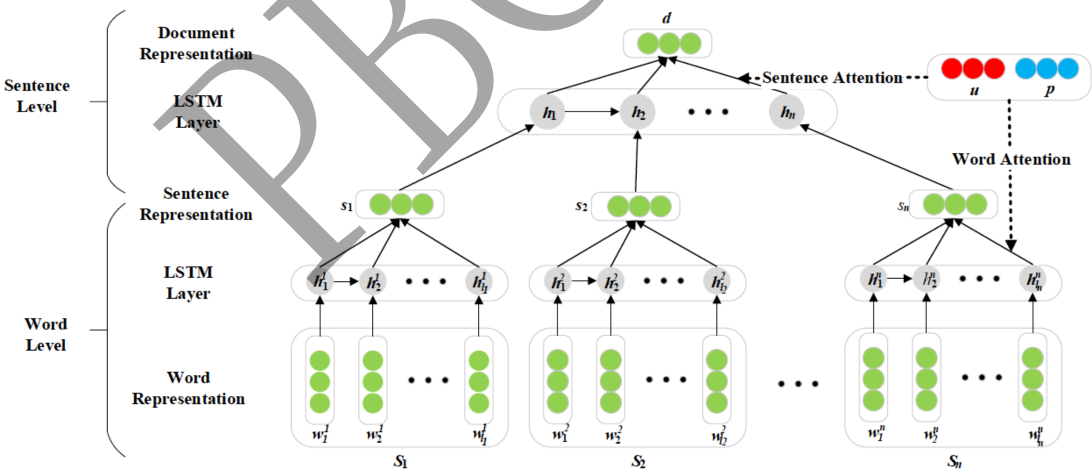


Figure 23: User product attention based neural sentiment classification (Chen et al., 2017)

Observing that the sentence-level polarity depends on both the content and the corresponding concerned aspect, Wang et al. (2016) investigate the connection between

PBCSF

an aspect and the content of a sentence. To do so, they propose an attention-based LSTM architecture for aspect-level sentiment classification. In particular, for different aspects in the input, the attention mechanism enables the model to focus on corresponding parts.

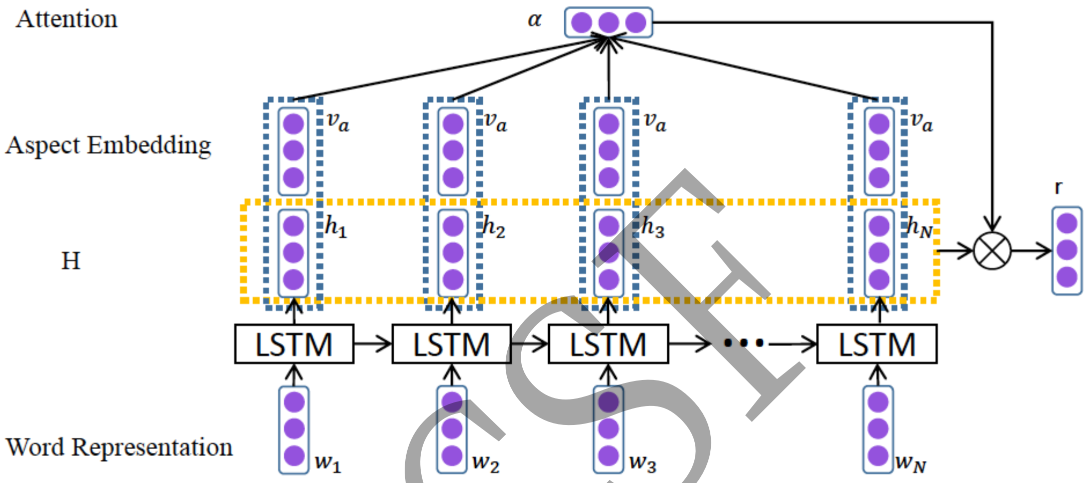


Figure 24: The architecture of attention-based LSTM

With the goal of identifying coherent aspects, He et al. (2017) use neural word embeddings to capture the distribution of word co-occurrences. In particular, they apply the attention mechanism in the training process to de-emphasize irrelevant words.

PBCSF

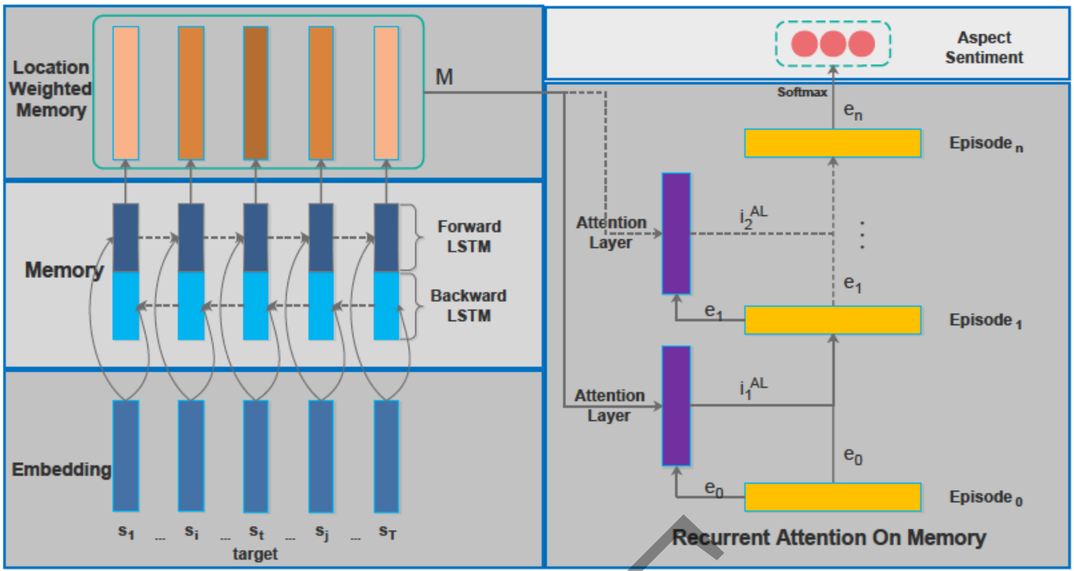


Figure 25: Model architecture (Chen et al., 2017)

As BERT has achieved breakthroughs in multiple NLP tasks, it is straightforward to exploit this powerful language representation model to tackle the challenging aspect sentiment classification. For instance, to use the pre-trained BERT for this subtask, Sun et al. (2019) construct an auxiliary sentence form the aspect and framed the aspect sentiment as a sentence-pair classification problem. The fine-tuned BERT has SOTA results on benchmark datasets.

4. Summary and Future Work

In this study, we focus on the applications of deep learning methods for sentiment analysis. We begin with an overview of the sentiment classification, which has been an ongoing and important research topic in the natural language processing field. The second section reviews the deep learning architectures with a special focus on recent breakthroughs. With this background, we then dive into representative literature that investigate deep learning approaches for three sentiment analysis tasks: the sentence-

PBCSF

level, the document-level, and the aspect-level.

We end this study by discussing some of the remaining challenges and pointing out some promising future directions. First of all, it is critical to standardize benchmark datasets and develop better performance measurements. Without this, it is difficult to effectively evaluate and compare the performance of various models and push the SOTA results. Second, developing automated sentiment analysis system is crucial to explore the exponentially increasing digital text effectively and efficiently. Last but not least, many high-stake fields, such as finance and medical care, require interpretability of the sentiment analysis model. However, the black-box nature of deep learning methods significantly limits its applications in such domains. Going forward, developing interpretable models will help release the power of machine learning methods in high-stake scenarios.

PBCSF

Reference

Mikolov, T., Chen, K., Corrado, G., and Dean, J., 2013, Efficient estimation of word representations in vector space. CoRR, abs/1301.3781.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of ACL, 2014.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In Proceedings of ACL: HLT, 142-150.

Igor Labutov, and Hod Lipson. 2013. Re-embedding words. In Proceedings of ACL: HLT, 489-493

Socher R, Pennington J, Huang E.H, Ng A.Y, and Manning C.D. Semi-supervised recursive autoencoders for predicting sentiment distributions. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2011), 2011.

Socher R, Huval B, Manning C.D, and Ng A.Y. Semantic compositionality through recursive matrix-vector spaces. In Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2012), 2012.

Socher R, Perelygin A, Wu J. Y, Chuang J, Manning C. D, Ng A. Y, and Potts C. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2013), 2013.

PBCSF

Glorot X, Bordes A, Bengio Y. Domain adaption for large-scale sentiment classification: a deep learning approach. In Proceedings of the International Conference on Machine Learning (ICML 2011), 2011.

Xin Wang, Yuanchao Liu, Chengjie Sun, Baoxun Wang, and Xiaolong Wang. "Predicting polarities of tweets by composing word embeddings with long short-term memory. 2015. In Proceedings of ACL, 1343-1353.

Zhiyang Teng, Duy Tin Vo and Yue Zhang. Context-Sensitive Lexicon Features for Neural Sentiment Analysis. In Proceedings of EMNLP 2016. Austin, Texas, USA, November.

Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. In Proceedings of ACL, 1556–1566.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In Proceedings of EMNLP, 1746-1751.

Y. LeCun, L. Bottou, Y. Bengio, P.Haffner. 1998. Gradient-based learning applied to document recognition. In Proceedings of the IEEE. 86(11): 2278-2324, November. https://ieeexplore.ieee.org/abstract/document/726791

Mikolov T, Sutskever I, Chen K, Corrado G, and Dean J. Distributed representations of words and phrases and their compositionality. In Proceedings of the Annual Conference on Advances in Neural Information Processing Systems (NIPS 2013), 2013.

PBCSF

Poria S, Cambria E, Gelbukh A. Deep convolutional neural text features and multiple kernel learning for utterance-level multimodal sentiment analysis. In Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2015), 2015.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In Proceedings of ACL, 655-665.

Wang J, Yu L-C, Lai R.K., and Zhang X. Dimensional sentiment analysis using a regional CNN-LSTM model. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2016), 2016.

Guggilla C, Miller T, Gurevych I. CNN-and LSTM-based claim classification in online user comments. In Proceedings of the International Conference on Computational Linguistics (COLING 2016), 2016.

Akhtar MS, Kumar A, Ghosal D, Ekbal A, and Bhattacharyya P. A multilayer perceptron based ensemble technique for fine-grained financial sentiment analysis. In Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2017), 2017.

Le Q, Mikolov T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning (ICML 2014), 2014.

Liu B. Sentiment analysis: mining opinions, sentiments, and emotions. The Cambridge University Press, 2015.

Tang D, Qin B, Liu T. Document modelling with gated recurrent neural network

PBCSF

for sentiment classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), 2015a.

Tang D, Qin B, Liu T. Learning semantic representations of users and products for document level sentiment classification. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2015), 2015b.

Zhang A, Liptton Z, Li M, Smola A, 2020. Dive into deep learning.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Advances in Neural Information Processing Systems, pages 6000-6010.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805, 2018.

Yang Z, Yang D, Dyer C, He X, Smola AJ, and Hovy EH. Hierarchical attention networks for document classification. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2016), 2016.

Katiyar A, Cardie C. Investigating LSTMs for joint extraction of opinion entities and relations. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2016), 2016.

Poria S, Cambria E, Gelbukh A. Aspect extraction for opinion mining with a deep convolutional neural network. Journal of Knowledge-based Systems. 2016.

PBCSF

Ruder S, Ghaffari P, Breslin JG. A hierarchical model of reviews for aspect-based sentiment analysis. In Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2016), 2016.

Chen P, Sun Z, Bing L, and Yang W. Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2017), 2017.

Zhou X, Wan X, Xiao J. Attention-based LSTM network for cross-lingual sentiment classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), 2016.

Li Z, Zhang Y, Wei Y, Wu Y, and Yang Q. End-to-end adversarial memory network for cross-domain sentiment classification. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2017), 2017.

Yin Y, Song Y, Zhang M. Document-level multi-aspect sentiment classification as machine comprehension. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2017), 2017.

Wang Y, Huang M, Zhu X, and Zhao L. Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), 2016.

He R, Lee WS, Ng HT, and Dahlmeier D. An unsupervised neural attention model for aspect extraction. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2017), 2017.

PBCSF

Chen H, Sun M, Tu C, Lin Y, and Liu Z. Neural sentiment classification with user and product attention. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2016), 2016.

Chen P, Sun Z, Bing L, and Yang W. Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP 2017), 2017.

Sun C, Huang L, Qiu X. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2019), 2019.

Kim Schouten, Flavius Frasincar, Survey on aspect-level sentiment analysis, IEEE, Volume: 28 Issue:3, 2015.

PBCSF

联系人: 张娜 邮 箱: zhangna@pbcsf.tsinghua.edu.cn