An Overview on Fine-grained Text Sentiment Analysis: Survey and Challenges

Xiaoting Guo*, Wei Yu, Xiaodong Wang

*Computer School, National University of Defence Technology, No.109, Deya Road, Changsha 410073, China

*guoxiaoting18@nudt.edu.cn

Abstract. Among various natural language process tasks, sentiment analysis has always been a research hotspot. From the initial sentence-level and document-level coarse-grained sentiment analysis to recent fine-grained sentiment analysis on the aspect word level, researchers are committed to applying diverse methods to obtain better sentiment analysis results, ranging from lexicon-based, statistical machine learning methods to deep learning models. In the change of technology, several benchmark datasets that can be used for model performance comparison are gradually yielded. This article summarizes the current research status of aspect-level text sentiment analysis from multiple dimensions such as dataset, mainstream methods, and evaluation indicators, finally, it puts forward the challenges facing and potential research directions from a unique perspective.

Keywords: aspect sentiment analysis, lexicon, machine learning, deep learning.

1. Introduction
With the advent of the era of artificial intelligence, digital information has shown an exponential growth trend. The seemingly isolated and scattered information hide complex and diverse relationships. Under the background of increasingly mature data mining technology, natural language processing (NLP) technology came into being. People need to help machines understand natural language, learn semantic relations, and even generate natural language text. Text sentiment analysis, also known as opinion mining, refers to using natural language processing, text mining, and computer linguistics to identify and extract subjective information in the original material. Sentiment analysis [1][2][3] has a very wide range of uses, such as customer preference analysis, public opinion mining, hot spot event position judgment, etc., which have extremely high potential application value. With the rise of natural language processing technology, researchers' passion for sentiment analysis has only increased.

According to the granularity of the analysis object, research can be divided into document level sentiment analysis[4][5], sentence level sentiment analysis[6][7], and aspect level sentiment analysis; The former coarse-grained analysis is designed to calculate the overall sentiment expressed by opinion holders based on the content of the text, which conducts an overall tendency with regard to all opinion words in the text. Pang et al.[6] and Kim et al.[7] respectively proposed supervised classification and had once been the baseline for subjective sentence sentiment analysis. When analysing the sentiment for something, one might be concerned about not only whether people are talking with a positive,
neutral, or negative attitude on the product, but also which specific aspects or features of the product people talk about. That’s what Aspect-Based Sentiment Analysis is about. We want to Fig out what is the top driver that makes them happy, satisfied, or not[8]. Apparently, aspect level sentiment analysis serves better on feedbacks, references, and vertical analysis.

According to whether the form of analysis content is unique, sentiment analysis can be divided into single-modal sentiment analysis and multi-modal sentiment analysis. The content of multi-modal data sentiment analysis may have multiple forms such as text, video, pictures, etc.[9][10]Multi-modal data can complement each other, but how to integrate multi-modal data and use the alignment information between different modal data to model the association between different modal data still faces great challenges. This study will focus on fine-grained text sentiment analysis.

The rest of the paper is structured as follows: Section 2 presents a comprehensive overview of the framework for aspect level sentiment analysis and reviews some of the most popular datasets that are commonly used in most existing studies. Section 3 introduces novel models and their characteristics. Section 4 discusses the main challenges and potential directions for deep learning-based text sentiment analysis. Section 5 concludes the paper.

2. Framework
Aspect level sentiment analysis aims at judging sentiment polarity for each aspect in a given review. It can be formulated as a triples \( \text{Sen} = (a, h, o) \), while \( a \) refers aspect term, \( o \) refers opinion on corresponding \( a \) and \( h \) refers opinion holder. In most cases, especially for a product review, feedback, and comments field, people tend to ignore the attribute of the opinion holder, since the opinion itself is much more important and meaningful than the opinion holder. Generally, aspect level sentiment analysis consists of two sub-tasks: detect the aspect terms in the given review and find the sentiment corresponding to the aspect terms detected, which is also called end-to-end aspect-based sentiment analysis[11].

2.1. Datasets
In early 2004, Hu et.al [12] came up with an aspect level dataset about digital products. Dong et al. [13] propose the largest corpus of Twitter phrase-level sentiment classification in 2014. At the same time, Semeval Put forward two important corpora of aspect-level sentiment analysis, becoming a benchmark for many studies, which greatly facilitates subsequent research. Zhou et al. [14]collected some of them and translated them into a uniform XML/JSON format. Compared with English datasets, Chinese datasets for sentiment analysis are fewer and scattered. In the early days, there were NLPCC datasets on microblogs sentiment recognition, and AI challenger datasets are not available now. Tab 1 collects the field and source of benchmark datasets recently used both in English and Chinese studies.

| Datasets  | Content | Source                                      |   |
|-----------|---------|---------------------------------------------|---|
| SemEval-14| Laptop, restaurant                           | http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools |
| Hu and Liu| Digital product                              | https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html |
| Movie Review| Movie                      | http://www.cs.cornell.edu/people/pabo/movie-review-data/ |
| SST       | Movie/TV                                     | https://nlp.stanford.edu/sentiment/ |
| Twitter   | tweets                                       | http://goo.gl/5Enpu7 |
| MPQA      | news                                          | http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/ |
| NLPCC     | microblog                                    | http://tcci.ccf.org.cn/conference/2013/pages/page04_tdata.html |

Tab 1 popular datasets for sentiment analysis and download source

2.2. Generalized process
Aspect terms can be implicit and explicit. No matter which kind of the term is, people usually map it into a pre-defined aspect category, so that they can obtain an overall sentiment distribution upon a specific aspect they concern about.

The methods can be grouped into the unsupervised lexicon-based method, supervised statistic machine learning, and deep learning-based method. The generalized process contains word representation, aspect, and opinion of term detection, sentiment judgment.

2.2.1. Word representation
Word presentation is a vital part of model language and capture features, measure the similarity between terms, and be the vector input for deep learning model to learn semantic relations. Word2vec[15], Glove[16], Bert[17] have been applied to sentiment analysis in existing research. As is known to us, word2vec and glove are context-free distributed representations that encode latent semantic information, while Bert, as an outstanding representative of the pre-trained model, provides context-dependent dynamic word vectors, and people can fine-tune the domain-adapted word vectors according to their needs. Tao[18] designed an experiment on text CNN, BiLSTM, AT-BiLSTM to compare how different word vector representations would affect the model’s performance. The experimental results show that the above three-word embedding methods all have achieved good results, among which Bert has the best effect, the glove has the shortest running time, and the word2vec effect and running time are both in the middle. Therefore, one can choose word2vec for the trade-off between performance and time. Bert has better performance in the field of generalization due to its powerful feature extraction capabilities, if you are pursuing the ultimate effect and have sufficient computing resources, you can choose Bert. Traditional word embeddings cannot address the polysemy problem and fail to capture the sentiment of words. Specifically, Naseem[19] studied the performance of transformer-based embedding, contextual embedding, glove embedding, part of speech (POS) embedding, and lexicon embedding on three airline datasets. Through their ablation experiments, contextual embedding ELMO plays a more important role in addressing such polysemy and assign different sentiment to words than transformer-based embedding Bert.

2.2.2. Aspect and opinion term detection
Term detection for both aspect and opinion is a fundamental part. Methods for this part can be grouped into the unsupervised lexicon-based method, statistical machine learning method, and deep learning-based model[20][21][22][23]. Unsupervised lexicon-based (WordNet, NTUSD[24], MPQA[25], Hownet, Chinese emotional vocabulary ontology[26]) method mainly utilizes given lexicon and syntactic analyser to detect opinion terms. Meanwhile, aspect terms, negative words, adverbs of degree, etc will be detected with dependency relations upon different dependency parser. Statistical machine learning and deep learning-based methods are mainly classified into supervised and unsupervised ones. The unsupervised approach includes topic modeling, frequent pattern mining, and label propagation, while supervised one always can be attributed to solving sequence labeling issues. He et al. [27] proposed an unsupervised neural attention model for aspect extraction and achieved high performance both on restaurant corpus and beer corpus.

2.2.3. Sentiment judgment
Sentiment judgment is the final procedure, aims at deducing the corresponding sentiment polarity or result with former information. In classical unsupervised lexicon-based methods, people obtain opinion term’s polarity according to the given lexicon and define rules for calculating sentiment scores regarding negative words, adverbs of degree, etc. Statistical machine learning methods like Naïve Bayesian (NB), maximum entropy model (ME), Support Vector Machine (SVM), always rely on the large labeled corpus to train a classifier model[28]. Deep learning outperforms them in many tasks of natural language processing due to its stronger feature extraction and learning capabilities, so people have started many attempts to apply deep learning to sentiment analysis. The basic combination of long-term and short-term memory networks and attention mechanisms became popular due to their
ability of capturing long-distance and important information[29][30][31][32]. Zhang et al.[33] and Zhou et al. organized and summarized the classic deep learning models base on CNN, RNN, Memory, and attention in detail. We will introduce the latest progress of deep learning-based methods in the past two years in Section 3.

2.3. Evaluation metrics
Sentiment analysis can be regarded as a classification problem to a certain extent, so its evaluation criteria can be accuracy, precision, recall, and F-score. And these indicators have been indeed widely used in the evaluation of various sentiment analysis. For sentiment analysis, the mapping between aspect words and aspect categories and the determination of sentiment polarity all involve multi-label classification issues. In order to measure the overall effect of the classifier more fairly and objectively, macro-average is introduced. Take the arithmetic average of F1 values of all classes to get macro-average, and macro-average indicators have been increasingly used in the evaluation of sentiment analysis in recent years.

3. Latest attempts in deep-learning-based methods

3.1. Capsule Network
The capsule network[34] proposed by Hinton compensates for the representation defects of CNN by extracting features in the form of vectors. The starting point is to build more structures in the neural network, and then hope that these new structures can help the model to generalize better. Wang et al. [35]made the first attempt to perform sentiment analysis by a capsule model based on RNN, he designed a simple capsule structure with an attribute, a state, three models respectively, and made each capsule focus on one specific sentiment category. Their work achieved SOTA performance without linguistic knowledge on benchmark RW and SST in 2018. Wang et al. [36]proposed the aspect-level sentiment capsules model (AS-Capsules) in the second year, which can perform aspect detection and aspect-level sentiment classification jointly and effectively. They utilized the correlation between aspects and corresponding sentiments through shared components. The experimental results showed a great performance in the SemEval14 restaurant corpus. Furthermore, they validated the robustness of the model by directly applying the model to other restaurant reviews on Yelp. Du et al. [37]proposed a capsule network with interactive attention (IACapsNet) to construct vector feature representations by EM routing algorithm to cluster features, which produced a SOTA result on benchmark SemEval14 and twitter. Most recently, Su et al.[38] combined XLNet and capsule network, called XLNetCN. They generated global sentence aspect-related representations by constructing auxiliary and local feature representations by employing a capsule network with a dynamic routing algorithm. Their work highly addressed aspect-specific and local feature-aware issues and achieved the new state-of-the-art result on SemEval14. In summary, the capsule model is a promising part that can be used in base models to capture hierarchical information.

3.2. Transfer Learning
Existing research on end-to-end sentiment analysis based on evaluation objects only considers the performance of a single domain, while ignoring the ability to migrate and generalize across domains. High-quality data annotation is expensive, the application of transfer learning to the adaption of the various domain is feasible. Fang et al.[39] used a transfer learning model base on Bert to address the problem of pre-defined knowledge structures missing. Li et al.[40] proposed a dual-memory interaction (DMI) mechanism to automatically capture the hidden relationship between the evaluator and the opinion word. DMI infers the relationship expression of each word by interacting multiple times with the local memory (LSTM hidden state) and the global evaluator memory, the point word memory so that it can be used to sequence the knowledge migrated under the task. Furthermore, they came up with selective adversarial learning (SAL) to align vital aspect terms in a sentence. All these
settings make a significant effect on leveraging knowledge from the relative domain to the target
domain, which is a meaningful direction for cross-domain migration and generalization capabilities.

3.3. Bert-based model
As we all know, Bert can generate context-related word vectors, so in terms of a better representation
of text semantic information, Bert can often achieve unexpected results. Recently, the research on
sentiment analysis has also made good progress. Li et al. used pre-trained Bert to generate context-
dependent word vectors, and then combined BERT with a standard neural sequence model to solve the
E2E-ABSA sequence labeling problem. Their experimental results found out that with Bert, the
simplest downstream task model could outperform the best model before[41]; continued to increase
the complexity of the downstream network, and the effect of the model could be further improved. Li
et al.[42] argued that Bert cannot provide enough auxiliary information to distinguish different aspects
restricted to its limited input. To solve that, Li et al. proposed to use a gating mechanism to control the
propagation of sentiment features from Bert output with context-related representations. They kept the
most relative context-aware information and achieved new SOTA performance on SemEval14. Based
on these results achieved by Bert, how to apply the lightweight Bert model to sentiment analysis is
also a direction worth exploring.

4. Challenges and opportunities

4.1. More intelligent embedding using external knowledge
Most studies focus on exploring the emotional tendency of all sentiment terms to the corresponding
aspect term, while ignoring the aspect terms themselves may have the characteristics of emotional
tendencies. Mass et al. learned word embeddings that can capture both semantic and sentiment
information. Tang also did some work on sentiment-specific word embeddings. However, once two
opposite aspect terms retouched by the same polarity opinion term will convey different emotions,
such as good endurance and fast battery consumption. Opinion terms may be both judged as positive,
and aspect terms will be mapped into the battery category. In such cases, the machine will wrongly
analyse the second description. Therefore, it is essential to introduce external knowledge to learn more
rich semantic word representation and consider the category alignment and emotional distinction
between aspect words. Naseem’s successful attempt for mixing diverse embedding methods to make
full use of contextual information, word relationships, and natural representation is also strong
evidence.

4.2. Uncertainty-aware for correcting analysis results
Standard deep learning models usually do not capture model uncertainty while Bayesian probability
theory considers model uncertainty with the high computational cost. Gal et al.[43] developed a new
theoretical framework that uses dropout training in deep neural networks (NNs) as approximate
Bayesian inference during deep Gaussian processes. Like the idea of latent Dirichlet allocation, the
model parameter of a Bayesian neural network is no longer a deterministic value, but a distribution.
The entropy is computed by minimizing the KL scatter of the parameter distribution from the real
model parameter distribution to obtain the uncertainty. Since Bayesian uncertainty has achieved high
fining capacity on sequence labeling and named entity recognition, joint extraction of aspects and
opinion terms, which can be treated as a sequence labeling to some extent, may also benefit from that.

5. Conclusion
We listed the most used datasets intuitively and introduced various novel methods and architectures
applied to diverse applications in aspect level sentiment analysis. Moreover, with a discussion of
possible limits upon existing works, we propose corresponding improvement directions. Text
sentiment analysis will show more promising results both on sentiment classification and the
prediction of sentiment changing trend with advanced research.
References
[1] Bo P and Lillian L 2008 Opinion mining and sentiment analysis Foundations and trends in information retrieval 2 pp 1-135
[2] Priyanka T and R C T 2019 A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)
[3] Minghao H, Yuxing P and Zhen H 2019 Open-Domain Targeted Sentiment Analysis via Span-Based Extraction and Classification 57th Annual Meeting of the Association-for-Computational-Linguistics pp 46-537
[4] Duyu T, Bing Q and Ting L 2015 Learning Semantic Representations of Users and Products for Document Level Sentiment Classification 53th Annual Meeting of the Association-for-Computational-Linguistics pp 23-1014
[5] Jiacheng X, Danlu Ch, Xipeng Q and Xuanqig H 2016 Proc. of the Conf. on Empirical Methods in Natural Language Processing pp 69-1660
[6] Bo P, Lillian L, Shivakumar V 2002 Proc. of the 7th Conf. on Empirical Methods in Natural Language Processing (USA) 10 pp 79–86
[7] Kim and Yoon 2014 Convolutional neural networks for sentence classification Proc. of the Conf. on Empirical Methods in Natural Language Processing pp 51-1746
[8] Hai Ha D, P W C P, Angelika M and Abeer A 2019 Deep learning for aspect-based sentiment analysis: a comparative review Expert Systems with Applications 118 pp 99-272.
[9] Qiuqian L, Dong Z, Liangming W and Shouzhan L 2019 Computer Science 46 pp 181-185
[10] Yongzhi T, Bin Z and Jianhui H 2018 Topic-Adaptive Sentiment Analysis on Tweets via Learning from Multi-sources Data. Proc. of 10th Int. Symp. on Computational Intelligence and Design pp 241-246
[11] Xin L, Lidong B, Wenxuan Z and Wai I 2019 Exploiting BERT for end-to-end aspect-based sentiment analysis Proc. of the 2019 EMNLP Workshop W-NUT pp 34-41
[12] Mingqiong H and Bing L 2004 Mining and summarizing customer reviews Proc. of the 10th ACM SIGKDD pp 77-168
[13] Li D, Furu W, Chuanqi T, Duyu T, Ming Z and Ke X 2014 Adaptive recursive neural network for target-dependent twitter sentiment classification Proc. of the 52nd annual meeting of the association for computational linguistics 2 pp 49-54
[14] Jie Z, Xiangji H, Qin C, Qinmin H, Tingting W and Liang H 2019 IEEE Access 7 pp 78454-83
[15] Tomas M, Ilya S, Kai C, Greg C, Jeffrey D 2013 Advances in neural information processing systems pp 19-3111
[16] Jeffrey P, Richard S and Christopher M 2014 Glove: Global vectors for word representation Proc. of the 2014 conf. on EMNLP pp 43-1532
[17] Jacob D, Mingwei C, Kenton L and Kristina T 2019 Proc. of the 2019 conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 1 pp 86-4171
[18] Yangming T 2020 deep learning-based fine-grained text sentiment analysis research [D] (Hangzhou) Zhejiang Gongshang University
[19] Usman N, Imran R, Kataryza M and Muhammad I 2020 Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis Future Generation Computer Systems 113 pp 58-69
[20] Fazal Masud K, Aurangzeb K, Shakeel A, Muhammad Zubair A 2014 Lexicon-based sentiment analysis in the social web Journal of Basic and Applied Scientific Research 4 pp 48-238
[21] Minglie H, Qiao Q, and Xiaoyan Z 2017 Encoding Syntactic Knowledge in Neural Networks for Sentiment Classification ACM Transactions on Information Systems 35 pp 1-27
[22] Duyu T, Bing Q and Ting L 2015 Document Modeling with Gated Recurrent Neural Network for Sentiment Classification Proc. of the conf. on Empirical Methods in Natural Language Processing pp 32-1422
[23] Xia H, Jialiang T, Huiji G and Huan L 2013 Unsupervised sentiment analysis with emotional signals Proc. of the 22nd int. conf. on World Wide Web pp 607-618
[24] Lunwei K, Hsinhsi C 2007 Journal of the American Society for Information Science and Technology 58 pp 50-1838
[25] Janyce W, Theresa W and Claire C 2005 Annotating expressions of opinions and emotions in language Language resources and evaluation 39 pp 165-210
[26] Jianmei C 2009 Construction and Application of Ontology of Chinese Emotional Vocabulary [D] (Dalian)
[27] Ruidan H, Wee Sun L, Hwee Tou N, Daniel D 2017 Proc. of the 55th Annual Meeting of the Association for Computational Linguistics 1 pp 388-397
[28] E. F. E. M and F.A. P 2014 Sentiment analysis: Bayesian ensemble learning Decision support systems 68 pp 26-38
[29] Xiaodan Z, Parinaz S and Hongyu G 2015 Long short-term memory over recursive structures Proc. of the 32nd Int. Conf. on Machine Learning pp 12-1604
[30] Yequan W, Minlie H, Xiaoyan Z and Li Z 2016 Attention-based LSTM for aspect-level sentiment classification Proc. of the conf. on EMNLP pp 15-606
[31] Christos B, Nikos P and Christos D 2017 Proc. of the 11th int. workshop on semantic evaluation pp 747-754
[32] Peng C, Zhongqian S, Lidong B and Wei Y 2017 Recurrent attention network on memory for aspect sentiment analysis Proc. of the conf. on EMNLP pp 452-61
[33] Lei Z, Shuai W and Bing L 2018 WIREs Data Mining Knowl Discov 8 p 1253
[34] Sara S, Nicholas F, Geoffrey E H 2017 Dynamic routing between capsules Advances in neural information processing systems pp 66-3856
[35] Yequan W, Aixin S, Jialong H, Y. L and Xiaoyan Z 2018 Sentiment analysis by capsules Proc. of the 2018 world wide web conf. pp 74-1165
[36] Yequan W, Aixin S, Minlie H and Xiaoyan Z 2019 Aspect-level sentiment analysis using as-capsules The World Wide Web Conf. pp 44-2033
[37] Chunning D, Haifeng S, Jingyu W, Qi Q, Jianxin L, Tong X and Ming L 2019 Capsule network with interactive attention for aspect-level sentiment classification Proc. of the conf. on EMNLP pp 501-5492
[38] Jindian S, Shanshan Y and Da L 2020 Enhancing Aspect-based Sentiment Classification with Auxiliary Sentence and Domain Knowledge J. IEEE Access 8 pp 61-100551
[39] Xing F and Jie T 2019 A Transfer Learning based Approach for Aspect Based Sentiment Analysis 6th Int. Conf. on SNAMS pp 83-478
[40] Zheng L, Xin L, Ying W, Lidong B, Yu Z and Qiang Y 2019 Transferable End-to-End Aspect-based Sentiment Analysis with Selective Adversarial Learning EMNLP pp 92-4582
[41] Huaishao L, Tianrui L, Bing L and Junbo Z 2019 DOER: Dual cross-shared RNN for aspect term-polarity co-extraction (preprint arXiv:1906.01794)
[42] Xinlong L, Xingyu F, Guangluan X, Yang Y, Jiuniu W, Li J, Qing L and Tianyuan X 2020 Enhancing BERT Representation with Context-Aware Embedding for Aspect-Based Sentiment Analysis IEEE Access 8 pp 76-46868
[43] Yarin G and Zoubin G 2016 Dropout as a bayesian approximation: Representing model uncertainty in deep learning Int. conf. on Machine Learning pp 59-1050