

Entity-level Sentiment Analysis in Contact Center Telephone Conversations

Xue-Yong Fu, Cheng Chen, Md Tahmid Rahman Laskar,
Shayna Gardiner, Pooja Hiranandani, Shashi Bhushan TN

Dialpad Canada Inc.

Vancouver, BC, Canada

{xue-yong, cchen, tahmid.rahman}@dialpad.com

{sgardiner, phiranandani, sbhushan}@dialpad.com

Abstract

Entity-level sentiment analysis predicts the sentiment about entities mentioned in a given text. It is very useful in a business context to understand user emotions towards certain entities, such as products or companies. In this paper, we demonstrate how we developed an entity-level sentiment analysis system that analyzes English telephone conversation transcripts in contact centers to provide business insight. We present two approaches, one entirely based on the transformer-based DistilBERT model, and another that uses a convolutional neural network supplemented with some heuristic rules.

1 Introduction

Businesses that provide Contact Center as a Service (CCaaS) often leverage Artificial Intelligence (AI) technologies to transcribe telephone conversations and generate aggregated insights reports for contact centers across various industry verticals. Customers occasionally make evaluative comments about specific products or companies during a customer support call to a contact center. These comments provide valuable competitive insights to the business, e.g. positive comments may provide information useful to a marketing department as it formulates an advertising campaign while negative comments may provide valuable insights that can be used to improve a product or a service. In such scenarios, building a system that can identify user sentiments towards entities like product or companies could be useful.

Though Aspect-Based Sentiment Analysis (ABSA) (Zhou et al., 2019, 2020) that aims to extract the user sentiment expressed towards a specific aspect associated with a given target could be a possible solution to solve such problems, it should be noted that ABSA has a key limitation in this regard. For example, in tasks like competitor analysis where the objective is to understand the overall user sentiment within a certain period

of time on specific products or general company services, ABSA techniques could not be useful as they provide more fine-grained opinions towards predefined aspects (e.g. features of a product, ease of use of a software, or aspects of restaurant experience) instead of providing a generic user sentiment towards a specific entity (Zhou et al., 2019).

While building the Entity-level Sentiment Analysis (ELSA) system for real-world contact center use-cases, we observe several key challenges. First of all, to the best of our knowledge, there are no existing public datasets available for the entity-level sentiment analysis task. Meanwhile, this task becomes more challenging when the requirement is to use a dataset constructing from telephone conversations since constructing a dataset from speech transcripts generated from telephone conversations is non-trivial as telephone transcripts are generated by automatic speech recognition (ASR) systems that have their own unique characteristics. For instance, an ASR system may have mistranscription errors as well as linguistic disfluencies (e.g. filler words) (Fu et al., 2021) that usually occur in a conversational speech dataset (Malik et al., 2021).

The factors mentioned above make the implementation of an entity-level sentiment analysis model very challenging to detect user opinions towards entities that appear in contact center calls. In this paper, we address the existing limitations behind developing an entity-sentiment model for commercial scenarios in the domain of business telephone conversation data in contact centers. Since there is no suitable publicly available dataset for the entity-level sentiment analysis task, we briefly describe how we sampled and annotated the data for this task. We then propose two approaches that leverage neural models for this task (i) one is based on the DistilBERT (Sanh et al., 2019) model in which we modify its architecture such that the model can also be utilized to extract the opinion term(s) while detecting the sentiment po-

larity towards a named entity; (ii) while the other approach uses a convolutional neural network (dos Santos and Gatti, 2014) supplemented with some pre-defined heuristic rules. We compare the effectiveness of both approaches through extensive experiments and discuss our findings to provide valuable insights for future developments of ELSA models for real world commercial scenarios.

2 Related Work

Since the entity-level sentiment analysis task is closely related to aspect-level sentiment analysis, in this section, we first briefly review the aspect-level sentiment analysis task followed by the entity-level sentiment analysis task in order to clarify the distinction between these two tasks while discussing our rationale behind developing an entity-level sentiment analysis model for contact centers.

2.1 Aspect-Based Sentiment Analysis (ABSA)

ABSA aims to classify the sentiment polarity of aspects of certain objects. Many previous studies are focused on this research (Sun et al., 2019; Tang et al., 2016; He et al., 2018; Zhao et al., 2020; Zhou et al., 2019). A more fine-grained related task is aspect sentiment triplet extraction (ASTE) (Peng et al., 2020; Xu et al., 2020), which extracts a triplet – aspect term, opinion term and sentiment – from the input. Detection of aspects in both ABSA or ASTE often relies on implicit lexical or semantic signs, for instance, *the food is too spicy* suggests that this comment is about the *taste* aspect. This is different from the entity recognition task where the goal is to detect the named entities in a given utterance based on the overall context.

2.2 Entity-level Sentiment Analysis (ELSA)

ELSA aims to predict the sentiment of named entities in a given text input (Steinberger et al., 2011; Saif et al., 2014). These named entities are usually application dependent. One recent work on ELSA is the work of Luo and Mu (2022), where they studied entity sentiment in news documents. Another prominent work on ELSA is the work of Ding et al. (2018), where an entity-level sentiment analysis tool was proposed for Github issue comments. Contrary to the above studies that focused on typed text, our focus is on noisy textual data (i.e., speech transcripts). Moreover, our proposed models can infer both entity sentiment and corresponding opinion terms for a better analysis of user

sentiments towards products or companies in business telephone conversations in contact centers.

3 Task Description

Let us assume that we have an utterance $U = w_1, w_2, \dots, w_n$ containing n words. The goal of the ELSA task is to identify m opinion words $O_W = ow_1, ow_2, \dots, ow_m$, (where $m < n$), and classify the sentiment of the identified opinion words towards the target entity e in the given utterance. In Table 1, we show some examples of the ELSA task to detect user sentiments towards products and organization type entities. In the first two examples, the customer is directly expressing positive sentiment about the named entity. For instance, (i) they say “I love it” indicating “Google” in context, or (ii) they are “very impressed” with “MAC”. In the third and fourth examples, customers are expressing negative sentiment about a product or facet associated with the company, e.g., “He has a hard time finding a good yogurt from Walmart” is a comment about the quality of Walmart’s service, not a comment about yogurt. Similarly, in the fourth example, difficulty navigating the Instacart app is indirectly an indication of negative sentiment concerning Instacart.

4 Dataset Construction

As noted earlier, there is no publicly available dataset for the ELSA task. We therefore had to create and annotate our own dataset. The first major issue that we observed while constructing a dataset for ELSA is that the entity-level sentiment events in our telephone transcripts are very infrequent. Hence, random data sampling techniques might yield an imbalanced dataset where most utterances would not have any positive or negative sentiments towards an entity. We therefore used two pre-existing models — a named entity recognition (NER) model based on DistilBERT (Sanh et al., 2019) that was trained to identify **Organization** and **Product** type entities and a convolutional neural network (CNN) (Krizhevsky et al., 2012; dos Santos and Gatti, 2014; Albawi et al., 2017) sentiment analysis model — to sample 13000 utterances that contained at least one named entity and one positive or negative sentiment predicted by these models. To balance the dataset, we sampled an additional 10000 utterances containing at least one entity and having no polarized sentiments (i.e., only neutral sentiment). The resulting 23000

I work at Google and I love it a lot.
She’s very impressed how MAC works so well.
He has hard time finding a good yogurt from Walmart .
It’s quite difficult to navigate the mobile app of Instacart .

Table 1: Examples for entity-level sentiment analysis. Words in color *blue* are target named entities. Words in color *teal* are positive opinion words. Words in color *purple* are negative opinion words.

utterances were manually annotated by independent annotators to determine the positive, neutral, or negative sentiment toward the target entity. The annotators also identified the opinion terms in the utterances.

5 Our Proposed Models

For performance evaluation, we propose two approaches: (i) *DistilBERT-based Model*, and (ii) *CNN-based Model with Heuristics Rules*. Below, we present our proposed approaches.

5.1 DistilBERT-based Model

For this approach, we leverage the DistilBERT model since this is a very lightweight model that does not require much computing power in production environments (Sanh et al., 2019). Below, we describe how we utilize this model for ELSA.

NER tagging: Given an utterance as input, we first run an NER model to determine if there is at least one entity (product or organization) detected. Our NER model is based on DistilBERT that is trained over business conversation data collected from call centers. During the training stage, we use the cross entropy (CE) loss as defined in Equation 1:

$$L_{CE} = -\frac{1}{N} \sum_{n=1}^N \log \frac{e^{\hat{y}_{n,y_n}}}{\sum_{c=1}^C e^{\hat{y}_{n,c}}} \quad (1)$$

Here, N is the number of samples in a batch, and C denotes the number of classes, $\hat{y}_{n,c}$ is the logit of the c -th class in the n -th example, and \hat{y}_{n,y_n} is the logit of the gold class in the n -th example.

Context Representation: We insert a special tag, *_NE_*, before any named entities detected in an utterance. This helps the model to identify which spans belong to the entity. For example, if the raw input is “*I really don’t like using Snapchat*”, we reformulated the input as “*I really don’t like using _NE_ Snapchat*”. Then, we send the pre-processed input to our entity sentiment detection model that we describe below.

Entity Sentiment Detection: Our entity sentiment detection model is also based on DistilBERT. However, for this task, we train DistilBERT over business telephone conversation data for a different task: the sentiment classification task. Meanwhile, our entity sentiment detection model can also extract the opinion word(s) in a given utterance. This is done by adding an additional prediction layer on top of the DistilBERT model to identify the opinion words. During the training phase, the model is fine-tuned on our entity sentiment dataset to predict the polarity of the opinion terms for a given utterance. If the target entity’s sentiment is positive or negative, the model will assign respective tags (**POS** for positive and **NEG** for negative) to the opinion token(s), while the remaining tokens will be assigned to the **O** tag.

Transfer Learning: To improve model performance, we introduce a transfer learning technique for our entity sentiment detection model, for which we first fine-tune the DistilBERT model for the sentence classification task (i.e., sentiment analysis) on the Stanford Sentiment Treebank (SST) dataset that contains 11,855 training examples. The assumption is that if a model is fine-tuned on a similar task, it is expected to perform better on related downstream tasks (Laskar et al., 2022c; Garg et al., 2020). Although the SST dataset is about predicting the general sentiment of a given text sequence, it requires the model to learn what words are associated with positive sentiments and what are associated with negative sentiment, which is essential for our task. The DistilBERT model trained on the SST dataset is then fine-tuned again on the processed input (pre-processed by using the *_NE_* tag obtained from our NER model) in the contact center conversation dataset. As mentioned earlier, in this stage of fine-tuning, the entity sentiment model can also extract the opinion word(s) via utilizing the additional prediction layer that we added on top of DistilBERT.

An overview of our proposed DistilBERT-based approach is shown in Figure 1.

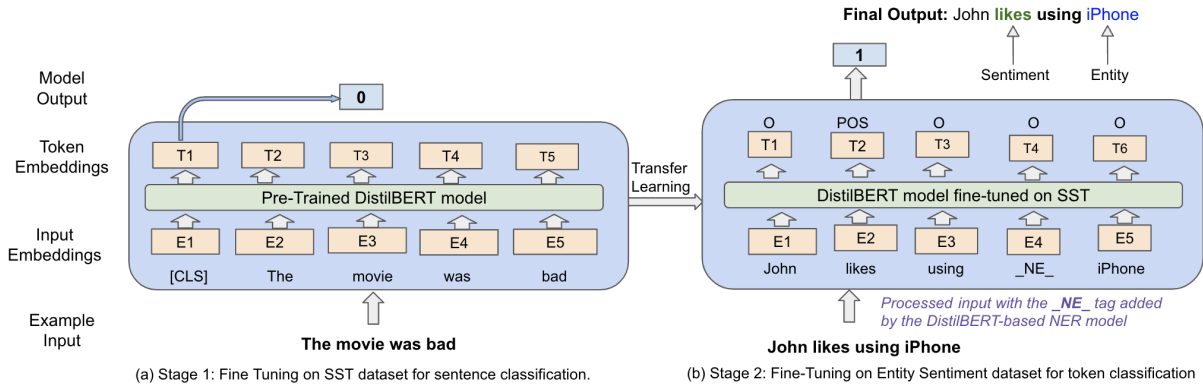


Figure 1: An overview of our proposed DistilBERT-based approach: (a) first, we do fine-tuning on the SST dataset for the generic sentiment analysis task, and (b) then, fine-tune on the in-domain Entity Sentiment dataset for entity level opinion extraction. Here, in the output layer, 1 denotes positive while 0 denotes negative sentiment.

5.2 CNN-based Sentiment Model Supplemented with Heuristic Rules

For this model, we employ a two-step approach. We first run a general sentiment analysis model that classifies the sentiment of a given utterance and also extracts the keywords that cause that sentiment (if the sentiment is positive or negative). We treat these sentiment keywords as opinion word candidates. Then we employ a set of linguistic heuristics that identify the opinion words that are associated with the entities mentioned in the input.

The sentiment analysis model is a multiclass, CNN-based classification model. We choose CNN here due to its effectiveness in related tasks (e.g., ABSA task) (Wang et al., 2021). However, for ELSA, we add an explainability layer on top of CNN that is tasked with explaining predictions. The explainability technique that we leverage is called Integrated Gradients, adopted from (Sundararajan et al., 2017). After a sentiment score is predicted by the model, the explainability layer emits words that are highly associated with the predicted sentiment. Note that we apply some heuristics to select the emitted words as candidates for opinion words.

Heuristics: For the extraction of opinion words, we utilize heuristics based on phrase structure types that are most likely to contain entity sentiment. We divide these phrase types into three categories to find which part of speech contains the potential opinion word: verb-based, adjective-based, and noun-based. Table 2 illustrates the possible syntactic patterns captured by these heuristics, all of which also allow for optional modifiers such as intensifiers (e.g. *really*, *very*, *so*), complementizers

(*that*, *which*) or stacked adjectives. Some example phrases that were captured include: *I’m so happy that Google made this*, *Android sucks*, *that was awesome of Netflix to do*, *Netflix is garbage*, *my hatred of LaTeX*, *classic LaTeX awesomeness*, etc.

6 Experiment

In this section, we present the training parameters, evaluation metrics, and the experimental results. In our experiments, we use the following three models:

- **DistilBERT:** This model does not leverage the SST dataset as the first stage of fine-tuning. Instead, it is fine-tuned only on our ELSA dataset.
- **DistilBERT + SST:** This model is initially fine-tuned on the SST dataset and then fine-tuned on our ELSA dataset.
- **CNN + Heuristics:** This is the model that leverages CNN and supplemented with some heuristics rules.

6.1 Training Parameters

For the DistilBERT model, we set the batch size to 32, learning rate to 5×10^{-5} , and employ early stopping with patience set to 5. The pretrained model is based on the HuggingFace Transformer (Wolf et al., 2020). While for the CNN model, we use 300 dimensional fastText embeddings (Bojanowski et al., 2016; Santos et al., 2017), global max pooling is utilized in the convolutional layer with filter sizes: 2, 3, 4, 5, 6. The fully connected layer is 128 dimensional.

Verb-Based	Adjective-Based	Noun-Based
sentiment verb + entity	sentiment adjective + entity	entity + sentiment noun
sentiment verb + + preposition + entity	sentiment adjective + <i>of/for</i> + entity	sentiment noun preposition + entity
entity + sentiment verb		entity + aux verb sentiment noun

Table 2: Heuristic Rules to Extract Opinion Words

Models	Precision		Recall		F1	
	Ent	OP	Ent	Op	Ent	Op
DistilBERT	74.43	59.99	73.77	69.44	73.70	64.35
DistilBERT + SST	74.83	68.21	74.69	63.02	74.72	65.48
CNN + Heuristics	77.48	97.65	58.23	16.78	50.07	28.64

Table 3: Experimental results for the Entity Level Sentiment (Ent) and Opinion Word Extraction (OP) tasks.

6.2 Evaluation Metrics

To evaluate entity sentiment classification and opinion word extraction, we define two kinds of evaluation metric. For polarity classification, we calculate precision, recall and F1 score for three sentiment categories: positive, negative and neutral. We then calculate the weighted value of these (support-based). For opinion word extraction, we evaluate it using the metrics that are usually used for named entity recognition (Li et al., 2020) and calculate precision, recall and F1 score for opinion words. The evaluation was done in a sample of 175 annotated utterances that were reviewed by another group of annotators.

6.3 Results

From Table 3, we find that in terms of the F1 metric, both variations of DistilBERT – (Vanilla **DistilBERT** and **DistilBERT + SST**) – outperform the **CNN + Heuristics** model by a huge margin. More specifically, **DistilBERT + SST** outperforms the **CNN + Heuristics** model by 15.75% of the F1 score. Comparing **DistilBERT** and **DistilBERT + SST**, we can see the effect of SST pre-training, which brings the F1 score up from 73.7% to 74.72%, with an increase of 1.38%. We also find that the **CNN + Heuristics** model obtains impressive precision score. This is because the heuristic rules used in the **CNN + Heuristics** model were developed to emphasize precision, but they do not handle linguistic variation well, resulting in poor Recall and F1 scores.

For opinion word extraction, which is noted as **OP**, the performance gap between the **DistilBERT** model and the **CNN + Heuristics** model is even larger. As shown in Table 3, **DistilBERT + SST**

outperforms the **CNN + Heuristics** by 38.48% F1 score. This is mainly because the **CNN + Heuristics** has very poor performance in recall: only 16%. Although the recall of **DistilBERT + SST** is lower than **DistilBERT**, its F1 score is still 1.07% higher than its counterpart.

Robustness Test: The overall metrics can’t identify if the performance of a model is robust in different situations. Thus, we investigate if our proposed model is robust against various kinds of input texts. For this purpose, we separate the test data into different sub-populations by the number of tokens and the number of entities. We then evaluate our models on sub-populations to see how they perform. Below, we define these sub-populations.

- (i) **1 entity:** *input text has only one target entity.*
- (ii) **> 1 entity:** *input text has more than one target entity.*
- (iii) **< 8 tokens:** *input text with less than eight tokens.*
- (iv) **> 45 tokens:** *input with more than forty five tokens.*

Table 4 contains the results of our proposed models in different data slices. We find that both models perform poorly when the input is long (> 46 tokens) compared to when the input is short (< 8 tokens). This could be because it is much harder to model long term dependencies when the sequence length is too long.

We also find that the **DistilBERT + SST** model is more sensitive to the number of target entities in the input compared to the **CNN + Heuristics** model. Its F1 score drops by 7.16% when the number of target entities increases from one to more than one.

Model	Precision	Recall	F1
CNN + Heuristics	97.65	16.78	28.64
CNN + Heuristics (< 8 tokens)	100	38.89	56
CNN + Heuristics (> 45 tokens)	100	8.16	14.99
CNN + Heuristics (= 1 entity)	97.54	16.86	28.75
CNN + Heuristics (> 1 entity)	100	15.56	26.72
DistilBERT + SST	68.21	63.02	65.48
DistilBERT + SST (< 8 tokens)	66.9	79.46	72.46
DistilBERT + SST (> 45 tokens)	62.18	26.32	36.92
DistilBERT + SST (= 1 entity)	68.28	63.67	65.87
DistilBERT + SST (> 1 entity)	68	52.94	58.71

Table 4: Robustness Report on the Opinion Word Extraction task.

7 Commercial Application

We have deployed the **DistilBERT + SST** model in our production system to generate entity sentiment data for contact centers as it has better accuracy than the **CNN + Heuristics** model. Due to the small model size and efficient inference of DistilBERT, each model instance is assigned 1 CPU and 1GB memory. Once there are enough entity-level sentiment predictions, there is a dedicated pipeline to aggregate entity sentiment in different granularity for each customer.

There are many use cases of the aggregated insights of entity sentiment. Contact center managers can use this information to improve contact center efficiency by investigating why customers are not happy with certain products (e.g., *itelephone 13 Pro Max*) and develop desired responses when the customer is complaining about it, so that the agents can handle the difficult situation more efficiently. The collected negative feedback can be used to inform the product team how to improve the products. The insights can also be used to conduct comparisons between several products or companies and help with competitor analysis.

8 Conclusion

In this paper, we described the creation of a task-specific dataset and a new model that extracts opinion words while performing entity sentiment polarity detection. The resulting DistilBERT-based model is currently deployed as a commercial application for entity-level sentiment analysis for English contact center conversations. In the future, we will investigate how to extend our proposed methods to other applications (Laskar et al., 2022a,b) of the entity recognition task (Fu et al., 2022) in telephone transcripts and explore how to improve model performance on utterances that contain more than one entity.

Limitations

As our entity sentiment models are trained on English business telephone conversations, they might not be suitable to be used in other domains, types of inputs (i.e written text), or languages. The NER component of **DistilBERT** based model has some limitations while detecting product and organization type entities. It is more biased towards detecting the entities that appear more frequently in the training data and misses rare entities. This could impact the overall performance of the model.

Ethics Statement

This data in this research is comprised of individual sentences that do not contain sensitive, personal, or identifying information. The entity sentiment model deployed in production is not used to attach any sentiment to people, only to non-human entities. Each machine-sampled utterance is labelled by annotators before the utterance is used as part of the training dataset. While annotator demographics are unknown and therefore may introduce potential bias in the labelled dataset, the annotators are required to pass a screening test before completing any labels used in these experiments, thereby mitigating this unknown to some extent. We paid adequate compensation to the annotators. Future work should nonetheless strive to improve training data further in this regard.

ACKNOWLEDGEMENTS

We would like to thank Simon Corston-Oliver for his helpful and detailed feedback on the paper. We also thank the reviewers for their excellent review comments that helped us improving the paper.

References

- Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. 2017. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching word vectors with subword information. arxiv 2016. *arXiv preprint arXiv:1607.04606*.
- Jin Ding, Hailong Sun, Xu Wang, and Xudong Liu. 2018. Entity-level sentiment analysis of issue comments. In *Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering, SEmotion@ICSE 2018, Gothenburg, Sweden, June 2, 2018*, pages 7–13. ACM.
- Cícero Nogueira dos Santos and Maira Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 69–78. ACL.
- Xue-Yong Fu, Cheng Chen, Md Tahmid Rahman Laskar, Shashi Bhushan, and Simon Corston-Oliver. 2021. Improving punctuation restoration for speech transcripts via external data. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, pages 168–174.
- Xue-Yong Fu, Cheng Chen, Md Tahmid Rahman Laskar, Shashi Bhushan Tn, and Simon Corston-Oliver. 2022. An effective, performant named entity recognition system for noisy business telephone conversation transcripts. In *Proceedings of the Eighth Workshop on Noisy User-generated Text (W-NUT 2022)*, pages 96–100.
- Siddhant Garg, Thuy Vu, and Alessandro Moschitti. 2020. Tanda: Transfer and adapt pre-trained transformer models for answer sentence selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7780–7788.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. Effective attention modeling for aspect-level sentiment classification. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 1121–1131. Association for Computational Linguistics.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, Nevada, United States*, pages 1106–1114.
- Md Tahmid Rahman Laskar, Cheng Chen, Jonathan Johnston, Xue-Yong Fu, Shashi Bhushan TN, and Simon Corston-Oliver. 2022a. An auto encoder-based dimensionality reduction technique for efficient entity linking in business phone conversations. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3363–3367.
- Md Tahmid Rahman Laskar, Cheng Chen, Aliaksandr Martsinovich, Jonathan Johnston, Xue-Yong Fu, Shashi Bhushan Tn, and Simon Corston-Oliver. 2022b. BLINK with Elasticsearch for efficient entity linking in business conversations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Track*, pages 344–352, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.
- Md Tahmid Rahman Laskar, Enamul Hoque, and Jimmy Xiangji Huang. 2022c. Domain adaptation with pre-trained transformers for query-focused abstractive text summarization. *Computational Linguistics*, 48(2):279–320.
- Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2020. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):50–70.
- Manman Luo and Xiangming Mu. 2022. Entity sentiment analysis in the news: A case study based on negative sentiment smoothing model (nssm). *International Journal of Information Management Data Insights*, 2(1):100060.
- Mishaim Malik, Muhammad Kamran Malik, Khawar Mehmood, and Imran Makhdoom. 2021. Automatic speech recognition: a survey. *Multimedia Tools and Applications*, 80(6):9411–9457.
- Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 8600–8607. AAAI Press.
- Hassan Saif, Yulan He, Miriam Fernandez, and Harith Alani. 2014. Semantic patterns for sentiment analysis of twitter. In *International Semantic Web Conference*, pages 324–340. Springer.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.
- Igor Santos, Nadia Nedjah, and Luiza de Macedo Mourelle. 2017. Sentiment analysis using convolutional neural network with fasttext embeddings. In *2017 IEEE Latin American conference on computational intelligence (LA-CCI)*, pages 1–5. IEEE.

- Josef Steinberger, Polina Lenkova, Mijail Kabadjov, Ralf Steinberger, and Erik Van der Goot. 2011. Multilingual entity-centered sentiment analysis evaluated by parallel corpora. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, pages 770–775.
- Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 380–385, Minneapolis, Minnesota. Association for Computational Linguistics.
- Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 3319–3328. PMLR.
- Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 214–224. The Association for Computational Linguistics.
- Xinyi Wang, Feng Li, Zequn Zhang, Guangluan Xu, Jingyuan Zhang, and Xian Sun. 2021. A unified position-aware convolutional neural network for aspect based sentiment analysis. *Neurocomputing*, 450:91–103.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2020 - Demos, Online, November 16-20, 2020*, pages 38–45. Association for Computational Linguistics.
- Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 2339–2349. Association for Computational Linguistics.
- Fei Zhao, Zhen Wu, and Xinyu Dai. 2020. Attention transfer network for aspect-level sentiment classification. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 811–821, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Jie Zhou, Jimmy Xiangji Huang, Qin Chen, Qinmin Vivian Hu, Tingting Wang, and Liang He. 2019. Deep learning for aspect-level sentiment classification: survey, vision, and challenges. *IEEE access*, 7:78454–78483.
- Jie Zhou, Jimmy Xiangji Huang, Qinmin Vivian Hu, and Liang He. 2020. SK-GCN: modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. *Knowl. Based Syst.*, 205:106292.