A Survey on Aspect Based Sentiment Analysis on Social Conversational Text

Guided Research

Department of Informatics
Technical University of Munich (TUM)

Supervisor PD Dr. Georg Groh
Research Group Social Computing

Advisor Gerhard Hagerer, M.Sc.
Research Group Social Computing

Author Sudeshna Dasgupta

Date Munich, October 3, 2019

Contents

1 Introduction

This report is a comprehensive review of the recent research efforts attempted for analyzing conversational social media corpora using aspect based sentiment analysis. Sentiment analysis is one of the most active research fields (Liu and Ren, 2019) in natural language processing (NLP) (Liu, 2012a), with social media as a major domain of application. Sentiment analysis at the aspect granularity, called aspect based sentiment analysis (ABSA), has emerged as a promising research direction. Social media platforms such as forum discussions and micro-blogs generate conversational text corpora principally from human interaction and collective behavior, hence garnering attention from disciplines like sociology, business, psychology, politics, economics, etc. The application of ABSA on social conversational text is therefore an encouraging attempt to understand social patterns, user attitude, and future behavioral trends.

This report is organized as follows. Section 2 presents a literature overview of the tasks in focus. Section 3 talks about related work. Section 4 discusses the focus of this report, Section 5 presents the open challenges in this domain, the relevant datasets, and reviews the relevant publications along with results. The discussed publications are categorized based on model architectures to obtain an overview of the prevailing trends in the field of our research topic and highlight the best performing architecture. Each publication has been analyzed under the framework: challenges, motivation, assumption and technique.

2 Literature Overview

2.1 Sentiment Analysis

Sentiment analysis is the computational study of opinions or sentiments expressed in text and detection of the sentiment polarity expressed in the text corpus as *positive, negative, neutral* (Liu, 2012b). The same can be performed at three granularities 1) document level 2) sentence level 3) target or aspect level.

2.2 Aspect Based Sentiment Analysis

The primary difference between target based sentiment analysis (TBSA) or aspect based sentiment analysis (ABSA), also termed targeted aspect based sentiment analysis (TABSA), and traditional sentiment analysis is that the former is aspect or target oriented. An opinion target is the linguistic expression used in the given text to refer to the reviewed entity or an aspect (part or attribute) of an entity. Hence, an opinion about an entity implies for the entity as a whole, while an opinion about an aspect only refers to that specific attribute of an entity (Shu et al., 2016). For example in the sentence "The salmon was good, but the manager was not polite", here *salmon* is an instance of the entity FOOD and *manager* is an instance of the entity SERVICE. Therefore, ABSA is a conglomeration of multiple sentiment analysis subtasks (Liu, 2012a) such as:

- Aspect extraction
- Opinion target extraction
- Entity extraction
- Aspect sentiment classification based on inferred aspect and its surrounding contexts.

The two major subtasks identified by all major research works are (1) identifying target words and phrases indicating aspect term termed as aspect extraction (2) identifying sentiment polarity based on inferred aspect and its surrounding contexts, termed as aspect sentiment classification.

Aspect extraction: *Aspect extraction* consists extraction of entity aspects on which opinions have been expressed.

Aspect sentiment classification: *Aspect sentiment classification* is the task of identifying fine-grained opinion polarity towards a specific aspect associated with a given target, assuming that a sentence might express different opinions towards different targeted entities (Ma et al., 2018b).

3 Related Surveys

Two relevant and comprehensive surveys published in the field of aspect based sentiment analysis are (Schouten and Frasincar, 2015) and (Zhou et al., 2019). (Schouten and Frasincar, 2015) review the state of the art in aspect-level analysis and its various sub-tasks till the year 2014. All the approaches use non-conversational social media data like product reviews collected from Amazon (Hu and Liu, 2004), hotel reviews from TripAdvisor, product reviews from Chinese Opinion Analysis Evaluation 2008 or comments retrieved from Chinese mobile telephone domain websites apart from the popular Semeval datasets. Additionally, two different datasets mentioned in the survey are blog entries returned from Google Blog Search[*] and phrases selected from dictated patient summaries at the Pediatric Environmental Health Clinic (PEHC). (Zhou et al., 2019) is an in-depth survey of deep learning-based aspect-level sentiment classification models. The data used constitute both conversational and non-conversational datasets, but only deep learning based approaches are considered.

[*] http://blogsearch.google.com

4 Research Overview

ABSA has garnered attention of research communities since SemEval 2014. In this paper we perform a survey of all the recent research efforts attempted to solve the problem of ABSA. We select the time interval of year 2016 to present (2019). Deep Learning being a flourishing field, has been applied to aspect based sentiment analysis majorly in recent years. Also, multiple surveys have been conducted in the field of aspect based sentiment analysis (Zhou et al., 2019, Schouten and Frasincar, 2015, Rana and Cheah, 2016). Our focus is **conversational text in social media**.

5 Synopsis of Recent Developments

5.1 Challenges

Apart from attempts to improve sentiment classification accuracy, a number of research questions in the domain of aspect based sentiment analysis have been addressed in the recent years. Some of the challenges undertaken are discussed below:

- Lack of awareness of the salience of each context word with respect to the given aspect in a sentence (Liu et al., 2018b).
- Need for improvement in multiple target extraction (Wang and Lu, 2018).
- One or more instances of the target are usually more tightly tied with sentiment than others, and hence need to be highlighted (Yang et al., 2019, Liu et al., 2018a, Chen et al., 2017, Ma et al., 2018b, Li et al., 2018a).
- Arbitrary number of aspect term-polarity pairs in sentences (Li et al., 2017).
- Lack of exploitation of the fact that aspect terms should co-occur with opinion indicating words.
- Unavailability of abundant aspect term(AT)-level corpora (Li et al., 2019b).
- While supervised sentiment analysis using domain adapted and sentence based analysis approaches are effective, gathering a substantial measure of sentiment from domain information is difficult (Kumar and Vardhan, 2019).
- High complexity and training cost of sequential models. Limited ability to capture long sentence information and non-negligible attention noises (Zhang et al., 2019a).
- Most existing methods assume that the target mentions are given, and only aim at aspect sentiment classification, which limits their practical use (Li et al., 2019a).

5.2 Datasets

The standard datasets for ABSA can be classified into two categories, conversational and non-conversational.

Conversational: Conversational-text datasets used majorly in the reviewed research publications are retrieved from a microblogging service Twitter[†]. The two listed datasets are a manually annotated dataset introduced by (Dong et al., 2014) and Mitchell[‡] dataset (Mitchell et al., 2013) which consists 10000 English tweets along with about 30000 Spanish tweets annotated by Amazon's Mechanical Turk. SentiHood (Saeidi et al., 2016) is a dataset retrieved from question answer forum Yahoo! Answers[§] in the domain of urban neighbourhoods. It is one of the largest knowledge exchange communities where users answer one another's questions.

Non-Conversational: Three popular datasets were released by the international workshop on semantic evaluation for the aspect-level sentiment analysis task SemEval 2014 , SemEval 2015 and SemEval 2016. Semeval14 consists of 7686 reviews from two domain-specific datasets Restaurant14 and Laptop14 respectively constituting restaurants and laptop reviews manually annotated with aspect term, aspect category and respective polarities. SemEval15 (Pontiki et al., 2015) consists of reviews for restaurants, laptop and hotels manually annotated with entity, attribute, polarity. SemEval16 (Pontiki et al., 2016) consists data from 7 domains and 8 languages.

5.3 Graph-Based Solution

(Li and Lu, 2017) formulate the targeted sentiment based analysis (TBSA) task in order that the subtasks target extraction, sentiment polarity detection, and detection of sentiment scope window size are performed in a joint manner. The publication introduces the utility of sentiment scope graphs where a sentiment scope is defined to be an unbounded and variable-size window of words surrounding each entity that determines its sentiment polarity. The window information is usually unavailable and learned from data. In this research, it's assumed that every target is associated with a sentiment scope but exactly one sentiment scope. Hence the sentiment scopes cover the complete sentence, and no overlapping is premised. Sequences are labelled with 9 types of nodes, B_+, E_+ and A_+ nodes for positive sentiment, B_-, E_- and A_- nodes for negative sentiment, as well as B_0, E_0 and A_0 nodes for neutral sentiment. The B nodes are used to denote that the current word is part of the sentiment scope of a certain sentiment polarity, and appears before the target or is exactly the first

[†]www.twitter.com
[‡]http://www.m-mitchell.com/code/index.html
[§]https://answers.yahoo.com/

word of the target expression.

The E nodes denote that the current word is part of a target expression. The A nodes denote that the current word belongs to a sentiment scope of a certain polarity and appears after the target or is the last word of the target expression. The nodes selected at different positions are connected to form a sentiment scope graph which therefore reveals the targets and their associated sentiment scopes with polarity information. As sentences have complex structures, there exists several possible sentiment scope graphs, each specifying a different possible entity and sentiment scope combination. Probability of predicting a possible output is as shown in the equation :

$$p(\mathbf{y}|\mathbf{x}) = \frac{\sum_{\mathbf{h}} \exp\left(\mathbf{w}^T \mathbf{f}(\mathbf{x}, \mathbf{y}, \mathbf{h})\right)}{\sum_{\mathbf{y}', \mathbf{h}'} \exp\left(\mathbf{w}^T \mathbf{f}\left(\mathbf{x}, \mathbf{y}', \mathbf{h}'\right)\right)} \tag{1}$$

where **w** is the weight vector, **f(x,y,h)** is the feature vector defined for sentence **x**, output structure **y**, and latent variable **h** which provides sentiment scope information for the **(x,y)** tuple. The model assigns high scores to those sentiment scope graphs with the correct named entity and sentiment scope information, and can assign lower scores to other incorrect sentiment scope graphs. The MPQA lexicon and SentiWordNet lexicon is used to obtain polarity for English words. The model obtains F1-Scores **56.83** and **40.11** for aspect extraction and sentiment analysis respectively.

5.4 Matrix Computation-Based Solutions

(Zhang et al., 2019a) propose a *co-attention* based network to attend the aspect term and its surrounding contexts simultaneously, and capture the correlation between aspect and contexts. Each input sentence is separated into three components based on its inferred aspect: left context, aspect term and right context. The co-attention network generates two feature vectors for each input sub-sentence, based on individual word embeddings, which may carry redundant information. As it is unnecessary to input all of them into the final prediction layer, a *Low-rank Bilinear Pooling* (LRBP) method based on Hadamard product, is used to reduce the dimension of the final input vector without losing discriminative power. As the proposed model is based on matrix computations and is not using any sequential components such as long short term memory (LSTM) or gated recurrent unit (GRU) to extract features, parallelized training can be performed to improve training efficiency.

(Kumar and Vardhan, 2019) address the challenge of gathering a substantial measure of sentiment from domain information by an unsupervised learning based sentiment analysis approach. Given corpus is split among different topics using *Probabilistic Latent Semantic Analysis* (pLSA), and the topics are utilized for tagging sentiment with the help of *Independent Component Analysis* (ICA). A word weight frequency-based approach is applied to create a bag of sentiment words, such as to form a vocabulary for a particular domain, which is used for sieving all the sentiment words from the corpus. ICA is applied to extract the sentiment words from the sieve, which hence are the ones with domain aspect sentiment. Polarities of these words are suggested to be more accurate than the lexicon-based sentiment analysis. Thus it allows the corpus to gain real polarity in the domain of topics.

5.5 Convolutional Neural Network-Based Solutions

(Chen and Qian, 2019) introduce the *TransCap* model to conduct aspect-level sentiment classification with the auxiliary knowledge transferred from document-level data. TransCap is motivated by the capsule network ((Hinton et al., 2011), (Sabour et al., 2017)), which uses *capsule vectors* and the *dynamic routing* approach to store and cluster features. The model consists *aspect routing* approach to compute the aspect weight for the context words of K-size window in learning task T_A, and the *dynamic routing* approach by adapting to the transfer learning framework. The *FeatCap* layer performs multiple convolution operations using various kernel groups, on N-gram extracted from sentence embedding, to produce feature capsules. One kernel group extracts one category of semantic meaning. The *SemanCap* layer accommodates the aspect routing mechanism which computes the aspect weight for the context words in a fixed-size window by a fusing convolution operation on the sentence embedding. The weight controls the information flow to the following layer, such that if aspect routing weight is zero, the feature capsule is blocked. All feature capsules of the same channel are aggregated to obtain global semantic representations. The *Class Capsule* layer performs the *dynamic routing* approach which hosts a logit to decide if a semantic capsule should pass to a class capsule.

(Zhang et al., 2019b) propose a *multi-layer attention based CNN* (Mul-AT-CNN) model which exploits convolutional neural network's (CNN) ability to receive data in parallel. The model uses a small receptive field to capture semantic representation of word and phrases. A deeper convolution layer, and hence a larger receptive field captures the semantic relationship of distant words. An attention mechanism uses semantic and location information to judge important features related to the target. The weights of each layer's output vectors are determined and a continuous vector is obtained from the sum of the vectors. Experimental results show that four or five layers are the best performing models.

5.6 Recurrent Neural Network-Based Solutions

(Luo et al., 2019) solve the two subtasks of ABSA, aspect term extraction (ATE) and aspect sentiment classification (ASC), jointly termed as aspect term-polarity co-extraction using a *Dual Cross-Shared RNN framework* (DOER). The tasks are individually formulated as sequence labeling problems to generate all aspect term polarity pairs of the input sentence simultaneously. DOER consists of stacked dual recurrent neural network (RNN) and a cross-shared unit, one RNN being for ATE and other for ASC. Each RNN layer consists of bidirectional Residual Gated Unit (ReGU). A ReGU involves a gate which mimics skip connection, that is transferring the input to the output, and thus is capable of training deeper and obtaining more useful features. Double embeddings are used as the initial word embeddings. They contain two types: *general-purpose* embeddings and *domain-specific* embeddings, distinguished by whether the embeddings are trained by an in-domain corpus or not. The labels of ATE and the labels of ASC are strongly related and have the information to imply the boundary of each aspect term. The cross-shared unit (CSU) functions upon the interaction of ATE and ASC. Two auxiliary tasks are also defined to enhance the representation of sentiment and alleviate the difficulty of long aspect terms. *Auxiliary Aspect Term Length Enhancement* task alleviates the concern of discontinuous labels, resulted from the length of the aspect term, by training the model to predict the average length of aspect terms in each sentence.

(Li et al., 2019a) proposes a unified framework to solve the task of target detection and sentiment prediction in an end-to-end fashion. The whole TBSA task is formulated as a sequence labeling problem which uses a unified tagging scheme γ^S ={B-POS, I-POS, E-POS, S-POS, B-NEG, I-NEG, E-NEG,S-NEG, B-NEU, I-NEU, E-NEU, S-NEU} $\cup$ {O} where a tag contains data about the boundary of target mention and the target sentiment. The model consists of two stacked RNNs with LSTM cells. One RNN layer predicts boundary tags of the target words and passes the hidden representations to second RNN layer which predicts unified tags as output. Three key components *Boundary Guidance, Sentiment Consistency* and *Opinion Enhanced Target Word Detection* perform auxiliary tasks. The Boundary Guidance component receives boundary information and updates a transition matrix to determine the final tagging decision. Initially, a non-zero element denotes the probability of the unified tags given the boundary tag and a zero element suggests that the unified tag does not correspond to the boundary tag. The final scores are obtained by combining the boundary-based and model-based unified tagging scores. As the task is formulated as a sequence tagging problem, the same word may be assigned different sentiment within a target sequence. The *Sentiment Consistency* module gated mechanism combines hidden state at previous timestep with current predictions. Opinion Enhanced Target Word Detection distills the boundary representations using a Softmax which functions as a token-level classifier that distinguishes target words from non-target words. Based on the assumption that opinion targets are always collocated with opinion words, a target word is one which should have opinion words with context window of fixed size. The model is the current state of the art for Targeted Sentiment Analysis and used as a baseline in (Hu et al., 2019).

(Zhang et al., 2016) is an extension over a baseline model (Zhang et al., 2015). The baseline model extracts features via pooling over left context, right context and target segment of the enclosed sentence for sentiment classification. Pooling functions can select useful features from a sequence of words, but fail to capture underlying tweet-level syntactic and semantic information, which may depict predicate-argument links, negations, coreferences and even sarcasm. Also, the interaction between the target and its contexts is not explicit. Two extensions are made to the baseline system to overcome the drawbacks. First, the addition of a recurrent hidden layer over the input layer. Rather than taking information from the current word alone, each node is also connected with its predecessor. A counterpart of the RNN is added in the reverse direction, thereby collecting information from left and right context of the target word. Pooling functions are performed over the vector concatenation of the hidden nodes. A second extension is the addition of gates to the recurrent hidden layers, which control information flow between nodes in the hidden and input layers. Gated neural networks reduce informal bias of vanilla recurrent neural networks towards the ends of a sequence by better propagation of gradients.

5.6.1 Gated Recurrent Unit-Based Solutions

(Liu et al., 2018b) propose a *content attention based aspect based sentiment classification model* called *Cabasc* which consists a *Content Attention Mechanism* (CAM) and a *Sentence-level Content Attention Mechanism* (SAM). CAM models the word order information, the aspect information and the correlation between the word and the aspect for attention weight computation. SAM captures important information in complex sentence structures from a global perspective. The contextual input representations are obtained from two GRU neural networks, and two multilayer perceptrons calculate the attention score over the hidden states. The attention weight corresponding to right and left context, and attention corresponding to aspect are concatenated to form the context attention weight vector. The vector is applied with the memory slices for words in the input sentence to obtain weighted memory. An average over the weighted memory vector functions as a sentence representation.

(Lei et al., 2019) propose a *Human-like Semantic Cognition Network* (HSCN) for aspect-level sentiment classification, inspired by the principles of the cognitive process of pre-reading, active reading and post-reading witnessed in human beings. Pre-reading is modeled by a *Word-level Interactive Perception* module which captures the correlation between context words and the given target words. A correlation matrix represents the relatedness between the context and the target. Mean pooling operation is performed on the values of the matrix to produce the context attention and target attention vectors. A *Target-aware Semantic Distillation* module consisting of *Skip-reading* module and *Semantic Composition* module mimics active reading. The first module which simulates skimming receives the context representation matrix and target representation matrix as inputs, and uses a semantic decision making scheme to determine whether context word should be deleted or retained. The semantic composition module receives the distilled context representation matrix and generates the target-specific context representation. A *Semantic Feedback* module evaluates the model's ability to comprehend the target-specific context semantics, by measuring the semantic deviation between the target-specific context representation and the given target. The obtained semantic residual vector is then used to fine-tune the aforesaid skip-reading step in a feedback regulation way.

(Ma et al., 2018a) propose a *Hierarchical Multi-layer Bidirectional Gated Recurrent Units* (HMBiGRU) network which automatically learns character-level features (e.g. capitalization, noun suffix, etc) on letter sequences, and models long-distance dependencies between words. The learned character features are capable of addressing out-of-vocabulary word problems. The character features extracted by the multilayer BiGRU, from character embeddings of each word, are concatenated with word embedding matrix which is utilized by another MBiGRU architecture to generate sentence representations. The target label information is introduced into predicting sentiment label, which helps the model learn to keep boundary information (B, I) of target label and sentiment label consistent. A transition matrix models the dependencies between labels, which is the probability of expecting label j after label i.

5.6.2 Long Short Term Memory-Based Solutions

(Hu et al., 2019) present a span-based extract then classify framework to avoid huge search space and sentiment inconsistency, two distinguished problems of sequence labeling. *Bidirectional Encoder Representations from Transformers* (BERT), is the backbone network of the model and is used to project input embeddings into contextualized token representations. An input sequence consists a $[CLS]$ token, the tokenized sentence, and a $[SEP]$ token. Each token is converted into vector space by summing the token, segment, and position embeddings. A series of *L* stacked Transformer blocks projects the input embeddings into a sequence of contextual vectors. Taking cue from extractive question answering, a multitarget extractor predicts multiple candidate targets by predicting their start and end positions in the sentence, as depicted in equation (2)

$$\begin{aligned} g^s &= w_s h^L \\ p^s &= \text{softmax}(g^s) \\ g^e &= w_e h^L \\ p^e &= \text{softmax}(g^e) \end{aligned} \tag{2}$$

where $w_s, w_e \in \mathbb{R}^h$ is a trainable weight vector, g^s and g^e are the unnormalized scores, and p^s and p^e are the probability distribution of the start and end positions. To adapt to multi-target scenarios, a *heuristic multi-span decoding* algorithm is proposed. The algorithm selects top-M indices from the two predicted scores g^s and g^e, and from multiple target candidate spans the one possessing the maximum heuristic regularized score is added to the set of outputs. Redundant spans are pruned using the Non-maximum Suppression Algorithm (Rosenfeld and Thurston, 1971). Overlapping spans are deleted. This process is repeated for all selected spans until the storage list is empty or top-K target spans have been proposed. A summarized vector is calculated using the attention mechanism over the tokens in a target span. The polarity probability is calculated for each candidate target span in the output set O, and the sentiment class that possesses the maximum probability is chosen. The framework has been studied using three approaches, *pipeline, joint and collapsed*, the pipeline model has the highest F1 score for Twitter dataset.

Previous research on targeted aspect based sentiment analysis (TABSA) extensively depended on feature engineering. Pre-trained language models such as ELMO, OpenAI and BERT have proven to be efficiently supplementing the efforts of feature engineering (Sun et al., 2019). **(Sun et al., 2019)** attempts to achieve state of the art results on the use of BERT model for TABSA. The task is formulated as sentence pair classification by constructing an auxiliary sentence followed by fine-tuning with pre-trained BERT. Four methods are used to convert the TABSA task into a sentence pair classification task. For a given token, its input representation is constructed by summing the corresponding token, segment, and position embeddings. For classification tasks, the first word of each sequence is a unique classification embedding ([CLS])

(Yang et al., 2019) set forth the disadvantage of using average pooling on multi-word target, and hence the plausible introduction of noise while obtaining target representation for sentiment classification. They recommend two models with *coattention* mechanism based on LSTM network and memory network, respectively named *Coattention-LSTM* and *Coattention-MemNet*. A co-attention mechanism alternates between target level attention and context level attention such as the target is attended based on the context summary vector, and the context is attended based on the attended target features. Both the attention processes include location prior information to assign more importance to context words proximal to the target, and to let the model depend on the adjoining context words to compute key features of the target. In the Coattention-MemNet model, the location weights are added into the target-level attention of 1-hop layer and the context-level attention of last-hop layer. This avoids the location weights to limit the iterative learning process of attention weights.

(Tang et al., 2016a) introduce two extensions to long short term memory (LSTM), the *Target-Dependent Long Short-Term Memory* (TD-LSTM) and its extension *Target-Connection Long Short-Term Memory* (TC-LSTM). Neural network models are observed to learn continuous features (representations) without tedious feature engineering and meanwhile capture the intricate relatedness between target and context words. The TD-LSTM uses two LSTM neural networks, a left one $LSTM_L$ and a right one $LSTM_R$, to model the preceding and following contexts respectively. The last hidden vectors are concatenated and fed to a Softmax layer to classify the sentiment polarity label. The TC-LSTM model extends TD-LSTM by incorporating a target connection component, which utilizes explicit connections between target word and each context word when composing the representation of a sentence. The prime difference between the models is that in TC-LSTM the input at each position is the concatenation of word embedding and target vector v_{target}, while TD-LSTM only includes the embedding of current word as input. The TD-LSTM model has been used as a baseline in (Liu et al., 2018b),(Ma et al., 2018b), (Li et al., 2018a).

(Li et al., 2019b) propose a *Multi-Granularity Alignment Network (MGAN)* to reduce the aspect granularity gap between tasks. Attention-based RNN models can achieve good performance usually only when large corpora are available. However, aspect-term (AT) level datasets are costly to obtain, as they require the aspect terms to be manually labeled or extracted by sequence labeling algorithms from the sentences. Aspect-category (AC) level corpora are more easily accessible as aspect categories are usually in a small predefined set. Hence a problem named *coarse-to-fine task transfer* is defined across domain and granularity, to borrow knowledge from an abundant source domain of the coarse-grained AC task, to a small-scale target domain of the fine-grained AT task. The model consists a source network, which consists of three-level attention hops *Context2Aspect (C2A)+Coarse2Fine (C2F)+Position-aware Sentiment (PaS)* for the AC task and acts as a teacher, while the target network is like a student, and uses two attention hops *C2A* and *PaS* for the AT task. The C2A attention measures the importance of each aspect word with regards to each context word. The C2F mechanism is designed to achieve task alignment between source and target tasks. Source aspect representation attends the target context, and the induced attention highlights the highly correlated aspect terms. Attention computation as shown in equation (3), (4)

$$z_i^f = (\mathbf{u}_f)^T \tanh\left(\mathbf{W}_f\left[\mathbf{h}_i; \mathbf{h}_s^a\right] + \mathbf{b}_f\right) \tag{3}$$

$$\beta_i^f = \frac{\exp\left(z_i^f\right)}{\sum_{i'=1}^{n} \exp\left(z_{i'}^f\right)} \tag{4}$$

where $\mathbf{W}_f \in \mathbb{R}^{d_u \times (2d_h + de)}, \mathbf{b}_f \in \mathbb{R}^{d_u}$ and $\mathbf{u}_f \in R^{d_u}$ are learnable weights, $\beta_i^f \in \mathbb{R}^n$ indicates the probability of each context word being an aspect term, h_i and $h_s{}^a$ are contextualized word representation of input and aspect representation of source domain respectively. PaS component uses the C2F attention weights to calculate position relevance of a context word from a location matrix, which represents proximity of each word in the sentence. A proximal context word with a large value β_i^f has a high position relevancy. The position attended aspect representation is passed to a $Softmax$ layer for sentiment classification. A *Contrastive Feature Alignment* unit consists of the modules *Semantic Alignment* (SA) and *Semantic Separation* (SS). The SA ensures identical distributions of feature representations when the domain is different, but the class is same. The SS ensures different distributions when both domain and class is different, hence mitigating false alignment. The model is used as a baseline in (Hu et al., 2019).

(Wang and Lu, 2018) propose an extension of the standard attention mechanism, termed as *segmentation attention*. The opinion expression associated with a target may be in the form of a chunk or can consist of multiple spans of texts. Standard attention mechanism does not model such structural information that exists in sentences, which however can be extremely crucial when the sentences contain multiple targets. The model concatenates a binary feature embedding to the word embedding to indicate whether each word is part of a target or not. Thereby the positional information of the target is encoded by the indicator sequence. A linear chain CRF specifies the structural dependencies between the latent variables. Based on its hidden states, the segmentation attention layer selects the opinion words of interest. The *segmentation attention* mechanism performs soft selections of opinion expressions, and incorporates a layer that is analogous to conditional random field (CRF) in the attention modeling process to capture the dependencies between adjacent words in the process. The opinion vector is derived from the representation of the latent opinions as denoted in equation (5):

$$\mathbf{z} = [z_1, \ldots, z_n]$$
$$\mathbf{m} = \sum_{\mathbf{z}} p(\mathbf{z}) g(\mathbf{R}, \mathbf{z}) \quad (5)$$

where $g(\mathbf{R}, \mathbf{z})$ is a feature function defined based on the selection of opinions, and $\mathbf{m}$ is the opinion vector. Compared with standard attention, segmentation attention has a high recall since the model captures a sequence of words rather than individual word thereby revealing more sentiment information to make the right prediction. But this depreciates the precision and two regularizers are introduces to balance them. Experiments reveal that as Twitter text is comparatively noisy and less structured, regularizers tend to lead the model to include wrong opinions.

(Yang et al., 2017) presents two *Attention-based Bidirectional LSTM (ABLSTM)* architectures for target-dependent sentiment classification. Unlike the models which only use the last hidden state (or mean pooling), the attention based model is capable of modelling long input sentences, when the target string is far from the most distinguishing features. Words are assigned attention scores according to their intent importance. ***ABLSTM1***: Weight for every word in the sequence is computed as the dot product of the hidden state of current input obtained by bidirectional LSTM (BiLSTM) at a time step and the hidden state of the target string. The scores are normalized across the context to obtain a proper probability distribution. ***ABLSTM2***:

$$\mathbf{o} = \sum_{t=1}^{T} a_t h_t, \text{ with } a_t = \operatorname{softmax}\left(h_t^T W_b h_{target}\right) \quad (6)$$

where $W_b \in R^{h \times h}$ is used in a bilinear term, used to compute the attention scores.

(Ma et al., 2018b) tries to overcome certain shortcoming in approaches where all instances of a given target are dealt with equal importance, and an average vector is computed over such instances. But, a given target might consist of multiple instances or multiple words in a sentence and one or more instances of the target might be more tightly tied with sentiment than others. The proposed model simultaneously learns a target-specific instance attention as well as a global attention. The sequence encoder, which is based on a BiLSTM, transforms the word embeddings into a sequence of hidden outputs. The hidden outputs corresponding to the location of consecutive or non-consecutive sequence of target words along with the target attention vector, termed as self attention vector, is used to compute the vector representation of the target. Similarly a sentence-level attention vector encodes the importance of each word in a sentence with respect to the target and aspect. The aspect-based sentence vector is then fed into the corresponding multi-class classifier to resolve the sentiment polarity. To include affective or commonsense knowledge from *SenticNet* (Cambria et al., 2016) knowledge base, a *concept vector* is added as extension to the forget, input and output gate of standard LSTM, now termed Sentic LSTM. The extension is made under the assumption that commonsense concepts contain information complementary to the textual word sequence, but are usually taken for granted and hence, are absent from the text.

5.7 Hybrid Solutions

In the Twitter dataset, sentiment words are usually far from person names. So a probable challenge is the word by word propagation of a sentiment feature far from the target, as loss of feature is highly probable. **(Chen et al., 2017)** introduce a multiple-attention mechanism *Recurrent Attention on Memory* (**RAM**) to incorporate important features in difficult sentence structures. A memory module is introduced between the attention module and the input module, to enable the framework to synthesize features of word sequences such as sentiment phrases. A normalized attention distribution attends information from memory, and the results of the attentions are combined in a nonlinear way. The architecture consists of five modules: *input module, memory module, position-weighted memory module, recurrent attention module, and output module.* The input module attains the word vector for an input sequence from an embedding lookup table, and gets a list of vectors. The states of time steps generated by BiLSTM from the input vectors are termed as memories, which are observed to be same for multiple targets in one comment, which is not ideal for predicting their respective sentiments. A position weighted strategy defined by the number of words between the word and the target is therefore used to assign weight to a memory slide. In the recurrent attention module, GRU architecture is used to update the episode e after each attention. The state of episode e_t is the interpolation of e_{t-1} and the candidate hidden vector $\tilde{e}_t$.

(Li et al., 2018a) propose target-specific *Transformation Networks* (TNet), which is the current state of the art for aspect sentiment classification. Instead of attention, a CNN layer extracts salient features from the transformed word representations received from a bi-directional RNN layer. The core of the model consists *L Context-Preserving Transformation* (CPT) layers, where each CPT module encapsulates a *Target-Specific Transformation* (TST) component. The TST component consists of a BiLSTM layer to generate the target word representations which are then dynamically associated with each word in the sentence as denoted in the equation (7) :

$$r_i^{\tau} = \sum_{j=1}^{m} h_j^{\tau} * \mathcal{F}\left(h_i^{(l)}, h_j^{\tau}\right) \tag{7}$$

where function $\mathcal{F}$ measures the relatedness between the j-th target word representation h_j and i-th sentence word representation h_i. The tailor-made target representation and the word representation is fed into a fully-connected layer to obtain a target-specific word representation. The non-linear activation function applied in the TST results in the loss of the context information in the contextualized representations received initially from the BiLSTM layer, as the mean and the variance of the features within the feature vector are changed. To utilize the context information, two strategies termed as *Lossless Forwarding* (LF) and *Adaptive Scaling* (AS) are proposed. In LF, the context representation is directly fed before the transformation to the next layer, as depicted in equation (8):

$$h_i^{(l+1)} = h_i^{(l)} + \tilde{h}_i^{(l)} \tag{8}$$

where $i \in [1, n]$ and $l \in [0, L]$ for an input sentence of n words and L number of CPT layers, $h_i^{(l)}$ and $\tilde{h}_i^{(l)}$ being the input of the l-th layer and TST output of the l-th layer respectively.

The second mechanism AS, is proposed to control the proportions of the input and transformed representations by applying a gating function and then performing convex combinations of the representations based on the gate. This thereby is an attempt to dynamically adjust the weights during LF. As CNN is capable of capturing multiple active local features holding different sentiments, it tends to associate it with the same target thus hindering prediction. So a position relevance between the word and the target is computed and added to the word representation and CPT module before the weighted representations are finally fed to a convolutional layer to generate the feature map. The model is used as a baseline in (Hu et al., 2019).

The attention mechanism tends to focus on frequent words with prominent sentiment polarities, while ignoring infrequent ones. However, implementing supervised attention is labor-intense work, as it is supposed to be manually annotated. **(Li et al., 2018b)** propose a *progressive attention supervision* mechanism as an extension to the aforementioned TNet model (Li et al., 2018a) and Memory Network model (Tang et al., 2016b). For each training instance, two word sets are initialized to record context words with active effects on the sentiment prediction, and context words with misleading effects. The most important context word is shielded with a special token "*<mask>*", thus creating a new sentence, before re-investigating effects of remaining context words of the training instance. The collected most impactful context word is dealt with based on two strategies, if the prediction is correct the model continues focusing on it, and add it to the list; otherwise the attention weight of the word is decreased and it's listed as a misleading context word. Hence, the maximum attention weight is exploited as attention supervision information to refine the model training.

6 Results

Table 1 , table 2 and table 3 list the results of the discussed approaches.

7 Future Research Directions

In the attempt to model the human reading cognition process, investigating the role of commonsense knowledge in modeling inter-target relations, and sentiment analysis over multiple targets co-occurring in especially long texts is a discernible future research direction. Other open research directions may include application of memory weighting strategies to distinguish multiple targets in a sentence, learning aspect categories across domains and transferring to a AT-level task, unsupervised detection of aspect terms and latent expression spans, modelling complex relationship between words such as negations modifiers, and identification of unknown sentiment words and phrases.

8 Conclusion

This survey intends to provide an extensive review of the most notable advancements on conversational text in the time frame 2016 to 2019. From the overview of the published approaches, it is apparent that deep learning architectures augmented with attention and memory network have demonstrated to be established techniques in the field. Furthermore, the acceptance of 4 publications on Twitter dataset based ABSA, in top tier conferences like AAAI, ACL and NAACL in the year 2019, depicts the growing interest in non-structured social media data. The discussed research publications are clustered for organization and highlighting existing and future research directions. Each publication has been analyzed under the framework: Challenges, Motivation, Assumption and Technique. Finally, the report discusses some of the open problems and promising future research directions.

Architecture	Model	Accuracy	Macro-F1	F1-Score
Graph-Based model	*SS*	-	-	56.83
RNN-Based models	*DOER*	-	-	71.35
GRU-Based models	*Joint Learning*	-	-	56.98
LSTM-Based models	*Sentic LSTM + TA + SA*	67.43	78.18	
	SPAN-pipeline	-	-	75.28
	BERT-pair-NLI-B	**79.8**	-	**87.5**

Table 1: Evaluation of results for aspect extraction. Best score in each column is marked in **bold**.

Architecture	Model	Accuracy	Macro-F1	F1-Score
Graph-Based model	*SS*	-	-	40.11
GRU-Based models	*Joint Learning*	-	-	**42.87**
LSTM-Based models	*Sentic LSTM + TA + SA*	89.32	-	
	SPAN-pipeline	75.16	-	-
	BERT-pair-NLI-B	**92.6**	-	-

Table 2: Evaluation of results for sentiment classification. Best score in each column is marked in **bold**.

Architecture	Model	Accuracy	Macro-F1	F1-Score
Matrix-Based Model	*pLSA*	**0.83**	0.4252	-
	Co-attention+LRBP	0.717	-	-
CNN-Based models	*Mul-AT-CNN*	0.713	-	-
	TransCap{T,T}†	0.739	-	0.701
RNN-Based models	*DOER*	-	-	0.514
	UNIFIED	-	-	0.48
	GRNN+G3	0.69	0.656	-
GRU-Based models	*Cabasc*	.715	-	-
	HSCN	0.69	0.661	-
LSTM-Based models	*Coattention-LSTM*	0.715	-	-
	Coattention-MemNet	0.705	-	-
	TD-LSTM	0.708	0.690	-
	TC-LSTM	0.715	0.695	-
	MGAN	0.746	0.735	-
	SA-LSTM	0.69	-	-
	AB-LSTM1	0.72	-	0.712
	AB-LSTM2	0.73	-	**0.722**
	SPAN-pipeline	-	-	0.577
Hybrid models	*RAM*	0.739	0.739	-
	TNet-LF	0.747	0.734	-
	TNet-AS	0.749	0.736	-
	TNet-ATT(+AS)	0.786	**0.777**	-

Table 3: Evaluation of results for aspect based sentiment analysis. Best score in each column is marked in **bold**. † Source document corpus is Twitter. Target aspect corpus is SemEval2014 Task 4 Laptop dataset

References

E. Cambria, S. Poria, R. Bajpai, and B. Schuller. Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives. In Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers, pages 2666–2677, 2016.

P. Chen, Z. Sun, L. Bing, and W. Yang. Recurrent attention network on memory for aspect sentiment analysis. In Proceedings of the 2017 conference on empirical methods in natural language processing, pages 452–461, 2017.

Z. Chen and T. Qian. Transfer capsule network for aspect level sentiment classification. In Proceedings of the 57th Conference of the Association for Computational Linguistics, pages 547–556, 2019.

L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu. Adaptive recursive neural network for target-dependent twitter sentiment classification. In Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers), pages 49–54, 2014.

G. E. Hinton, A. Krizhevsky, and S. D. Wang. Transforming auto-encoders. In International Conference on Artificial Neural Networks, pages 44–51. Springer, 2011.

M. Hu and B. Liu. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168–177. ACM, 2004.

M. Hu, Y. Peng, Z. Huang, D. Li, and Y. Lv. Open-domain targeted sentiment analysis via span-based extraction and classification. arXiv preprint arXiv:1906.03820, 2019.

P. Kumar and M. Vardhan. Aspect-based sentiment analysis of tweets using independent component analysis (ica) and probabilistic latent semantic analysis (plsa). In Advances in Data and Information Sciences, pages 3–13. Springer, 2019.

Z. Lei, Y. Yang, M. Yang, W. Zhao, J. Guo, and Y. Liu. A human-like semantic cognition network for aspect-level sentiment classification. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 6650–6657, 2019.

C. Li, X. Guo, and Q. Mei. Deep memory networks for attitude identification. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, pages 671–680. ACM, 2017.

H. Li and W. Lu. Learning latent sentiment scopes for entity-level sentiment analysis. In Thirty-First AAAI Conference on Artificial Intelligence, 2017.

X. Li, L. Bing, W. Lam, and B. Shi. Transformation networks for target-oriented sentiment classification. arXiv preprint arXiv:1805.01086, 2018a.

X. Li, L. Bing, P. Li, W. Lam, and Z. Yang. Aspect term extraction with history attention and selective transformation. arXiv preprint arXiv:1805.00760, 2018b.

X. Li, L. Bing, P. Li, and W. Lam. A unified model for opinion target extraction and target sentiment prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 6714–6721, 2019a.

Z. Li, Y. Wei, Y. Zhang, X. Zhang, and X. Li. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 4253–4260, 2019b.

B. Liu. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012a. ISBN 1608458849, 9781608458844.

B. Liu. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1):1–167, 2012b.

F. Liu, T. Cohn, and T. Baldwin. Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis. arXiv preprint arXiv:1804.11019, 2018a.

N. Liu and F. Ren. Emotion classification using a cnn_lstm-based model for smooth emotional synchronization of the humanoid robot ren-xin. PloS one, 14(5):e0215216, 2019.

Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu. Content attention model for aspect based sentiment analysis. In Proceedings of the 2018 World Wide Web Conference, pages 1023–1032. International World Wide Web Conferences Steering Committee, 2018b.

H. Luo, T. Li, B. Liu, and J. Zhang. Doer: Dual cross-shared rnn for aspect term-polarity co-extraction. arXiv preprint arXiv:1906.01794, 2019.

D. Ma, S. Li, and H. Wang. Joint learning for targeted sentiment analysis. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 4737–4742, 2018a.

Y. Ma, H. Peng, and E. Cambria. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In Thirty-Second AAAI Conference on Artificial Intelligence, 2018b.

M. Mitchell, J. Aguilar, T. Wilson, and B. Van Durme. Open domain targeted sentiment. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1643–1654, 2013.

M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos. Semeval-2015 task 12: Aspect based sentiment analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 486–495, 2015.

M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, A.-S. Mohammad, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, et al. Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pages 19–30, 2016.

T. A. Rana and Y.-N. Cheah. Aspect extraction in sentiment analysis: comparative analysis and survey. Artificial Intelligence Review, 46(4):459–483, 2016.

A. Rosenfeld and M. Thurston. Edge and curve detection for visual scene analysis. IEEE Transactions on computers, (5):562–569, 1971.

S. Sabour, N. Frosst, and G. E. Hinton. Dynamic routing between capsules. In Advances in neural information processing systems, pages 3856–3866, 2017.

M. Saeidi, G. Bouchard, M. Liakata, and S. Riedel. Sentihood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. arXiv preprint arXiv:1610.03771, 2016.

K. Schouten and F. Frasincar. Survey on aspect-level sentiment analysis. IEEE Transactions on Knowledge and Data Engineering, 28(3):813–830, 2015.

L. Shu, B. Liu, H. Xu, and A. Kim. Lifelong-rl: Lifelong relaxation labeling for separating entities and aspects in opinion targets. In Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing, volume 2016, page 225. NIH Public Access, 2016.

C. Sun, L. Huang, and X. Qiu. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint arXiv:1903.09588, 2019.

D. Tang, B. Qin, X. Feng, and T. Liu. Effective LSTMs for target-dependent sentiment classification. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 3298–3307, Osaka, Japan, Dec. 2016a. The COLING 2016 Organizing Committee.

D. Tang, B. Qin, and T. Liu. Aspect level sentiment classification with deep memory network. arXiv preprint arXiv:1605.08900, 2016b.

B. Wang and W. Lu. Learning latent opinions for aspect-level sentiment classification. In Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

C. Yang, H. Zhang, B. Jiang, and K. Li. Aspect-based sentiment analysis with alternating coattention networks. Information Processing & Management, 56(3):463–478, 2019.

M. Yang, W. Tu, J. Wang, F. Xu, and X. Chen. Attention based lstm for target dependent sentiment classification. In Thirty-First AAAI Conference on Artificial Intelligence, 2017.

M. Zhang, Y. Zhang, and D. T. Vo. Neural networks for open domain targeted sentiment. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 612–621, 2015.

M. Zhang, Y. Zhang, and D.-T. Vo. Gated neural networks for targeted sentiment analysis. In Thirtieth AAAI Conference on Artificial Intelligence, 2016.

P. Zhang, H. Zhu, T. Xiong, and Y. Yang. Co-attention network and low-rank bilinear pooling for aspect based sentiment analysis. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 6725–6729. IEEE, 2019a.

S. Zhang, X. Xu, Y. Pang, and J. Han. Multi-layer attention based cnn for target-dependent sentiment classification. Neural Processing Letters, pages 1–15, 2019b.

J. Zhou, J. X. Huang, Q. Chen, Q. V. Hu, T. Wang, and L. He. Deep learning for aspect-level sentiment classification: Survey, vision and challenges. IEEE Access, 2019.