Aspect-based Sentiment Analysis with Graph Convolution over Syntactic Dependencies

Anastazia Žunić, Padraig Corcoran, Irena Spasić*

School of Computer Science & Informatics, Cardiff University, The Parade, Cardiff, CF24 3AA, United Kingdom

Abstract

Aspect-based sentiment analysis is a natural language processing task whose aim is to automatically classify the sentiment associated with a specific aspect of a written text. In this study, we propose a novel model for aspect-based sentiment analysis, which exploits the dependency parse tree of a sentence using graph convolution to classify the sentiment of a given aspect. To evaluate this model in the domain of health and well-being, where this task is biased towards negative sentiment, we used a corpus of drug reviews. Specific aspects were grounded in the Unified Medical Language System, a large repository of inter-related biomedical concepts and the corresponding terminology. Our experiments demonstrated that graph convolution approach outperforms standard deep learning architectures on the task of aspect-based sentiment analysis. Moreover, graph convolution over dependency parse trees (F-score of 0.8179) outperforms the same approach over a flat sequence representation of sentences (F-score of 0.7332). These results bring the performance of sentiment analysis in health and well-being in line with the state of the art in other domains.

Keywords: sentiment analysis, natural language processing, dependency parsing, neural network, graph convolutional network.

---

*Corresponding author
    Email address:* `spasici@cardiff.ac.uk` (Irena Spasić)

1. Introduction

Sentiment analysis (SA) is a natural language processing (NLP) task which aims to classify the sentiment expressed or implied by a given piece of text. It can be applied at different levels of text organisation: the whole document [1], an individual paragraph or sentence, or a specific aspect [2]. This study focuses specifically on aspect-based SA. This task is particularly difficult in the domain of health and well-being, where the performance of sentiment analysis was found to lag behind the state of the art [3].

Recent proliferation of online platforms designed to share health-related information with other users sparked research interest in sentiment analysis in this domain. However, existing research in aspect-based SA is typically conducted using user reviews of products and services such as mobile devices and restaurants, but also pharmaceutical drugs, which are related to one's health and well-being. Sentiment analysis of drug reviews can be used to support pharmacovigilance by detecting new adverse drug reactions [4].

In the case of drug reviews, current research efforts in sentiment analysis are focusing on the whole document (i.e. review) and not on an individual aspect. This is partly related to the availability of annotations that can be used to train supervised classification approaches. Reviews typically come together with star rating, which can be easily converted into sentiment labels. Aspect-based SA requires manual data annotation, which has been identified as one of the key obstacles to machine learning approaches in clinical NLP [5]. SentiDrugs [2] represents a dataset relevant to the current study: it consists of drug reviews in which aspects were manually identified and annotated for sentiment. Unfortunately, this dataset is not publicly available.

In terms of methods used to support sentiment analysis in health and well-being, a recent systematic review revealed that rule-based and traditional machine learning approaches are used most commonly [3]. Both approaches require manual engineering of either rules or features, which limit their portability across different tasks and domains. On the other side, deep learning does not suffer

from these limitations and has demonstrated considerable success in a variety of NLP tasks including SA. For example, deep learning has been successfully applied to support SA of user reviews of hotels and restaurants as well as product reviews (phones, cameras, laptops, etc.) [6, 7]. Incorporating syntactic structure into the deep learning process to model sentiment compositionality can improve the performance of sentiment analysis by almost 10 percent points as confirmed by Socher et al. [8] who trained a recursive neural network on constituency parse trees. Dependency parse trees directly capture syntactic dependencies between the words and in that respect may be better placed to support aspect-based SA by traversing dependencies associated with the target word or phrase. From the existing variety of neural network architectures, graph convolutional networks (GCN) are most naturally suited to traversing the graph structure of syntactic dependencies.

In this study, we hypothesise that a GCN approach should outperform traditional neural network architectures on the task of aspect-based SA. To test this hypothesis, we developed a new approach to aspect-based SA based on graph convolution, where the aspect of interest is represented as a vertex in the graph representation of a sentence and convolution performed along its edges and those of its neighbours. To examine the effect of syntactic dependencies on sentiment polarity of a given aspect, we tested the proposed approach on two graph representations of a sentence including a simple sequence and a dependency parse tree. The experiments asserted the importance of features incorporating syntactic dependencies over sequential order for aspect-based SA.

This paper is organized as follows. Section 2 provides an overview of relevant literature. Section 3 describes the methodology. In Section 4, we evaluate the approach. Finally, Section 5 concludes the paper and outlines the future work.

2. Related work

A recent systematic review paper [3] provided evidence that a vast majority of approaches to SA in health and well-being is based on rule-based and tra-

3

ditional machine learning, whose results are sub-par to those achieved in other domains. These approaches tend to employ simple features such as distribution of words and n-grams, but which fail to capture syntactic dependencies between words and, consequently, the effect of syntax on semantics or, in this case, sentiment. Deep learning, on the other hand, can discover structural dependencies in the form of graphs or, when these dependencies are already known, easily integrate them into the learning process.

Traditional rule-based and machine learning approaches require features to be engineered manually. In contrast, deep learning applies layers of linear and non-linear data transformations to learn a representation of the problem that is best suited for the end task. Popular deep learning architectures include recurrent neural network (RNN) [9], long short-term memory (LSTM) [10], bidirectional LSTM (BiLSTM) [11], gated recurrent unit (GRU) [12], convolutional neural network (CNN) [13] and graph convolutional network (GCN) [14].

The RNN architecture is designed to process data sequentially, where the hidden state of the current element represents the memory of the network at the particular time step by capturing information about all previous time steps. The LSTM architecture is a type of RNN, which improves modelling of long range dependencies. This architecture contains three types of gates (input, forget and output gates), which are used to calculate the hidden state. BiLSTM consists of two LSTMs, where information is propagated in both forward and backward direction. GRU has a slightly simpler architecture than LSTM as it contains two gates (reset and update gates) which are used to memorize relevant information. The CNN architecture is translation invariant and it can extract important local features. Its subtype, GCN acts as a message passing algorithm, where the information between vertices in the graph is propagated along the edges, allowing the vertices to aggregate information from their neighbours.

Different deep learning architectures, although most commonly LSTM, have been applied to solve the problem of aspect-based SA. A groundbreaking approach employed hierarchical BiLSTM on constituency parse trees to perform aspect-based SA on restaurant and laptop reviews [6]. Similarly, an LSTM

4

approach achieved the accuracy of 84.0% and 89.9% on classifying customer reviews into positive and negative sentiment, respectively [7]. More recently, an LSTM-based architecture was combined with a lexicon to analyse the sentiment of restaurant reviews with an accuracy of 82.86% [15]. A deep learning approach to aspect-based SA in health and well-being performed below 80% accuracy as illustrated by a GRU architecture that achieved an accuracy of 78.26% on drug reviews [2]. Although such underperformance may be attributed to differences in architectures used and specific properties of the training data, it is in line with a previous finding from [3] that SA in health and well-being does lag behind the state of the art in other domains. This is mainly due to generally negative connotation of health-related concepts, which tends to skew the results of SA toward negative polarity. It is, therefore, ever so important to carefully examine the context when such concept are used as aspects in SA.

Bidirectional Encoder Representations from Transformers (BERT) is representative example of pre-trained language models in NLP [16], which can easily be fine-tuned using additional text to solve specific NLP tasks such as aspect-based SA. BERT itself provides contextual word representations, which are generated using transformers [17], which use an attention mechanism rather than recurrence. The attention mechanism is used to determine which words and sequences are important for the overall context. These features can then be exploited by an additional output layer on top of the base BERT model to make it task specific. The aspect-based SA task can be formulated as a sentence-pair classification task, such as question answering, by using the aspect as an auxiliary sentence [18, 19, 20]. Attention may not necessarily coincide with direct syntactic relations. To explicitly capture the compositional sentiment semantics and improve performance on phrase-level sentiment classification, SentiBERT was implemented as a variant of BERT that uses a recursive constituency parse tree structure to learn to predict the sentiment of the phrase nodes [21].

The sentiment of a specific aspect is directly influenced by its modifiers and not necessarily the entire context, which may end up adding unnecessary noise to the problem representation. Proximity and order are often used as proxies in

5

lieu of explicit dependencies. Therefore, most often, context will be represented using n-grams or sequences. Constituency parse trees also take advantage of the notions of proximity and order to group words together into coherent phrases, which represent key features for semantic analysis and thus can support applications such as question answering and information extraction [22]. However, applications such as aspect-based SA depend crucially on the use of modifiers, which can change or emphasise a particular word in a sentence. Universal dependencies represent grammatical relations between words in a sentence [23] in the form of triplets (name of the relation, governor and dependent), which give rise to a graph representation of a sentence, which is often collapsed into a tree [24]. In aspect-based SA, such structure can be used to explore those words that are logically associated with the given aspect regardless of their physical proximity and ignore (or downplay) those that are less relevant with respect to the expression of sentiment.

GCN which performs convolution over a graph, is a most naturally suited neural network architecture for this type of sentence representation. In NLP, this type of architecture has previously been employed to extract semantic relations from syntactic dependency trees in [25]. Most recently, GCN has been used to perform aspect-based SA on product reviews in [26] and on Twitter data in [27]. In [27] GCN was additionally combined with the syntactic dependencies.

In this study, we propose an approach to aspect-based SA based on GCN over a dependency parse tree of the sentence and compare its performance against other deep learning architectures.

3. Methodology

Aspect-based SA is a fine-grained sentiment analysis task, where the goal is to identify the sentiment of the specific aspect, rather than the overall sentiment of the document. This study considers binary classification of the given aspect, where the goal is to classify the sentiment of the given aspect into one of the two classes, positive or negative.

The overall system design is shown in Figure 1. A set of drug reviews was collected from Drugs.com [28], is the largest, most widely visited, independent pharmacologic drug information web site, which allows its users to post reviews describing personal experience about their use. All reviews were automatically annotated with concepts from the Unified Medical Language System (UMLS), a large repository of inter-related biomedical concepts and the corresponding terminology [29]. These concepts represent aspects whose sentiment needs to be classified. Input documents are processed by Stanford CoreNLP [30] to convert individual sentences into dependency graphs. Individual words representing vertices in such graphs were mapped onto their embeddings, which were pretrained on web data from Common Crawl using the GloVe method [31]. Each input sentence is represented as a sequence of tokens $S = (w_1, w_2, ..., w_n)$. When combined with word embeddings, this representation gives rise to a matrix of dimensions n x d, where n is the total number of tokens and d is the size of the embedding vector space. These data were combined with sentiment labels to train a neural network to classify the sentiment associated with individual UMLS concepts. The architecture of this neural network, shown in Figure 2, was based on graph convolution. The input consists of the dependency graph of the sentence whose vertices are convoluted by propagating information from other vertices across the edges of the graph. After two successive convolutions, the vertex corresponding to the aspect is mapped onto 2-dimensional classification space whose dimensions correspond to positive and negative polarity, respectively. This means that the two polarities produced as an output are extracted directly from the aspect and indirectly from its neighbours via convolution, and therefore their values may vary across different aspects within the same sentence.

3.1. Sentence representation

Sentence S is a sequence of tokens that can be represented as a graph $G = (V, E)$. Graph is an ordered pair (V, E), where V is a set of vertices (nodes) and E is a set of edges that represent pairs of vertices. Graph $G = (V, E)$ of
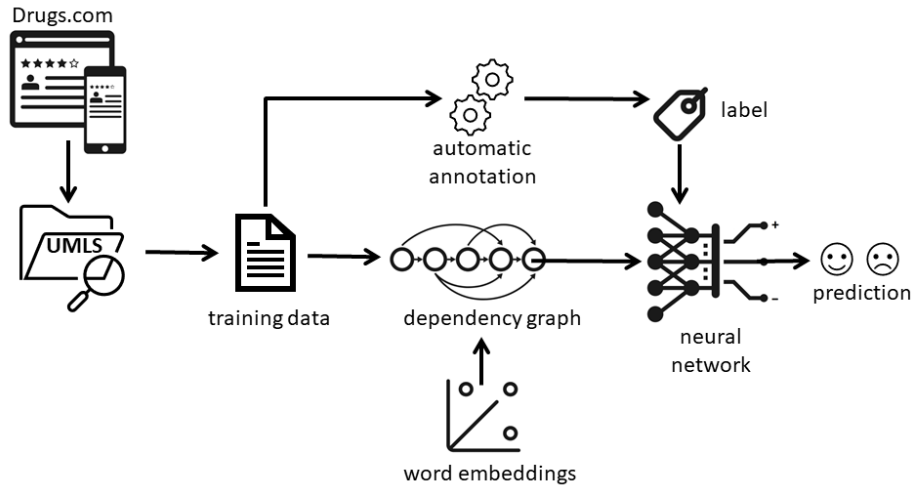
7

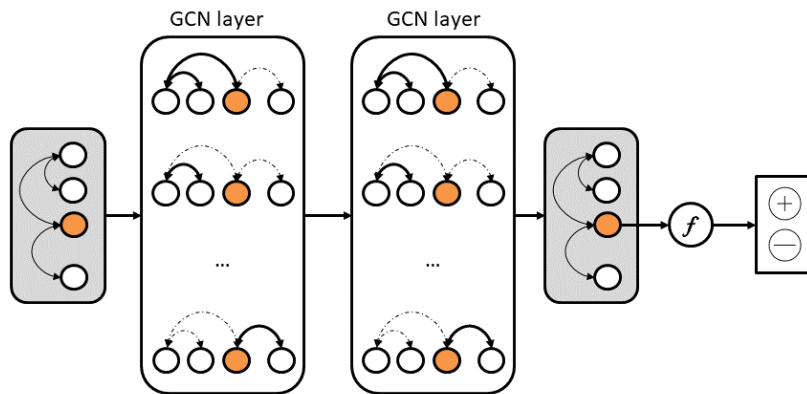Figure 1: System design for aspect-based SA of drug reviews.

Figure 2: Graph convolutional network over dependency graphs. The sentiment of an aspect, highlighted in the graph, is classified.

a sentence can be modelled in different ways, for example sequence graph, dependency graph, relational dependency graph, or constituency graph. Vertices represent the tokens within the sentence, and edges connect the vertices, i.e. tokens. The structure of sequence, dependency and relational dependency graph can be seen in Figure 3.

We decided to represent a sentence using its dependency parse tree, which is a special case of a directed dependency graph. Here, each vertex has a single parent except for the root. Figures 4 and 5 provide illustrations of dependency parse trees that are used as sentence representation. In order for the convolution to flow in both directions across a dependency graph, we chose to ignore the direction.

3.2. Graph convolution

Graph convolutional network (GCN) is a neural network architecture that takes a graph as an input and outputs an updated representation of each vertex in the given graph. The updated representation of each vertex is a function of the current representation of the given vertex and its neighbouring vertices. Various GCN architectures have been proposed [32]. In this study, we used two types of GCN architectures, which were applied to the sentence dependency tree. The architectures in question are *GraphSAGE* GCN (GS-GCN) [33] and relational (R-GCN) [34]. These are commonly used GCN architectures which have been found to provide good performance in empirical comparisons of different GCN architectures [35]. We now describe the details of each GCN architectures considered.

The input to each GCN architecture is the undirected dependency graph of the sentence, where each vertex is initialized with the corresponding 300-dimensional word embedding vector. Vertex representation is updated with each convolution over the graph. A sequence of k convolutions will propagate information across the graph to the k-th order neighbour. Our architectures each contain two graph convolution layers, which means that every vertex will get information from the second order neighbour. Specifically, the first convolution
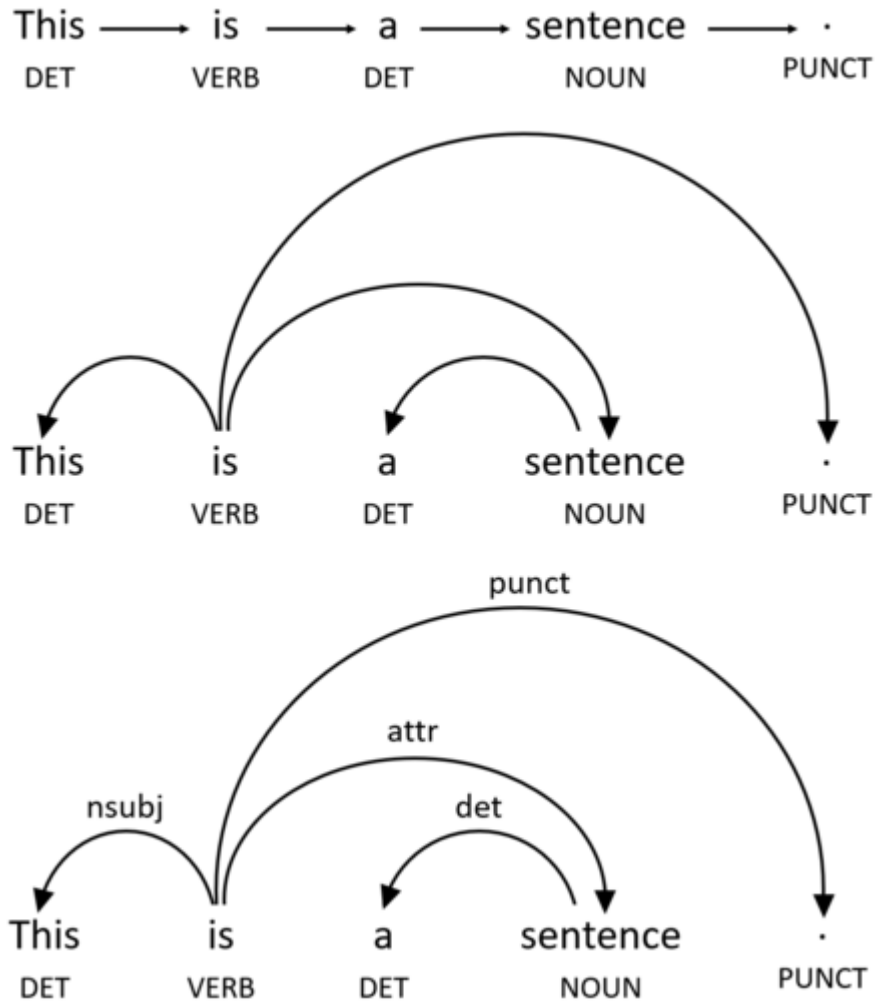
9

Figure 3: Three types of graphs that are considered for sentence representation: sequence graph, dependency graph, and relational dependency graph with different edge types from top to bottom respectively.
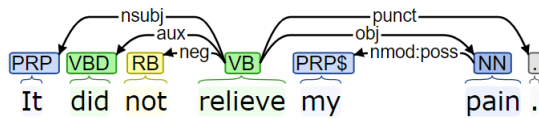

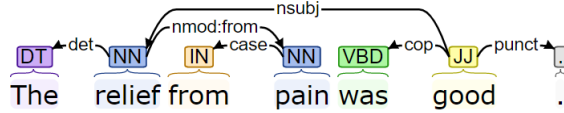Figure 4: Example of negation in a dependency parse tree.

Figure 5: Example of the dependency parse tree.

layer reduces the dimensionality of the input vectors from 300 to 125 and the second layer will outputs a 100-dimensional vector. In the GS-GCN architecture the hidden state h_i^t corresponding to vertex i and layer t after one convolution is updated as follows:

$$h_i^t = \sigma(W^t \cdot concat(h_i^{t-1}, aggregate(h_j^{t-1}, \forall j \in N(i)) + b^t) \tag{1}$$

where $N(i)$ denotes a set of neighbors of vertex i, $\sigma(\cdot)$ is a non-linear function, $concat(\cdot)$ represents concatenation of vectors horizontally and $aggregate(\cdot)$ indicates the summation of the neighbours of the corresponding vertex. Moreover, W^t and b^t are weight matrix and bias, respectively, for the t-th convolutional layer, which are the parameters of the model that are learned. Non-linear function $\sigma(\cdot)$ in this architecture is rectified linear unit (ReLU). The potential downside of this method can be that the edge types are not taken into consideration. The solution to that is to use another graph convolution algorithm known as relational graph convolution (R-GCN) [34]. In order to perform R-GCN, a graph is represented as a triplet $G = (V, E, R)$, where R is a set of relations r_i corresponding to each edge e_i. In the R-GCN architecture the hidden state h_i^t corresponding to vertex i and layer t after one convolution is updated as follows:

$$h_i^t = \sigma\left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^{t-1} h_j^{t-1} + W_0^{t-1} h_i^{t-1}\right) \tag{2}$$

where N_i^r denotes a set of neighbors of vertex i that are connected with relation $r \in R$, $c_{i,r}$ is a normalization constant. Relations r are directly extracted are extracted from the output of the Stanford CoreNLP dependency parser. Figures 4 and 5 provide examples of different edge types including *nsubj* - nominal

subject, *neg* - negation modifier, *obj* - object, etc. More detailed explanation about the syntactic relations can be found in Universal Dependencies documentation [24, 36]. An overview of the different graph types used in our experiments for sentence representation is shown in Figure 3.

3.3. Classification

After two successive graph convolutions, the hidden state of the sentiment aspect is retrieved and mapped into the classification space by a linear transformation, reducing the dimensionality of its embedding from 100 to 2, where the two dimensions corresponds to positive and negative polarity, respectively. The 2-dimensional vector is then passed through the softmax layer, which provides the probability distribution over the two sentiment polarities.

4. Experimental results and analysis

To validate the model proposed in this work we performed two sets of experiments. The first set compares different types of graphs as input to the GCN models, whereas the second set compares different neural network architectures including GCN, RNN and LSTM. All methods were implemented in the Python programming language using the PyTorch library for deep learning. The source code is available at `https://github.com/ispasic/ABSA-with-GC-over-Syntactic-Dependencies`.

4.1. Data

The raw data was originally collected from Drugs.com [28], the largest independent pharmacologic drug information web site, which allows its users to post their reviews. We re-used a subset of these reviews described in [37]. This subset was previously annotated for aspect-based SA [2], but unfortunately it is not publicly available. The choice of aspects in this study was motivated by the likely practical applications of SA on this dataset. The most obvious applications are related to the drugs' efficacy and safety, which could be inferred from the sentiment associated with the signs and symptoms discussed in drug reviews. To extract signs and symptoms, we automatically annotated a total

12

lexicon	description
AFINN [39, 40]	A list of 2477 words and phrases with an integer value as a sentiment score between -5 (negative) and 5 (positive).
EmoLex [41, 42]	A lexicon where items are annotated with 8 basic emotions and sentiment score of either positive or negative.
Harvard General Inquirer [43, 44]	Lexicon that provides 1915 positive and 2219 negative words.
MPQA [45, 46]	A lexicon of around 8000 items that provides sentiment scores.
Opinion lexicon [47, 48]	A list of approximately 6800 positive and negative words.
Wordnet Affect [49, 50]	An extension of WordNet, each item in the lexicon is labeled as positive, negative, ambiguous, or neutral.

Table 1: A selection of sentiment lexicons.

of 128,581 reviews with concepts from the Unified Medical Language System (UMLS), a large repository of inter-related biomedical concepts and the corresponding terminology [29]. We focused on a single subclass of UMLS concepts that represents clinical signs and symptoms [38]. We then cross-referenced these concepts against six lexicons described in Table 1 to identify those with negative sentiment. This choice was based on a previous finding that the negative connotation of health symptoms tends to skew the SA results toward the negative spectrum [3]. In other words, the sentiment of such aspects is more challenging to classify. In this study, we want to focus specifically on this bias by exploring the ways in which the context (represented by syntactic dependencies) can modify the negative polarity associated with signs and symptoms.

The most frequently mentioned signs and symptoms were selected to represent aspects whose sentiment needs to be classified: *burning, constipation, dizziness, dizzy, dry, fatigue, headache, nausea, nauseous, nauseated, pain, painful, sick, sickness, symptom, tired* and *tiredness*. The silver-standard sentiment of each aspect was inferred from the corresponding user review's star rating ranging from 1 and 10. To easily convert star rating into sentiment, we annotated reviews with star rating of 1 or 2 with negative sentiment, those with rating of 9 or 10 with positive sentiment and removed the remaining reviews from further consideration. The overall distribution of sentiment before the selection of short reviews is shown in Figure 6.
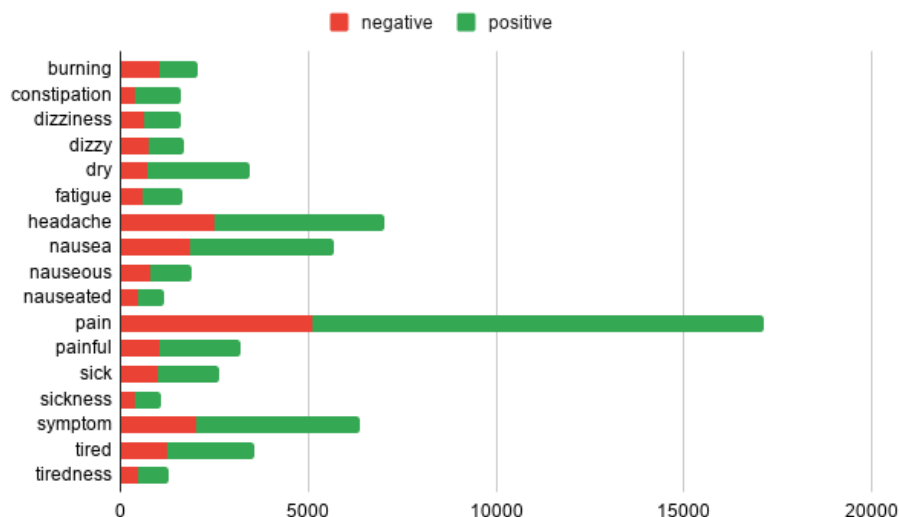
13

Figure 6: Distribution of aspect-based sentiments.

To obtain a reliable sentence-level sentiment, which may vary across a long document, we only used reviews consisting of a single sentence. This reduced the need for manual annotation albeit at the cost of reducing the size of the dataset. As a result, we ended up with a set of 806 positive and 612 negative reviews, which were then curated manually. After incorrectly annotated reviews were removed, 79.28% of positive reviews and 96.89% of negative reviews were retained. Finally, each aspect (i.e. a reference to a sign or symptom from Figure 6) was mapped to the sentiment of the corresponding sentence, providing a total of 1,232 annotated aspects out of which 639 were positive and 593 negative, which represents a well-balanced set despite the fact that the chosen signs and symptoms are otherwise inherently negative.

Data were split randomly to use 80% for training and keep the remaining 20% for testing. From the training subset, 20% was used to tune hyperparameters prior to training the model on the remaining 80%. The distribution of sentiment within training, validation and testing is provided in Table 2.

14

	positive	negative	total
train	410	378	788
validation	99	98	197
test	130	117	247
total	639	593	1232

Table 2: The distribution of sentiment in the training, validation and test sets.

4.2. Hyperparameters

The model was trained with backpropagation by minimising cross entropy loss function, and optimized with the Adam optimizer. Learning rate was set to 0.001. To avoid overfitting we applied early stopping, where the training was stopped if the validation loss increased in five consecutive epochs, therefore the patience was set to 5. Apart from early stopping, dropout was used as regularization and set to 0.2, which means that during the training process 20% of the weights were set to zero.

4.3. Evaluation measures

The measures used to evaluate the performance of the model included accuracy, F-score and cross entropy loss. Accuracy is calculated as the percentage of correctly classified instances. Precision and recall are calculated using true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) in the following way: $P = TP/(TP + FP)$ and $R = TP/(TP + FN)$. These two values are combined into the F-score as follows: $F = 2 * P * R/(P + R)$. Cross entropy loss for binary classification is defined as:

$$loss = -\frac{1}{n} \sum_{i=1}^{n} ln(p_i) \tag{3}$$

4.4. Sentence representation

To investigate whether dependency graph is better suited for modelling aspect-based SA than a simple sequence (also represented by a graph), we used both and compared the results. In both cases, we performed experiments with

15

graph type	loss	F-score	accuracy (%)
sequence	0.6348	0.6529	65.59
undirected sequence	0.5635	0.7332	73.68
dependency	0.5953	0.7001	70.04
undirected dependency	**0.4570**	**0.8179**	**81.78**

Table 3: Evaluation of the GS-GCN architectures over sequence and dependency graphs with different edges.

both directed and undirected graphs. The results achieved using the GS-GCN model over different types of input graphs are reported in Table 3. The best performance across all considered measures were achieved when an undirected dependency graph was used for sentence representation.

4.5. Neural network architectures

To establish the baseline, we also performed experiments with two architectures, RNN and LSTM, which have previously been used to support aspect-based SA [7, 15]. They used a sequence, which is equivalent to a sequence graph we defined earlier. Note that, CNN was not considered because it operates on a feature vector used to represent the whole sentence rather than each token individually [51].

Apart from the standard RNN, LSTM and two GCN architectures, we also performed experiments with BiLSTM-GS-GCN, in which the input was passed through a BiLSTM as proposed in [26] prior to performing graph convolution. These models were trained using the Adam optimizer with learning rate of 0.001. Dropout of 0.2 and early stopping with patience of 5 were applied to reduce overfitting. The results are presented in Table 4. The four architectures based on graph convolution outperform the rest. The best results were achieved using the GS-GCN over the undirected dependency graph. These results indicate that a dependency graph enriches sentence representation, which in turn enables the model to exploit syntactic dependencies for semantic reasoning.

We used the Z-test for the equality of two proportions [52] to test statistical

method	sentence representation	loss	F score	acc (%)	pos acc (%)	neg acc (%)
RNN	sequence	0.6202	0.6725	67.61	76.92	57.26
LSTM	sequence	0.6259	0.6725	67.61	76.92	57.26
R-GCN	undirected dependency	0.6265	0.6833	68.42	76.92	58.97
GS-GCN	undirected sequence	0.5635	0.7332	73.68	77.85	68.39
BiLSTM+GS-GCN	undirected dependency	0.5095	0.7566	75.71	76.92	70.09
GS-GCN	undirected dependency	**0.4570**	**0.8179**	**81.78**	**78.46**	**85.47**

Table 4: Evaluation of the baseline models and of the proposed model (GS-GCN-undirected dependency).

significance of differences in accuracy between GS-GCN against that of the five baseline methods, respectively. Table 5 shows that all p-values are ≤ 0.05, hence we reject the null hypothesis and conclude that there is significant difference be-
340   tween the given architectures mechanisms in terms of their accuracy. Therefore, the GS-GCN over the undirected dependency graph is indeed significantly more accurate than any of its counterparts.

Evaluation of the baseline architectures and of the proposed model over positive and negative sentiments can be seen in Table 4. Most of the architectures
345   perform better on classification of positive sentiment. This trend was overturned by GCN over the dependency graph, while still providing the best results on classification of positive sentiment, suggesting that this approach makes the best utilisation of context to perform aspect-based SA.

method	accuracy (%)	p-value
RNN	67.61	0.0001
LSTM	67.61	0.0001
R-GCN	68.42	0.0003
GS-GCN (undirected sequence)	73.68	0.0152
BiLSTM+GS-GCN	75.71	0.0496

Table 5: Comparison of the GS-GCN over the undirected dependency graph with accuracy of 81.78% against the baseline methods.

5. Conclusion

We proposed a new approach to aspect-based SA based on graph convolution over the dependency parse tree of the sentence. The experimental results show that, relative to other neural network architectures and sentence representations, this approach makes the best utilisation of context to perform aspect-based SA. We specifically looked at the sentiment surrounding medical signs and symptoms because of the negative sentiment underlining their semantics, which makes SA in the domain of health and well-being challenging. For instance, for someone suffering from a chronic condition, having a good quality of life is not necessarily measured by the absence of associated signs and symptoms, but rather by the extent to which they can be successfully managed and controlled. However, the negative connotation of signs and symptoms tends to skew the results of SA toward negative polarity. This is one of the reasons SA in health and well-being is performing below the F-score of 60% on average, lagging behind the state of the art in SA on service and product reviews, where F-score is found to be above 70% and 80%, respectively [3]. In this study, we successfully tackled this bias by exploring the ways in which the context (represented by syntactic dependencies) can modify the negative polarity associated with signs and symptoms and effectively closed this performance gap by achieving the state-of-the-art results regardless of the domain. Future work will consider incorporation of the attention mechanism into the model to better differentiate

₃₇₀ among syntactic dependencies that act as sentiment modifiers.

References

[1] S. Yadav, A. Ekbal, S. Saha, P. Bhattacharyya, Medical sentiment analysis using social media: Towards building a patient assisted system, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation, European Language Resources Association, Miyazaki, Japan, 2018.

[2] Y. Han, M. Liu, W. Jing, Aspect-level drug reviews sentiment analysis based on double bigru and knowledge transfer, IEEE Access 8 (2020) 21314–21325.

[3] A. Žunić, P. Corcoran, I. Spasić, Sentiment analysis in health and well-being: Systematic review, JMIR Medical Informatics 8 (1) (2020) e16023.

[4] I. Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou, G. H. Gonzalez, Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts, Journal of biomedical informatics 62 (2016) 148–158.

[5] I. Spasić, G. Nenadić, Clinical text data in machine learning: Systematic review, JMIR Medical Informatics 8 (3) (2020) e17984.

[6] S. Ruder, P. Ghaffari, J. G. Breslin, A hierarchical model of reviews for aspect-based sentiment analysis, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, TX, USA, 2016, pp. 999–1005.

[7] Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in: Proceedings of the 2016 conference on empirical methods in natural language processing, Association for Computational Linguistics, Austin, TX, USA, 2016, pp. 606–615.

[8] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Seattle, WA, USA, 2013, pp. 1631–1642.

[9] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536.

[10] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[11] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional lstm and other neural network architectures, Neural networks 18 (5-6) (2005) 602–610.

[12] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1724–1734.

[13] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[14] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 2017.

[15] L. Bao, P. Lambert, T. Badia, Attention and lexicon regularized lstm for aspect-based sentiment analysis, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Re-

20

search Workshop, Association for Computational Linguistics, Florence, Italy, 2019, pp. 253–259.

[16] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding (2019). `arXiv:1810.04805`.

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 6000–6010.

[18] C. Sun, L. Huang, X. Qiu, 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, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 380–385.

[19] H. Xu, B. Liu, L. Shu, P. Yu, BERT post-training for review reading comprehension and aspect-based sentiment analysis, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 2324–2335.

[20] M. Hoang, O. A. Bihorac, J. Rouces, Aspect-based sentiment analysis using bert, in: Proceedings of the 22nd Nordic Conference on Computational Linguistics, 2019, pp. 187–196.

[21] D. Yin, T. Meng, K.-W. Chang, SentiBERT: A transferable transformer-based architecture for compositional sentiment semantics, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 3695–3706.

[22] D. Jurafsky, J. H. Martin, Constituency parsing, in: Speech and Language Processing (3rd Edition), draft, 2019, Ch. 13, pp. 232–245.

[23] M.-C. de Marneffe, T. Dozat, N. Silveira, K. Haverinen, F. Ginter, J. Nivre, C. D. Manning, Universal Stanford dependencies: A cross-linguistic typology, in: Proceedings of the Ninth International Conference on Language Resources and Evaluation, European Language Resources Association, Reykjavik, Iceland, 2014, pp. 4585–4592.

[24] M.-C. De Marneffe, C. D. Manning, Stanford typed dependencies manual, Tech. rep., Technical report, Stanford University (2008).

[25] Y. Zhang, P. Qi, C. D. Manning, Graph convolution over pruned dependency trees improves relation extraction, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 2205–2215.

[26] K. Sun, R. Zhang, S. Mensah, Y. Mao, X. Liu, Aspect-level sentiment analysis via convolution over dependency tree, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 5683–5692.

[27] X. Bai, P. Liu, Y. Zhang, Exploiting typed syntactic dependencies for targeted sentiment classification using graph attention neural network, arXiv preprint arXiv:2002.09685.

[28] Drugs.com, `https://www.drugs.com/`.

[29] O. Bodenreider, The unified medical language system (umls): integrating biomedical terminology, Nucleic acids research 32 (suppl_1) (2004) D267–D270.

[30] D. Chen, C. D. Manning, A fast and accurate dependency parser using neural networks, in: Proceedings of the 2014 conference on empirical methods

480   in natural language processing, Association for Computational Linguistics,
        Doha, Qatar, 2014, pp. 740–750.

[31]   J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word
        representation, in: Proceedings of the 2014 conference on empirical meth-
        ods in natural language processing, Association for Computational Linguis-
485   tics, Doha, Qatar, 2014, pp. 1532–1543.

[32]   Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, S. Y. Philip, A comprehensive
        survey on graph neural networks, IEEE Transactions on Neural Networks
        and Learning Systems (2020) 1–21.

[33]   W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on
490   large graphs, in: Advances in neural information processing systems, Cur-
        ran Associates Inc., Long Beach, California, USA, 2017, pp. 1024–1034.

[34]   M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov,
        M. Welling, Modeling relational data with graph convolutional networks,
        in: European Semantic Web Conference, Springer International Publishing,
495   Heraklion, Greece, 2018, pp. 593–607.

[35]   F. Errica, M. Podda, D. Bacciu, A. Micheli, A fair comparison of graph
        neural networks for graph classification, in: Proceedings of the 8th Inter-
        national Conference on Learning Representations (ICLR), 2020.

[36]   Universal Dependencies, `https://universaldependencies.org/`.

500 [37] F. Gräßer, S. Kallumadi, H. Malberg, S. Zaunseder, Aspect-based senti-
        ment analysis of drug reviews applying cross-domain and cross-data learn-
        ing, in: Proceedings of the 2018 International Conference on Digital Health,
        Association for Computing Machinery, Lyon, France, 2018, pp. 121–125.

[38]   L.-T. T. Tran, G. Divita, M. E. Carter, J. Judd, M. H. Samore, A. V.
505   Gundlapalli, Exploiting the UMLS Metathesaurus for extracting and cate-
        gorizing concepts representing signs and symptoms to anatomically related
        organ systems, Journal of biomedical informatics 58 (2015) 19–27.

[39] F. Å. Nielsen, A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, arXiv preprint arXiv:1103.2903.

[40] AFINN, `http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010`.

[41] S. M. Mohammad, P. D. Turney, Crowdsourcing a word–emotion association lexicon, Computational Intelligence 29 (3) (2013) 436–465.

[42] NRC Emotion Lexicon, `http://sentiment.nrc.ca/lexicons-for-research/`.

[43] P. J. Stone, D. C. Dunphy, M. S. Smith, The general inquirer: A computer approach to content analysis.

[44] Harvard General Inquirer, `http://www.wjh.harvard.edu/~inquirer/homecat.htm`.

[45] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of human language technology conference and conference on empirical methods in natural language processing, Association for Computational Linguistics, Vancouver, British Columbia, Canada, 2005, pp. 347–354.

[46] MPQA, `http://mpqa.cs.pitt.edu/`.

[47] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Association for Computing Machinery, Seattle, WA, USA, 2004, pp. 168–177.

[48] Opinion Lexicon, `https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon`.

[49] C. Strapparava, A. Valitutti, WordNet affect: an affective extension of WordNet, in: Proceedings of the Fourth International Conference on Lan-

guage Resources and Evaluation, European Language Resources Association, Lisbon, Portugal, 2004.

[50] WordNet Affect, `http://wndomains.fbk.eu/wnaffect.html`.

[51] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1746–1751.

[52] G. K. Kanji, 100 Statistical Tests, Sage, 2006.