1st International Conference on Data Science and Intelligent Computing 2022- (ICDSIC2022)

An improved Bi-LSTM performance using Dt-WE for implicit aspect extraction

1st Omar Mustafa Al-Janabi
school of computer sciences
Universiti Sains Malaysia
Penang, Malaysia
Alnukhba University College,
Baghdad, Iraq
o.mustafa@alnukhba.edu.iq

2nd Mohammed K. Ibrahim
school of computer sciences
Universiti Sains Malaysia
Penang, Malaysia
mohammed_9218@student.usm.my

3rd Abdulkarim Kanaan-Jebna
Faculty of Information and Communication Technology
Universiti Tunku Abdul Rahman
Perak, Malaysia
a.kanaan@msn.com

4th Osamah Mohammed Alyasiri
school of computer sciences
Universiti Sains Malaysia
Penang, Malaysia
Al-Furat Al-Awsat Technical University
Karbala, Iraq
osama.alyasiri@atu.edu.iq

5th Hiba J. Aleqabie
Computer Science Department
Faculty of Computer Science and Information Technology
Karbala, Iraq
hiba.jabbar@uokerbala.edu.iq

***Abstract*—In aspect-based sentiment analysis ABSA, implicit aspects extraction is a fine-grained task aim for extracting the hidden aspect in the in-context meaning of the online reviews. Previous methods have shown that handcrafted rules interpolated in neural network architecture are a promising method for this task. In this work, we reduced the needs for the crafted rules that wastefully must be articulated for the new training domains or text data, instead proposing a new architecture relied on the multi-label neural learning. The key idea is to attain the semantic regularities of the explicit and implicit aspects using vectors of word embeddings and interpolate that as a front layer in the Bidirectional Long Short-Term Memory Bi-LSTM. First, we trained the proposed domain-trained word embeddings (Dt-WE) model using explicit and implicit aspects. Second, interpolate Dt-WE model as a front layer in Bi-LSTM. Finally, extract implicit aspects by testing the trained architecture using the opinionated reviews that comprise multiple implicit aspects. Our model outperforms several of the current methods for implicit aspect extraction (Source Code).**

Keywords—Word Embeddings, Bi-LSTM, Multiple Implicit Aspect Extraction, Sentiment Analysis

I. INTRODUCTION

The exponential growth of the online product reviews on the Web encouraged the sentiment analysis researchers to extract the aspect-terms and predict it is semantic polarity. However, the difficulty of extracting the aspect-terms vary according to the aspect-type in the opinionated reviews. Adequate number of methods proposed for the extraction of explicit aspects in comparison to the proposed methods for implicit aspects extraction, as the latter is tough task to accomplish. Example 1 shows a review text, in which the **'picture quality'** is an explicit aspect terms in "camera" domain-data.

- **Example 1**: a few of my work constituents owned the g2 and highly recommended the canon for picture quality.

The appeared aspect "picture quality" is explicitly appeared in the review and extracting that has been previously accomplished using several methods like sequential rules [1], neural network [2], topic models [3][4].

While Example 2 stated a sentimental review that carries hidden aspect (i.e. size) in it is context.

- **Example 2**: while light, it will not easily go in small handbags or pockets.

The extraction of the implicit aspect "size" that represented by an opinion-word/clue "small", is a challenging task, and required a specially designed method for the extraction of such aspects, as it being extracted using handcrafted rules [5]. Other researchers relied on the explicit aspects for the extraction of implicit one. For instance, a co-occurrence between the explicit aspects and the opinion-words is being used for the implicit aspects extraction [6],[7]. On the other hand, a handcrafted rules were used for the extraction of the implicit aspects as in [8],[9]. Further, Example 3, and Example 4 shows a multiple implicit aspect in a single review.

- **Example 3**: Pictures taken can get blurred because of lack of image stabilizer but overall a great option for given budget
- **Example 4**: It seems quite small to me and very light

Multiple implicit aspect in example 3, they are "camera quality", and "price". While it is "size" and "weight" in example 4. However, few effort has accomplished the extraction of multiple implicit aspect like [6], [10]. To the best of our knowledge, this is the first study used multi-label recurrent neural model proposed for the extraction of multiple implicit aspects in online reviews. Our model relied on the vectors of word embeddings and the Bidirectional Long Short-Term Memory (Bi-LSTM). We massively evaluate the performance of the model using text-data from different domains. The experimental results shown the superiority of our model over the current efforts with minimal cost. Firstly, we accomplish the multiple implicit aspects extraction.

Secondly, we reduced the needs for the manually labeled handcrafted rules or co-occurrence methods for the extraction of the implicit aspects. Lastly, we achieved the task of implicit aspects extraction in different domains.

The rest of the paper organized as follow: Section (II) stated the related methods for the implicit aspects and multi-label learning algorithms. Section (III) stated our model for the extraction of the multilabel implicit aspects. Section (IV), and (V) stated the experimental results and the conclusion of the work.

II. Related Methods

A. *Implicit Aspects Extraction*

The methods for implicit aspects extraction mainly categorized under three main categories: Supervised, Un-supervised, and Semi-supervised machine learning methods [11] [12] [13]. The given terminologies for these categories are based on the necessity for the labelled and unlabeled datasets. For instance, in the supervised methods, they are in need for the labelled datasets for the training and testing of the models, whereas, it is not for the Un-supervised methods, which required no class label for the training. Further, Semi-supervised methods fall between these two methods (i.e. supervised, and unsupervised) in which, the dataset may have a bunch of labelled data points (class label) with a lot of unsupervised data for the training of the Semi-supervised methods and that is atypical.

1) *Supervised Methods*

As a supervised method, Mowlaei, et al., [14] proposed Particle Swarm Optimization algorithm that being combined with a lexicon-based method for the extraction of implicit aspects. In [15], they have utilized a classification method for the extraction of the explicit aspects, then identify the implicit aspects by matching the opinion-words to the extracted explicit aspects, while lexicon-based (i.e. WordNet) used to find the matching using a similarity measure. Whilst, in [16], the implicit aspects extracted based on the similarity measure between the opinion-words and clusters of explicit aspects.

In [17], first, they extracted the explicit aspects using a syntactic-rules. Second, train a classification model using the extracted explicit aspects. Finally, the classification model tested using the opinionated sentences for the extraction of the implicit aspects. Besides, a handcrafted rules were interpolated into a classification methods for the extraction of implicit aspects, and that is a rule-based crafted using a dependency-parser interpolated into a Convolutional neural network [18].

2) *Unsupervised Methods*

The widely used unsupervised methods for the extraction of implicit aspects are the co-occurrence methods. The co-occurrence methods basically works as follows, extracting the explicit aspects using a certain method like rule-based or frequency-based then find the score of co-occurrence between the opinion-words in the implicit aspects reviews and the extracted explicit aspects [7] [19] [20].

In the other hand, a lexicon-based [21] and the hierarchy lexicon-based [20] methods were also used as an unsupervised methods for the extraction of implicit aspects.

B. *Multi-label learning algorithms*

Multi-label learning can be designed to learn from variant sources of data (e.g. text, image, videos). However, Multi-label text classification has appeared as a problem of assigning each document to a subset of categories [22][23]. Therefore, multi-label learning algorithms emerged for the classification of the text data with multiple class-label. The approaches to solve the multi-label problem can be categorized into problem transformation approaches, and algorithm adaption approaches [24].

III. Our Model

For the extraction of the multiple implicit aspects, we proposed a multi-label learning algorithm that is the word embedding model (i.e. CBoWs) [25] along with the deep recurrent neural network (i.e. LSTM) [26] [27]. Therefore, we first describe word embeddings model in section A. Then, in section B, describe the LSTM neural network architecture (hidden layers). Fig. 1. Shows the framework architecture of the proposed model.

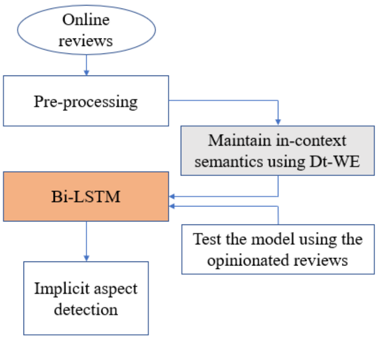

Fig. 1. Framework architecture of the proposed model

A. *Word Embeddings*

Word embedding is a vector representation of words learned using the trained words, have been shown to convey semantic relations of the trained words. In practice, CBoWs [25] accepts only one-word per-context. It will predict one target word for one context word. It can be thought of as a bigram model considering the semantic associations of the words. In bigram model, two words are linked together if they frequently occur in the context, but that lacks when it comes to the aspects extraction; not all the frequently occurred pair of words are aspect-term, also, bigram fall short for the semantic association of the words. In contrast, the CBoWs technique considered the frequent and infrequent words and also consider the semantic association between them.

In practice, for a given sequence of training words $w_1, w_2, w_3, \dots, w_T$, the word vector model is used to maximize the average log probability, as equation (1) shows.

$$\frac{1}{T}\sum_{t=k}^{T-k} log\ p(w_t|w_{t-k}, \dots\dots\dots, w_{t+k}) \quad (1)$$

For every word in the vocabulary V, each word is mapped to a unique vector, that is represented by a column in a matrix W. While the column is indexed by the position of the word in the vocabulary. The prediction task performed using a multiclass classifier, that is Softmax as in equation (2). However, the sum or the concatenation of the vectors is then used as features for prediction of the next word in the sentence.

For instance, the context of three words ("the," "food," "was") is used to predict the fourth word ("amazing").

$$p(w_t|w_{t-k}, \ldots\ldots, w_{t+k}) = \frac{e^{y_{wt}}}{\sum_i e^{y_i}} \quad (2)$$

For each output word i, un-normalized log-probability for each y_i, is computed as in equation (3).

$$y = b + Uh(w_{t-k}, \ldots\ldots\ldots, w_{t+k}; W) \quad (3)$$

Where U, b are the Softmax parameters. h is created by an average/concatenation of word vectors extracted from W.

B. Bi-LSTM neural network architecture

The proposed Bi-LSTM architecture is distinctive in that it has context neuron that represent the concept of a short memory, it holds values between calls between the neural networks. These contexts start by zero value and then update the values while moving to the next sequences.

LSTM is made up of three gates:

Forget Gate f_t – Controls if/when the context is forgotten.

Input Gate i_t – Controls if/when a value should be remembered by the context.

Output Gate o_t – Controls if/when the remembered value can pass from the unite.

First, calculate the forget gate based on the Sigmoid function S that flips into that 0 or 1 range by considering the weight for forgetting W_f. Zero (0) means we should forget, while one (1) means we should remember.

$$f_t = S\left(W_f.[\hat{y}_{t-1}, x_t] + b_f\right) \quad (4)$$

f_t is the forget gate, it is the result of S function of multiplying the W_f by the previous output $\hat{y}_{t-1}$ and the current input x_t (the input is a vector), and adding the bias which is also a learning parameter. So, by adjusting the weights and the biases we will be able to learn when to forget.

Second, calculate the input gate i_t, which is exactly the same function for the output gate, and it will remember when the results are one (1).

$$i_t = S(W_i.[\hat{y}_{t-1}, x_t] + b_i) \quad (5)$$

Similarly, the input weight W_i and the biases b_i learning parameters are affecting the adjusted architecture when to remember.

Context $\tilde{c}_t$ is the value that is remembering, which is the output from the neuron, that can be (-1) to (1) as it use tanh function, and it can't be the S function, because it clip off the values if anything below zero. tanh function have wider range of values that can we deal with if the values are lower than zero.

The context function $\tilde{c}_t$ (which also named as the candidate context) calculates the weight and the biases at the same way it being calculated in the other gates (e.g. input and output gates).

$$\tilde{c}_t = \tanh\left(W_c \,.\, \left[\hat{Y}_{t-1}, x_t\right] + b_c\right) \quad (6)$$

While C_t determine the new context, which is a switch gate, that have the forget and the input gates. Input here means which should go into the context, forget means should we remember the previous context. Plus (+) is piping the forget and the input gates together. In other words, if the forget f_t is zero (0) it will wipeout the previous context C_{t-1} because it is coefficient. Input gate i_t here is multiplied by the candidate context $\tilde{c}_t$

$$C_t = f_t \,.\, C_{t-1} + i_t \,.\, \tilde{c}_t \quad (7)$$

o_t is the output gate in the followed equation, which is also calculated using S activation function.

$$o_t = S\left(W_o \,.\, \left[\hat{Y}_{t-1}, x_t\right] + b_o\right) \quad (8)$$

$\hat{y}_t$ is the actual output in architecture, that is calculated by multiplying the output gate o_t and the tanh of the context C_t.

$$\hat{y}_t = o_t \,.\tanh(C_t) \quad (9)$$

Even though, LSTM model overcome the performance of standard RNN, where it overcome the problem of vanishing gradient problem, by using nonlinear activation function, and incorporating gating functions into their states. Yet, LSTM lack the ability to take contextual information of the reverse order of the sequences. Therefore, we have proposed a Bidirectional LSTM (Bi-LSTM), which read the sequences from both directions. However, reading the sentences from it is both directions is needed to detect the implicit aspects in the trained reviews.

IV. Experiments and Results

A. Data set

We chose two datasets that being manually annotated with multiple aspect-terms to evaluates the performance of our model. The first dataset is SemEval-2014 Restaurant [28] that contains Restaurant online reviews. Each sentence in the corpus is assigned with one or more aspect (which make it multiple aspects), and these aspects further being assigned into aspect-category. Second dataset is SemEval-2015 [29]. Table I stated the characteristic of the used datasets.

TABLE I. TRAINING DATASETS.

Dataset	Short form	#Category	#Reviews	
			Train	*Test*
SemEval-2014 Restaurant	"R:14"	5	3041	800
SemEval-2015 Restaurant	"R:15"	5	1315	761

B. Results and Discussion

This section report series of experiments that compared our model with the current methods for implicit aspect extraction (e.g. supervised, unsupervised, rule-based methods).

The advised method in this work, (the supervised method) is a representation learning based on neural network. That reduced the needs for rule-based methods for the extraction, that is by replacing the task-specific feature engineering by a continuous real-valued vector. We have presented a Domain-trained Word Embedding (Dt-WE: is word embedding model (i.e. CBoWs) trained using the same domain data for the training of Bi-LSTM (e.g. SemEval-2014 dataset) model to be interpolated to the Bi-LSTM neural learning for the extraction of implicit aspects. Given a collection of implicit aspects, Dt-WE is able to learn the syntactic and semantic meaning of the words.

The dimension size of the Dt-WE is set to 300. The window size and number of epochs set to 7 and 32, respectively. The trained Dt-WE embedding model is initiated as a first layer in neural model. Followed by a masking and dropout layers, then followed by the Bi-LSTM layer. In Bi-LSTM model, there are forward transfer layer and backward transfer layer. The bidirectional structure is considering a full account of the contextual information in the encoding of the sentimental reviews. Followed by a dense unite used 'sigmoid' as an activation function. Finally, the model compiled using non-linear 'nadam' optimizer on four aspect categories.

Table II, III stated the performance of implicit aspect extraction using a different configuration of a supervised learning. As can be seen in the tables, the performance of using Dt-WE model interpolated into the neural network (Bi-LSTM) has achieved the highest accuracy compared to the pre-rained word embedding (GloVe) in LSTM and Bi-LSTM models. Precision as an evaluation metric is the most crucial evaluation method that assess the performance of the truly predicted implicit aspects compared to the actual implicit aspects in the class label.

Table III highlighted the best performance using bold font. On which, the performance of Bi-LSTM and Dt-WE has achieved the highest accuracy in terms of Precision, Recall, and F-score that is nearly 0.80% for all of them. Whilst the performance of using LSTM with the pretrained word embeddings (GloVe) has achieved lower accuracy, that is 0.62%, 0.79%, and 0.70% for the Precision, Recall, and F-score respectively. Consequently, Table III presented the performance of the supervised methods against R:15 dataset. BiLSTM model with the Dt-WE model has achieved highest accuracy compared to the other methods.

Fig. 2 and 3 show the performance of the proposed model in terms of Precision, Recall, and F-score for implicit aspect detection. In the four settings, the LSTM model interpolated with the Dt-WE model has achieved highest accuracy.

TABLE II. SUPERVISED LEARNING ON THE R:14 DATASET FOR IMPLICIT ASPECT EXTRACTION.

Model	Precision	Recall	F-score
LSTM + GloVe	0.6296	0.7951	0.7027
Bi-LSTM + GloVe	0.5777	0.8113	0.6748
LSTM + Dt-WE	0.8026	0.8059	0.8043
Bi-LSTM + Dt-WE	**0.8059**	**0.8059**	**0.8059**

TABLE III. SUPERVISED LEARNING ON THE R:15 DATASET FOR IMPLICIT ASPECT EXTRACTION.

Model	Precision	Recall	F-score
LSTM + GloVe	0.668	0.8128	0.7333
Bi-LSTM + GloVe	0.6217	0.8128	0.7045
LSTM + Dt-WE	0.8413	0.8186	0.8297
Bi-LSTM + Dt-WE	**0.8474**	**0.8186**	**0.8327**

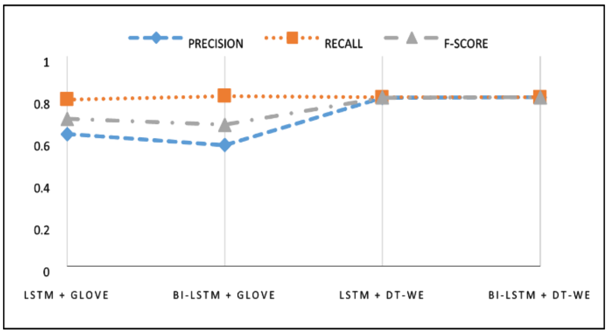

Fig. 2. Performance of LSTM architecture trained by various GloVe and Dt-WE word embedding models using R:14 dataset.

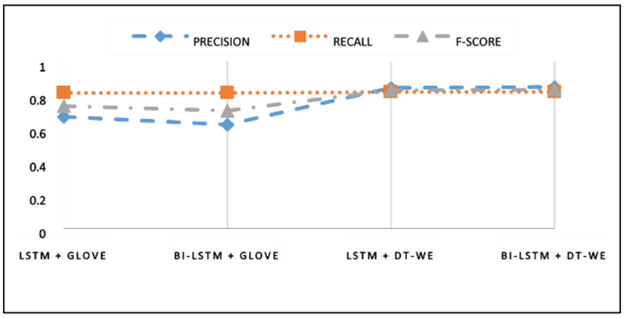

Fig. 3. Performance of LSTM architecture trained by various GloVe and Dt-WE word embedding models using R:15 dataset.

1) Comparing with Previous Methods

Three major methods were previously proposed for the extraction of the implicit aspects, co-occurrence matrixes, classification methods, and hand-crafted rules. Our model is compared with several methods that are summarized as follows:

Hai et al., [30] originally presented **CoAR**, that is later being replicated in [16]. It is a co-occurrence matrix between explicit aspects and opinion words. They first generated set of rules to synthesis the relation between the explicit aspects and the opinion words, then a co-occurrence matrix being used to find the relation between the explicit aspects and the opinion words based on a threshold value between them. Similarly, **CRSA** [31] is being replicated in [16], that is a frequency-based method relied on the co-occurrence frequency in the sentences.

CBA [32] stands for a Classification-based Approach being implemented in [16]. They have grouped the aspects that co-occur with the same opinion word into same category, than a classifier is implemented to assign the sentence into one aspect-opinion in the test stage.

Similar to **CoAR**, Sun, et al., [33]presented a co-occurrence matrix named **NCBA** between the explicit aspects and the opinion words to find the implicit aspects based on the similarity threshold between them. Xu, et al., [16] re-constructed **NCBA** model for the extraction of the implicit aspects in R:15 dataset.

Yan et al., [34] introduced a PageRank algorithm named **EXPRS** for the extraction of implicit aspects based on a dependency relation between the aspect-opinion pairs and the selected candidate aspects. The selected candidate aspects are chosen to be expanded with synonyms using the PageRank, and the candidate with highest rank are the implicit aspects as in [16].

The presented method in [35] contains two distributions of opinion words: the context distribution that is derived by co-occurrence matrix and topic distribution of opinion words is learnt through Latent Dirichlet Allocation (**LDA**) topic model. This approach called **CWC** replicated in [16] for implicit aspect extraction.

N-NM [16] is a non-negative matrix factorization is used for the implicit aspect extraction based on the explicit aspects and opinion words extraction. The synonymous explicit aspects were first grouped into categories, then the opinion words in the sentences are mapped into the semantically relevant category based on similarity measure.

CNN [36] is a deep learning approach used to tag each word in an opinionated sentence as either aspect or non-aspect word. Then a set of manually crafted patterns are combined into the deep learning approach for the for the extraction of implicit aspects.

CRF [37] stands for the conditional random field classifier combined with word features (i.e. PoS tags, dependency relation) for implicit aspect extraction.

Poria et al., [5] proposed hand-crafted rules that are formulated via a dependency parser between the aspect-opinion pairs. The implicit aspects extracted using an implicit aspect clues and the opinion words that have no explicit aspects related to them.

TABLE IV. BASELINE METHODS FOR IMPLICIT ASPECT EXTRACTION ON THE R:14 DATASET COMPARED TO OURS.

Model	Precision	Recall	F-score
CNN + Rule-based	0.8827	0.861	0.8815
CRF	0.8535	0.8272	0.8401
(Poria et al., 2014)	0.8521	0.8815	0.8665
Ours	**0.8059**	**0.8059**	**0.8059**

TABLE V. BASELINE METHODS FOR IMPLICIT ASPECT EXTRACTION ON THE R:15 DATASET COMPARED TO OURS.

Model	Precision	Recall	F-score
CoAR	0.7571	0.4741	0.5831
CRSA	0.7689	0.9854	0.8638
CBA	0.7605	0.9919	0.861
NCBA	0.7443	0.6828	0.7122
EXPRS	0.7331	0.4623	0.567
CWC	0.7407	0.2913	0.4181
N-NM	0.7661	0.9887	0.8633
Ours	**0.8474**	**0.8186**	**0.8327**

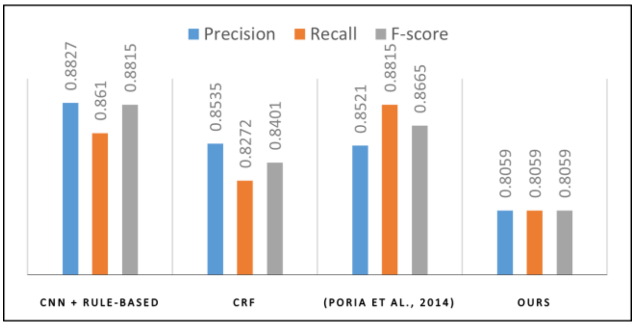

Fig. 4. Comparison of the baseline methods to our model "Bi-LSTM + Dt-WE" on the R:14 Dataset.

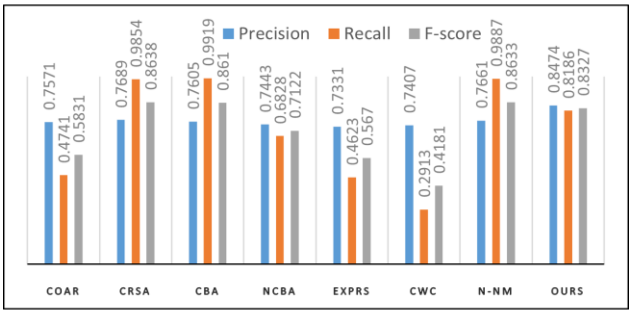

Fig. 5. Comparison of the baseline methods to our model "Bi-LSTM + Dt-WE" on the R:15 Dataset.

In contrary to the previously proposed methods, our proposed model outperforms several methods for the implicit aspect extraction on the R:14, and R:15 datasets as shown in Table IV,V. Therefore, Fig. 4 and 5 visualize the performance of the proposed model compared to the baseline methods,

The previously proposed models mostly supervised machine learning methods. For instance, **CNN** is a supervised neural network that achieved the highest accuracy in terms of aspects extraction. However, our model outperforms the two other methods which **CRF** and Rule-based [5] that are also supervised machine learning methods. Besides, for R:15 dataset, our model outperforms most of the presented methods for the extraction in terms of Precision score, yet **CBA** presents the highest value in terms of Recall metric as it is meant for single implicit aspect extraction.

V. CONCLUSION

In the breadth of massively generated sentimental reviews across all public platforms about any particular products or services, the need for a multi-label neural network model capable of extracting implicit aspect terms automatically without necessitating a handcrafted rules cannot be denied. To this end, a domain-trained word embeddings is proposed to be interpolated into a multi-label neural network model is proposed for the extraction of multiple implicit aspects. Nevertheless, the proposed methodology for implicit aspect extraction is built based on a supervised neural network (i.e. Bi-LSTM) which suffer from a limitation, that make it not sufficient for unannotated text data. Because the problem of aspect extraction relies on a data that is raw, that have no class label. So, the future direction of this research is focusing on using unsupervised machine learning like non-parametric topic model.

REFERENCES

[1] T. A. Rana and Y.-N. Cheah, "A two-fold rule-based model for aspect extraction," *Expert Syst Appl*, vol. 89, pp. 273–285, 2017.

[2] S. Poria, E. Cambria, and A. Gelbukh, "Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-level Multimodal Sentiment Analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, no. September, pp. 2539–2544.

[3] O. M. AL-Janabi, N. H. A. H. Malim, and Y.-N. Cheah, "Unsupervised model for aspect categorization and implicit aspect extraction," *Knowl Inf Syst*, Apr. 2022, doi: 10.1007/s10115-022-01678-5.

[4] O. M. Al-Janabi, N. H. A. H. Malim, and Y.-N. Cheah, *Aspect Categorization Using Domain-Trained Word Embedding and Topic Modelling*, vol. 619. 2020. doi: 10.1007/978-981-15-1289-6_18.

[5] S. Poria, E. Cambria, L.-W. Ku, C. Gui, and A. Gelbukh, "A Rule-Based Approach to Aspect Extraction from Product Reviews," in

second workshop on natural language processing for social media (SocialNLP), 2015, pp. 28–37. doi: 10.3115/v1/w14-5905.

[6] R. Panchendrarajan, N. Ahamed, B. Murugaiah, P. Sivakumar, S. Ranathunga, and A. Pemasiri, "Implicit Aspect Detection in Restaurant Reviews," in *Proceedings of NAACL-HLT 2016*, 2016, pp. 128–136.

[7] J. Feng, S. Cai, and X. Ma, "Enhanced sentiment labeling and implicit aspect identification by integration of deep convolution neural network and sequential algorithm," *Cluster Comput*, vol. 22, no. 3, pp. 5839–5857, 2019, doi: 10.1007/s10586-017-1626-5.

[8] P. Soujanya, N. Ofek, A. Gelbukh, A. Hussain, and R. Lior, "Dependency Tree-Based Rules for Concept-Level Aspect-Based Sentiment Analysis," *Communications in Computer and Information Science*, vol. 475, pp. 41–47, 2014, doi: 10.1007/978-3-319-12024-9.

[9] R. Benkhelifa, N. Bouhyaoui, and F. Z. Laallam, "A real-time aspect-based sentiment analysis system of youtube cooking recipes," in *Hassanien A. (eds) Machine Learning Paradigms: Theory and Application. Studies in Computational Intelligence*, vol. 801, Springer Cham, 2019. doi: 10.1007/978-3-030-02357-7_11.

[10] N. Dosoula *et al.*, "Sentiment analysis of multiple implicit features per sentence in consumer review data," *Frontiers in Artificial Intelligence and Applications*, vol. 291, pp. 241–254, 2016, doi: 10.3233/978-1-61499-714-6-241.

[11] M. Tubishat, N. Idris, and M. A. M. Abushariah, "Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges," *Inf Process Manag*, pp. 545–563, 2018, doi: 10.1016/j.ipm.2018.03.008.

[12] T. A. Rana and Y.-N. Cheah, "Aspect extraction in sentiment analysis: comparative analysis and survey," *Artif Intell Rev*, vol. 46, no. 4, pp. 459–483, 2016.

[13] O. M. Al-Janabi, N. H. A. H. Malim, and Y. N. Cheah, "Analogy-based Common-Sense Knowledge for Opinion-Target Identification and Aggregation," in *2018 2nd International Conference on Imaging, Signal Processing and Communication, ICISPC 2018*, Jul. 2018, pp. 1–5. doi: 10.1109/ICISPC44900.2018.9006687.

[14] M. E. Mowlaei, M. Saniee Abadeh, and H. Keshavarz, "Aspect-based sentiment analysis using adaptive aspect-based lexicons," *Expert Syst Appl*, vol. 148, p. 113234, 2020, doi: 10.1016/j.eswa.2020.113234.

[15] H. El Hannach and M. Benkhalifa, "Using synonym and definition wordnet semantic relations for implicit aspect identification in sentiment analysis," in *ACM International Conference Proceeding Series*, 2019, vol. Part F1481, pp. 1–5. doi: 10.1145/3320326.3320406.

[16] Q. Xu, L. Zhu, T. Dai, L. Guo, and S. Cao, "Non-negative matrix factorization for implicit aspect identification," *J Ambient Intell Humaniz Comput*, vol. 11, no. 7, pp. 2683–2699, 2020, doi: 10.1007/s12652-019-01328-9.

[17] T. Gaillat, B. Stearns, G. Sridhar, R. McDermott, M. Zarrouk, and B. Davis, "Implicit and Explicit Aspect Extraction in Financial Microblogs," in *Proceedings of the First Workshop on Economics and Natural Language Processing*, 2019, pp. 55–61. doi: 10.18653/v1/w18-3108.

[18] P. Ray and A. Chakrabarti, "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis," *Applied Computing and Informatics*, no. xxxx, 2019, doi: 10.1016/j.aci.2019.02.002.

[19] F. A. Vargas and T. A. S. Pardo, "Aspect Clustering Methods for Sentiment Analysis," in *International conference on computational processing of the Portuguese language.*, 2018, pp. 365–374. doi: 10.1007/978-3-319-99722-3_37.

[20] M. Afzaal, M. Usman, A. C. M. Fong, and S. Fong, "Multiaspect-based opinion classification model for tourist reviews," *Expert Syst*, vol. 36, no. 2, p. e12371, 2019, doi: 10.1111/exsy.12371.

[21] S. Areed, O. Alqaryouti, B. Siyam, and K. Shaalan, "Aspect-Based Sentiment Analysis for Arabic Government Reviews," in *Studies in Computational Intelligence*, vol. 874, Springer, 2020, pp. 143–162. doi: 10.1007/978-3-030-34614-0_8.

[22] O. M. Alyasiri, Y. Cheah, A. K. Abasi, and O. M. Al-Janabi, "Wrapper and Hybrid Feature Selection Methods Using Metaheuristic Algorithms for English Text Classification: A Systematic Review," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3165814.

[23] A. Elnagar, R. Al-Debsi, and O. Einea, "Arabic text classification using deep learning models," *Inf Process Manag*, vol. 57, p. 102121, 2020, doi: 10.1016/j.ipm.2019.102121.

[24] B. Al-Salemi, M. Ayob, G. Kendall, and S. A. M. Noah, "Multi-label Arabic text categorization: A benchmark and baseline comparison of multi-label learning algorithms," *Inf Process Manag*, vol. 56, no. 1, pp. 212–227, 2019, doi: 10.1016/j.ipm.2018.09.008.

[25] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Adv Neural Inf Process Syst*, pp. 1–9, 2013.

[26] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/NECO.1997.9.8.1735.

[27] A. S. Shamsaldin, P. Fattah, T. A. Rashid, and N. K. Al-Salihi, "A Study of The Convolutional Neural Networks Applications," *UKH Journal of Science and Engineering*, vol. 3, no. 2, pp. 31–40, Dec. 2019, doi: 10.25079/UKHJSE.V3N2Y2019.PP31-40.

[28] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in *International Workshop on SemanticEvaluation*, 2015, pp. 27–35. doi: 10.3115/v1/s14-2004.

[29] 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)*, 2015, pp. 486–495. doi: 10.18653/v1/s15-2082.

[30] Z. Hai, K. Chang, and J. J. Kim, "Implicit feature identification via co-occurrence association rule mining," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6608 LNCS, no. PART 1, pp. 393–404. doi: 10.1007/978-3-642-19400-9_31.

[31] K. Schouten and F. Frasincar, "Finding implicit features in consumer reviews for sentiment analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer*, 2014, pp. 130–144.

[32] L. Zeng and F. Li, "A Classification-Based Approach for Implicit Feature Identification," in *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, 2013, pp. 190–202. doi: 10.1007/978-3-642-41491-6_18.

[33] L. Sun, S. Li, J. Li, and J. Lv, "A novel context-based implicit feature extracting method," in *DSAA 2014 - Proceedings of the 2014 IEEE International Conference on Data Science and Advanced Analytics*, 2015, pp. 420–424. doi: 10.1109/DSAA.2014.7058106.

[34] Z. Yan, M. Xing, D. Zhang, and B. Ma, "EXPRS: An extended pagerank method for product feature extraction from online consumer reviews," *Information and Management*, vol. 52, no. 7, pp. 850–858, 2015, doi: 10.1016/j.im.2015.02.002.

[35] J. Chen, L. Sun, Y. Peng, and Y. Huang, "Context weight considered for implicit feature extracting," in *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, (DSAA)*, 2015, pp. 0–4. doi: 10.1109/DSAA.2015.7344860.

[36] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowl Based Syst*, vol. 108, pp. 42–49, 2016, doi: 10.1016/j.knosys.2016.06.009.

[37] Z. Toh and W. Wang, "DLIREC: Aspect Term Extraction and Term Polarity Classification System," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, pp. 235–240. doi: 10.3115/v1/s14-2038.