

TopFuzz4SA: An Integrated Fuzzy Neural Network with Topic-aware Auto-encoding for Sentiment Analysis

Tham Vo (✉ thamvth@tdmu.edu.vn)

Thu Dau Mot University <https://orcid.org/0000-0001-7291-4168>

Research Article

Keywords: sentiment analysis, topic modeling, seq2seq, fuzzy neural network

Posted Date: July 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-623864/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Soft Computing on January 7th, 2022. See the published version at <https://doi.org/10.1007/s00500-021-06520-8>.

TopFuzz4SA: an integrated fuzzy neural network with topic-aware auto-encoding for sentiment analysis

Tham Vo^{1,*}

¹Thu Dau Mot University, Binh Duong, Vietnam

¹thamvth@tdmu.edu.vn

(*) *Corresponding author*

ABSTRACT

Recent advanced deep learning architectures, such as neural seq2seq, transformer, etc. have demonstrated remarkable improvements in multi-typed sentiment classification tasks. Even though recent transformer-based and seq2seq-based models have successfully enabled to capture rich-contextual information of texts, they are still lacking of attention on incorporating the global semantic information, such as topic, in order to sufficiently leverage the performance of downstream SA task. Moreover, emotional expressions of users are normally in forms of natural human-written textual data which might consist a lot of noise and ambiguity which impose great challenges on the processes of textual representation learning as well as sentiment polarity prediction. To meet these challenges, we propose a novel integrated fuzzy-neural architecture with a topic-driven textual representation learning approach for handling SA task, called as: TopFuzz4SA. Specifically, in the proposed TopFuzz4SA model, we first apply a topic-driven neural encoder-decoder architecture with the incorporation of latent topic embedding and attention mechanism to sufficiently learn both rich contextual and global semantic information of the given textual data. Then, the achieved rich semantic representations of texts are fed into a fused deep fuzzy neural network to effectively reduce the feature ambiguity and noise, forming the final textual representations for sentiment classification task. Extensive experiments in benchmark datasets demonstrate the effectiveness of our proposed TopFuzz4SA model in comparing with contemporary state-of-the-art baselines.

Keywords: sentiment analysis; topic modeling; seq2seq; fuzzy neural network.

1. INTRODUCTION

Normally considering as an important task of natural language processing (NLP) domain, sentiment analysis (SA) [1] [2] aims to automatically analyze the underlying opinions/emotions towards specific entities, e.g., news, products, services, etc. Recently, thanks to the tremendous development of Internet, there are a huge number of people who are frequently actively involving in multiple activities in multi-typed digital platforms, such as: social networks, ecommerce platforms, etc. As a result, a large amount of data is generated every day which contains emotional expressions, opinions, attitudes etc. of people about entities that they have been interacted with. These emotional expressions can be considered as valuable resources for analyzing and supporting companies/organizations can deeply gain the insights of their provided products or services. Thus, sentiment analysis a.k.a. opinion mining is considered as an important task of NLP due to its primitive applications in multiple areas. To identify the sentiment polarity of a user toward a specific product/service, multiple machine learning (ML) based techniques have applied to formulate, model and characterize the underlying emotional aspects from the raw data, normally in form of texts (e.g., comments, reviews, micro-blogs, etc.). In fact, textual data in form of natural language is considered as a common source for expressing opinions, emotions or feelings of users upon their interacted entities in online platforms. From the past, there are multiples researches which formulate the SA task as the text classification problem in which hand-crafted textual representation learning techniques (e.g. BOW and its family) and ML-based classification algorithms (SVM, Logistic Regression, etc.) have been applied to predict the sentiment polarity from texts. Specifically, a SA task is designed as the text analysis and

binary classification model to categorize the given user's comments/reviews into the positive or negative emotional states. However, these traditional techniques encountered limitations related to ambiguity and sparsity of short textual data to effectively exploit the sentimental aspects. Recently, with the emergence of deep learning which has shown promising performances in multiple domains of computer science, including the NLP. The utilization of neural network architecture in text analysis has dramatically alleviate the efforts of hand-crafted feature engineering for the process of text representation learning. There are advanced deep neural architectures, such as: recurrent neural network (GRU, LSTM, Bi-LSTM, etc.) [3] [4] [5] and convolutional neural network (CNN) [6] have been utilized to deeply learn and characterize the sentimental aspects from the given texts to handle SA task. There are complex deep neural-network based approaches [3] [4] [6] [5] have demonstrated state-of-the-art performances in multiple downstream sub-tasks of SA (aspect-level, sentence-level, document-level, multiple domains, etc.). However, previous deep learning based model still have drawbacks related to the capability of capturing rich semantic and contextual information of texts in order to better fine-tune for multiple downstream SA sub-tasks. Moreover, the application of deep neural architectures in the processes of sequential textual representation learning and SA-oriented neural training objective might severely suffer extracted feature ambiguity and noise which can dramatically reduce the accuracy performance of sentimental polarity prediction.

1.1. Recent progress & existing challenges

To deal with recent challenges of rich semantic representation learning, several advanced neural sequence-to-sequence (seq2seq)/auto-encoding based architectures [7] [8] [9] with attention mechanism have been proposed to effectively capture the sequential representation of texts. Among significant achievements in NLP, attention is considered as the most important framework to facilitate multiple RNN-based textual representation learning models. For sentiment classification task, there are multiple integrated attention with LSTM/Bi-LSTM approaches, such as well-known works of Bi-LSTM+CRF [10], Sentic-LSTM [11], CDSC [12], etc. have been utilized to address the challenge of multiple emotional aspect learning objective for SA task. In fact, the neural attention framework has become an effective mechanism for most of proposed SA models in recently. However, contemporary integrated attention with RNN-based architectures are still insufficient to capture the diversity in categories/attributes of texts (a.k.a. sentimental aspects) which might implicitly express different user's emotional expressions toward a specific entity. Moreover, the application of attention-based mechanism in SA task to for the system to fully concentrate on sentimental aspects is also considered ineffective and time-consuming.

Recently, with the appearance of pre-trained language models a.k.a. transformer (e.g., ELMo [13], GPT-2 [14], BERT [15], etc.), multiple NLP tasks, including sentiment classification task, have been significantly improved in both efficiency and accuracy performance aspects. These pre-trained language architectures can effectively support to capture the rich syntactic and contextual information in the circumstance of large-scale text corpora of specific language which have been carefully and well trained, previously. For SA task, the utilization of pre-trained language model such as the well-known BERT can sufficiently characterize sentimental features from the input texts and achieve the state-of-the-art accuracy performance on various downstream sub-tasks of SA, such as recent works in BERT4ABSA [16], ABSA-BERT-pair [17], SentiLARE [18], etc. These pre-trained BERT based model have demonstrated significant improvements on the accuracy performance of sentiment polarity identification within the situation of context/aspect-varied textual representation learning. However, recent pre-trained BERT based SA model still suffered limitations, mostly in the capability of achieving the global semantic information of texts to better fine-tune for the SA task as well as the elimination of feature noise and ambiguity from the achieved text representations. In fact, most of recent pre-trained BERT based opinion mining models such as BERT4ABSA [14] and SentiLARE [16] mostly focused on enriching the captured latent features of semantic and syntactic relationships between words in a given text corpus to correctly identify the emotional aspects to predict the sentiment polarity rather than focus on the global semantic information such as topics/latent semantic structures of texts. Moreover, due to the ambiguity of emotional aspects in forms of natural language, the learnt sentimental aspect-aware representations of texts, which are achieved by using

pre-trained language models, normally contain a lot of noises and uncertainties in the extracted latent features. Thus, it might lead to the significant downgrades in the accuracy performance of sentiment polarity prediction task in the after all. Recently, there are proposals [17] of using the fuzzy neural learning concept for eliminating the feature noise and ambiguity from the learnt data representation. Such as the recent proposal of using fuzzy neural CNN model (FCNN) [18] to learn the high-quality representations of texts to leverage the performance of sentiment analysis task. However, the previous proposed FCNN-based SA model [18] still lacked of thorough evaluations on the global and sequential semantic information of texts which is considered as unable to deal with aspect-based sentiment classification problem.

1.2. Motivations & our contributions

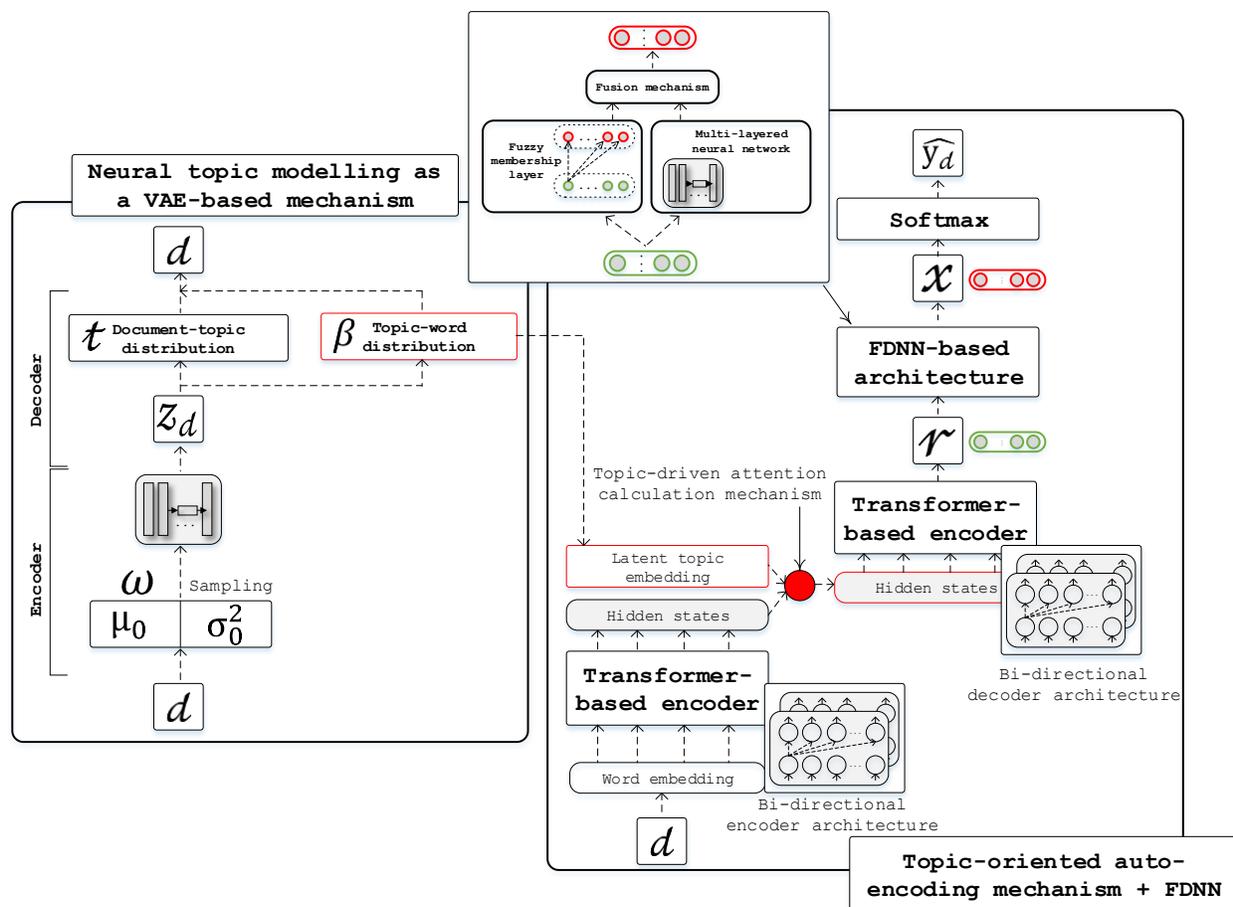


Figure 1. Illustrations of our proposed TopFuzz4SA model in this paper

To meet above mentioned challenges, in this paper we proposed an integrated fused fuzzy deep neural network with a topic-driven transformer-based encoder-decoder architecture to handle multiple downstream tasks of SA, called TopFuzz4SA (as illustrated in Figure 1). First of all, to effectively learn the latent representations of distributed latent topics in a given text corpus, we applied the neural topic modelling architecture as a variational auto-encoding mechanism which is mainly inherited from previous works [19] [20] to capture all global semantic information of texts. Then the learnt topic representations are used to facilitate the self-attention mechanism of the given neural encoder-decoder architecture. In our TopFuzz4SA model, for dealing with challenges related to the context-varied and sentimental aspect-diversified understanding from a given text corpus, a deep neural transformer-based encoder-decoder architecture is applied to comprehensively capture complicated syntactical and sequential features from the input texts. Moreover, the utilization of an integrated topic-aware attention mechanism with the

transformer-based network also can help to sufficiently capture both global salient latent topic and rich contextual relationships between words in order to enrich the quality of textual representations. Then, the achieved rich semantic representations of input texts are fed into a fused fuzzy deep neural network (FDNN) to produce the final high-quality representations of sentences/documents for dealing with the sentiment classification problem. Motivated by the use of fuzzy learning concept in the data uncertainty and noise reduction of previous works [17] [18], we applied a combined fuzzy and deep neural architecture with a fusion mechanism to learn and merge the achieved deep textual representations in previous steps and pass them through a full-connected layer with the softmax classification function to proceed sentiment polarity prediction. In general, our contributions in this paper can be summarized as threefold, which are:

- First of all, we apply a neural topic modelling architecture with an auto-encoding mechanism [20] to efficiently learn the representation of the latent distributed topics over input texts. The learnt latent topic embedding vectors are used to facilitate the attention mechanism of our designed transformer-based encoder-decoder architecture to effectively extract topic-oriented global and sequential semantics of texts.
- Secondly, the previous achieved topic-oriented rich semantic representations of input texts are then fed into a FDNN-based architecture to significantly alleviate feature noise and ambiguity before feeding to a full-connected layer to handle SA task. The proposed FDNN architecture in our paper is designed as two separated components, which are the fuzzy learning based and deep learning neural components. These two components are utilized to simultaneously capture latent features of input embedding vectors from both fuzzy and deep learning aspects. Then, the learnt latent features of these two components are fused together by using a fusion mechanism to produce the final representations of input texts.
- Finally, we conducted extensive experiments in benchmark SA-based datasets to demonstrate the effectiveness of our proposed ideas in this paper for multiple downstream sentiment classification tasks. The experimental outputs show the outperformances of our proposed TopFuzz4SA in comparing with recent state-of-the-art baselines in SA task.

In overall, the left contents of our paper are organized into 4 sections. In the second section, we briefly review about recent works in SA task and discuss about pros/cons of these works. Next, we formally present the methodology and implementation of our proposed TopFuzz4SA model in the third section. In the fourth section, we conduct extensive experiments in benchmark datasets to demonstrate the effectiveness of our proposed model in comparing with recent state-of-the-art baselines. In the last section, we conclude our works and highlight some potential improvements for the future works.

2. RELATED WORKS

Recently, most of proposed SA models adopted advanced deep learning architecture such as RNN and CNN to effectively handle the rich semantic representation learning and sentiment polarity prediction problems from texts. There are RNN-based SA models which utilize the LSTM/Bi-LSTM to dynamically characterize and extract sentimental features from the input documents by leveraging the capability of sequentially encoding and transforming words into fixed dimensional embedding vectors without the interactions of hand-crafted feature engineering techniques. Such as the proposed IAN model [21] of Ma, D. et al. which applied an interactive attention-based dual LSTM architecture to efficiently model input sentences and existing sentimental aspects for handling aspect-based sentiment analysis task. Similar to that, Rao, G. et al. with SR-LSTM [4] applied an two-layered LSTM based network to capture the semantic relationships between sentences to deal with the length-varied document-level sentiment analysis problem. With the integration of attention-based mechanism, RNN-based models which are for aspect-based SA task can effectively learn and identify sentimental categories/attributes from input sentences/documents, such as recent proposed models: Bi-LSTM+CRF [8], Sentic-LSTM [9] and CDSC [10]. Moreover, beside the only use of RNN, researchers also tried to combine various deep learning architecture to find better ways for sufficiently capture context-varied and aspect-varied representations of texts to handle multiple

downstream SA sub-tasks, such as the integrations of LSTM+CNN [3] and Bi-LSTM+CNN [1] [22]. However, these contemporary deep learning based SA models still encountered limitations on the ability of capturing rich contextual information of texts to thoroughly characterize the sentimental aspects.

In recent years, with the emergence of pre-trained transformer-based language models, such as: GPT-2 [12], BERT [13], etc. have demonstrated remarkable performances in multiple primitive tasks of NLP domain. These pre-trained models are used as the latent feature extractors to achieve the rich contextual information from texts by well-designed and trained upon large-scale text corpora. Taking advantages of pre-trained BERT model, several transformer-based architectures for SA task have been proposed, such as: BERT4ABSA [14], ABSA-BERT-pair [15] and SentiLARE [16]. Among BERT-based SA model, SentiLARE [16] by Ke, P. et al. recently has shown state-of-the-art improvements in multiple downstream sub-tasks of SA domain by integrating the pre-trained BERT model with the lexical linguistic knowledge to enrich the textual representations of texts for leveraging the performance of SA task. However, recent pre-trained language model for SA task still suffered drawbacks related to the capability of capturing the global semantic information such as topics of input texts in order to better fine-tune for handling context-varied and topic-diversified sentiment classification task. Moreover, recent deep RNN/transformer based models for SA still suffered problems related to the feature noises and ambiguities in the learnt representations of input texts which might lead to the downgrades in the overall accuracy performance of the sentiment polarity prediction process. Different from recent approaches, our works in this paper majorly focus on the utilization of fuzzy neural learning concept to reduce the uncertainty and noise from learnt textual representations which are obtained by using a topic-aware neural encoder-decoder architecture. Our designed transformer-based auto-encoding mechanism can sufficiently capture both rich global and contextual semantics of texts by integrating with the topic-aware attention mechanism which are facilitated by previous neural topic modelling approach [19] [20].

3. METHODOLOGY & IMPLEMENTATION

In this section, we formally present the methodology and detailed descriptions on the implementation of our proposed TopFuzz4SA model. In the first section, we introduce the use of neural topic modelling in learning the representations of distributed latent topics up the given text corpus. Then, the achieved topic representations are used to facilitate the topic-driven attention mechanism of a neural auto-encoding mechanism to fully capture the global and contextual semantic information of the input documents. Then, these rich semantic textual representations are fed into a fused fuzzy neural network (FDNN) architecture to reduce the feature noise and ambiguity in order to improve the performance of multi-task sentiment classification which are taken in charge by a full-connected neural layer with the softmax classification function.

3.1. Topic-oriented neural auto-encoding for sentimental analysis

3.1.1. Neural topic model encoding

To efficiently extract and learn the representations of distributed latent topics upon a given text corpus, we apply a neural variational auto-encoding (VAE) mechanism which is majorly inherited from previous works [19] [20] with the involvement of analyzed (z) latent topics from the input documents. Within the topic modeling paradigms, each document (d) with a set of words, as: $d = \{w_1, w_2, \dots, w_{|d|}\}$, with: $\mathcal{W}, w \in \mathcal{W}$ is the vocabulary set of all unique words in a given text corpus, has a corresponding distributed topic proportion, denoted as: $z_d \in \mathbb{R}^{1 \times K}$ with (K) is the predefined number of latent topics. On the other hand, for each latent topic, we have the t_n as the topic assignment for an observed words, denoted as: w_n . In our applied neural topic modelling approach, we used two separated neural network architectures which are: generative network (is played as the encoder) and inference network (is played as the decoder). For the generative network, it is used to encode the input texts into the latent topic representations. Then, these latent topic representations from the generative encoder are reconstructed to the original input texts at the inference/decoder network. The ultimate purpose of using the VAE-based architecture in this case is used

to learn and parameterize the multinomial probabilistic distributions of latent topics in which the predefined distributions is not required to efficiently guide the topic generative process. In general, the generative process for each input document (d) can be formulated as the following (as shown in equation 1):

$$\begin{aligned}
\omega &\sim \mathcal{N}(\mu_0, \sigma_0^2) \\
z_d &= \text{softmax}(W_\omega \omega + b_\omega) \\
t_n &\sim \text{Mult}(z_d) \\
w_n &\sim \text{Mult}(\beta_{t_n})
\end{aligned} \tag{1}$$

Following previous works [19] [20] [23], we apply a diagonal Gaussian distribution to parameterize the topic distribution of each document and achieve the variable distribution with the unbiased gradient estimator. In the equation 1, the μ_0 and σ_0^2 present for the mean and variance of the Gaussian distribution, W_ω and b_ω are the trainable weighting and bias parameter matrices. In overall, the loss function of a given VAE-based neural topic modelling architecture is formulated as: $\mathcal{L}_{\text{VAE-NTM}} = D_{\text{KL}}(q(z_d)||p(z_d|w)) - E_{q(z_d)}[p(w|z_d)]$, with $p(z_d|w)$, $p(w|z_d)$ and $q(z_d)$ are the corresponding probabilistic distributions for the generative/inference networks, and standard Normal prior: $N(0, I)$, respectively. After this process, we obtain the latent topic-word distributions, as: $\beta \in \mathbb{R}^{K \times |W|}$.

3.1.2. Topic-oriented transformer-based encoder-decoder mechanism

From the latent topic embedding matrix, as: β which is achieved in previous steps, we use these distributed latent topic embedding vectors to calculate the attention weights, denoted as: (a) of the given topic-oriented attention mechanism as the following (as shown in equation 2):

$$\begin{aligned}
Z &= \text{softmax}(\text{Linear}(\beta)) \\
\mathcal{P} &= \text{Linear}(\beta) + \text{Norm}(Z) \\
a &= \mathcal{H}^{\text{att}} \cdot \mathcal{P}^T
\end{aligned} \tag{2}$$

The \mathcal{P} is the transformation component of the given latent topic embedding matrix (β), as: $\mathcal{P} \in \mathbb{R}^{K \times h}$ with (h) is the dimensionality of the RNN-based hidden state vector in a given neural encoder-decoder architecture. Then, for each i^{th} word, as: (w_i), in the input document (d), the average attention weight over K latent topics are identified as: $\alpha_i = \frac{1}{K} \sum_{i=1}^K a_{t,i}$. Finally, the average attention weight for each i^{th} word (w_i) is normalized with a softmax function, denoted as: $\hat{\alpha}_i = \text{softmax}(\alpha_i) = \frac{\exp(\alpha_i)}{\sum_{j=1}^{|\mathcal{d}|} \exp(\alpha_j)}$, the final achieved

topic-oriented attention is represented as: $\alpha = \{\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_{|\mathcal{d}|}\}$. Then, the final topic-oriented attention scores are used to facilitate the neural encoder-decoder framework for handling sentiment classification task. In the encoder part, a transformer-based architecture is applied to learn and transform the input word embedding vector of document (d), as: $\{e_1^w, e_2^w, \dots, e_{|\mathcal{d}|}^w\}$ into the latent contextual representations through (k) multi-layered transformer-based layers. Specifically, for each l^{th} layer, the previous input is passed to through and generate a corresponding hidden states, denoted as: $\mathcal{H}^{\text{enc},[l]} = \{\mathcal{H}_1^{\text{enc},[l]}, \mathcal{H}_2^{\text{enc},[l]}, \dots, \mathcal{H}_{|\mathcal{d}|}^{\text{enc},[l]}\}$.

Then, the last hidden state matrix of the given encoder, denoted as: $\mathcal{H}^{\text{enc},[k]}$ is combined with the calculated topic-oriented attention scores (α), to form the final contextual representation, denoted as: (s). Finally, the encoded contextual representation vector: (s) is fed into the decoder part to reconstruct the original input as the embedding vector: (r) with a full-connected layer, as: $\text{FFN}(\cdot)$ at the end. The general process of these steps can be formulated as the following (as shown in equation 3):

$$\begin{aligned}
\mathfrak{s} &= \alpha^T \cdot \mathcal{H}^{\text{enc},[k]} \\
\mathcal{H}^{\text{dec},[0]} &= \mathfrak{s} \\
\mathcal{H}^{\text{dec},[i]} &= f_{\text{att}}(\hat{y}_{[i-1]}, \mathcal{H}^{\text{dec},[i-1]}, \mathfrak{s}) \\
\mathcal{r} &= \text{FNN}(\mathcal{H}^{\text{dec}})
\end{aligned} \tag{3}$$

In the equation 3, the $f_{\text{att}}(\cdot)$ is the calculation of the given topic-oriented attention based mechanism which is implemented as the self-attention mechanism with output word embedding ($\hat{y}_{[i-1]}$) and input contextualized representation of the encoder (\mathfrak{s}).

3.2. Integrated FDNN for feature noise and ambiguity reduction

Finally, from the achieved embedding output (\mathcal{r}) of the decoder, we feed it to a fused fuzzy deep neural architecture to alleviate the feature noise and ambiguity before applying the sentiment classification with a full-connected layer at the end. Our proposed FDNN is designed with two components, the first component is in charge of fuzzification/de-fuzzification input embedding vectors of the decoder part. This fuzzy neural learning component contain the separated multi-layered membership and fuzzy rule layers to handle the fuzzy logic representation learning. Then, the fuzzy-based learning outputs are fused with the deep learning based transformed embedding vectors to form the final representation of (\mathcal{r}). In the fuzzy learning based component, each l^{th} membership layer takes each i^{th} dimension of the vector (\mathcal{r}), denoted as: r_i as the input variable and passes it through a fuzzy-neuron, denoted as: $u_i(\cdot)$ with activation function is the Gaussian membership function to calculate the fuzzy degree, as: (σ_i) of each i^{th} input variable, as: $u_i(r_i): \mathcal{R} \rightarrow [0,1]$. Each l^{th} fuzzy membership layer of the given FDNN is generally formulated as the following (as shown in equation 4):

$$\sigma_i^{\text{fuzz},[l]} = u_i(r_i^{[l]}) = e^{-\frac{(r_i^{[l]} - \mu_i)^2}{\sigma_i^2}}, \forall i \tag{4}$$

For each fuzzy neuron, the activation function is defined as the Gaussian membership function with μ and σ^2 are the mean and variance, respectively. Then, these output fuzzy degrees are fed into the fuzzy rule layer and perform the ‘‘AND’’ logic operation, as: $\sigma_i^{\text{fuzz},[l]} = \prod_j \sigma_j^{\text{fuzz},[l-1]}$, $\forall j \in \Omega_i$ with Ω_i is the set of all input output nodes in the $(l - 1)^{\text{th}}$ layer which are connected to the input variable (i). In the deep neural learning component which is also designed as a multi-layered neural architecture in which each input vector (\mathcal{r}) is passed through different full-connected layer with the sigmoid activation function, in order to learn and exploit the high-level textual representations of previous neural encoder-decoder architecture. In general, each l^{th} neural layer of the given deep neural learning component is formulated as: $\sigma^{\text{deep},[l]} = \sigma(W^{\text{deep},[l]} \cdot \mathcal{r} + b^{\text{deep},[l]})$, with W^{deep} and b^{deep} are the trainable weighting and bias parameter matrices of each full-connected neural layer. Finally, to fuse different representations of both fuzzy and deep learning based mechanism in two components, we applied a neural network-based fusion mechanism to effectively merge type-varied representations of (\mathcal{r}) into a unified embedding, denoted as: (x) which is later used for handling sentiment classification task. The neural fusion mechanism is defined as the following (as shown in equation 5):

$$x = W^{\text{fuse},\text{fuzz}} \cdot \sigma^{\text{fuzz}} + W^{\text{fuse},\text{deep}} \cdot \sigma^{\text{deep}} + b^{\text{fuse}} \tag{5}$$

This neural fusion mechanism has a set of trainable parameters, as: $\Theta^{\text{fuse}} = \{W^{\text{fuse},\text{fuzz}}, W^{\text{fuse},\text{deep}}, b^{\text{fuse}}\}$ which are simultaneously optimized with the previous encoder-decoder architecture. Finally, the feature noise-reduced representations of each input document (d), as: x_d is feed into a full-connected layer with

the softmax activation function to handle the sentiment polarity prediction task with the training objective is formulated as the cross entropy loss, as the following (as shown in the equation 6):

$$\begin{aligned} \widehat{y}_d &= \text{softmax}(x) \\ \mathcal{L}_{\text{TopFuzz4SE}} &= -\frac{1}{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} [y_d \cdot \log(\widehat{y}_d) + (1 - y_d) \cdot \log(1 - \widehat{y}_d)] \end{aligned} \quad (6)$$

In this equation, the \mathcal{T} , y_d and \widehat{y}_d are the training set, vector-based encoded data of ground-truth and predicted sentiment polarity of the given input document (d), respectively. To train our proposed architecture, we apply the stochastic gradient descent (SGD) to optimize all model’s parameters upon the training objective of $\mathcal{L}_{\text{TopFuzz4SE}}$, as: $\Theta^{\text{TopFuzz4SE}} = \text{argmin}_{\Theta = \{\Theta^{\text{fuse}}, \Theta^{\text{enc-dec}}\}} (\mathcal{L}_{\text{TopFuzz4SE}}, \eta)$ with η is the predefined learning rate. In general, our proposed TopFuzz4SE is motivated by the success of previous achievements of neural topic modelling, transformer-based encoder-decoder architecture in rich-contextual text representation learning and utilization of fuzzy learning concept in feature noise and ambiguity reduction in order to leverage the performance of sentiment classification task.

4. EXPERIMENTS & DISCUSSIONS

In this section, we conduct extensive experiments in benchmark datasets (SST, AR, MR, IMDb, Yelp and SemEval2014) to demonstrate the effectiveness of our proposed TopFuzz4SA model in comparing with recent state-of-the-art baselines, including: SR-LSTM [4], Sentic-LSTM [9], general BERT [13], BERT4ABSA [14], ABSA-BERT-pair [15] and SentiLARE [16] for both sentence-level and aspect-level sentiment analysis tasks.

4.1. Experimental settings & datasets

4.1.1. Dataset descriptions

For our experiments in this paper, we used different benchmark datasets which are commonly applied to evaluate the SA model in previous works, which are:

- **SST (Stanford Sentiment Treebank)**^[1]: is considered as a common dataset for evaluating the performance of SA model. This dataset contains 11K labelled sentences/documents as movie reviews which are collected from the <https://www.rottentomatoes.com/>. The SST dataset for SA task includes 8,544 for training, 1,101 for testing and 2,210 for validation with five sentiment-level labels for classification.
- **AR (Amazon Reviews)**^[2]: is a large-scale dataset for SA task which contains > 34M user’s reviews and ratings on specific products (2.4M) in different categories which are collected from the Amazon e-commerce platform. For experiments in this dataset, we randomly selected 500K, 50K and 50K reviews for training, testing and validation sets, respectively.
- **MR (Movie Reviews)**^[3]: is a traditional dataset for SA task with 10K short/single-sentence based reviews of users on specific movies. These reviews are labelled as positive (5K) and negative (5K) classes. The MR dataset is considered as less challenging than other datasets in which the SA task is considered as a classical binary classification task. For this dataset, we divide it into three parts: training (8,534), testing (1,078) and validation (1,050).
- **Yelp (Yelp-5)**^[4]: is a well-known dataset for multiple disciplines including SA task. The Yelp-5 dataset belongs to the Yelp Challenge Dataset collection which contains > 1.6M reviews of 366K users upon 61K local businesses/companies. The Yelp-5 reviews are categorized into 5 classes (1-

¹ SST: <https://nlp.stanford.edu/sentiment/index.html>

² AR dataset: <http://snap.stanford.edu/data/web-Amazon-links.html>

³ MR: <http://www.cs.cornell.edu/people/pabo/movie-review-data/>

⁴ Yelp: <https://www.yelp.com/dataset>

5). Similar to previous works [16], we randomly divided this dataset into 3 parts: training (594,000), testing (56,000) and validation (50,000).

- **IMDb**⁵: is similar to the MR dataset, this dataset contains 50K user’s reviews upon specific movies which are categorized as positive and negative attitudes. For this dataset, we also applied the same split of Ke, P. et al. in [16], as randomly selected 22,500 for training, 25K for validation and 2.5K for testing.
- **SemEval-2014** [24]⁶: is a common dataset for aspect-level sentiment analysis task. For the sentimental aspect-based analysis problem with this dataset, we selected the “*laptop*” (contains 3,045 for training and 800 for testing) and “*restaurant*” (contains 3,041 for training and 800 for testing) categories for evaluating performance or implemented SA models.

Dataset pre-processing steps & configurations. For basic textual pre-processing steps, such as stop-word removal, word tokenization, stemming, etc. we mainly used the Stanford CoreNLP library⁷ [25] to handle textual data in each dataset. For setup of general BERT [13] which are used for SA task, we reused the general pre-trained BERT model (large/uncased version) which is released by Google at this repository⁸. For the setup of SentiLARE [16], we used the SentiWordNet 3.0 [26]⁹ which is similar to the original implementation of Ke, P. et al. [16].

4.1.2. Experimental setups & evaluation method usage

For the setup of TopFuzz4SA model, we implemented it by using Python programming language with the support of Tensorflow machine learning library. Our TopFuzz4SA model and other SA comparative baselines are deployed to a single server with the Intel Xeon SKL-SP 4210 CPU and 64Gb in memory. For the detailed configurations of our proposed TopFuzz4SA model, we configured the number latent topic for the neural topic modelling architecture (described in section 3.1.1) is 10 or $K = 10$. The general dimensional size of word embedding vector for our transformer-based encoder-decoder architecture (described in section 3.1.2), $d^w = 300$ and dimensionality of hidden state vector (or number of used RNN-based cells) of each transformer-based layer in the given topic-oriented auto-encoding architecture, as: $h^{RNN} = 256$. For the number of layers which are used in both transformer-based encoder and decoder parts, we configured them as $k^{enc-dec} = 7$. For the number of fuzzy-based and deep learning based layers in the FDNN architecture (described in section 3.2), we set them as: $k^{FDNN} = 5$. Table 1 lists other configurations of our TopFuzz4SA model which are implemented for experiments in this paper. Similar to previous works [14] [15] [16], to evaluate the accuracy performance of different models for SA tasks as a primitive classification task, we mainly applied the Accuracy and F-1 evaluation metrics.

Table 1. Configurations for our proposed TopFuzz4SA model for experiments

Model’s parameter	Value
Number of latent topic for the neural topic modelling (described in section 3.1.1), (K)	10
Dimensionality of word embedding vector in the topic-oriented auto-encoding mechanism (described in section 3.1.2), (d^w)	300
Dimensionality of hidden state vector in the topic-oriented auto-encoding mechanism, (h^{RNN})	256
Number of transformer-based layers for encoder and decoder in topic-oriented auto-encoding mechanism, ($k^{enc-dec}$)	7

⁵ IMDb: <https://www.imdb.com/interfaces/>

⁶ SemEval2014 (Task 4) dataset: <https://alt.qcri.org/semeval2014/task4/>

⁷ Stanford CoreNLP: <https://stanfordnlp.github.io/CoreNLP/>

⁸ Pre-trained BERT (large, uncased): <https://github.com/google-research/bert>

⁹ SentiWordNet 3.0: <https://github.com/aesuli/SentiWordNet>

Number of FDNN-based layers for both fuzzy learning and deep learning based components, (k^{FDNN})	5
Number of training epochs	60
Learning rate (η)	0.0003
Model's optimizer	Adam (weight decay = 0.005)
General dropout rate	0.5

4.1.3. Comparative baselines

To compare the accuracy performance of our proposed TopFuzz4SA model with other baselines for both sentence-level and aspect-level SA tasks, we implemented different well-known SA models, which are:

- **SR-LSTM** [4]: is a dual LSTM based architecture which supports to effectively learn the sequential representations of texts for dealing with document-level sentiment classification problem. In this SR-LSTM model, Rao, G. et al. [4] proposed a two-layered LSTM neural network to jointly learn the semantic sequential representations and latent relationship features between sentences of the input documents. As the sentence/document-level SA model, SR-LSTM is unable to deal with the aspect-based SA task.
- **Sentic-LSTM** [9]: is a recent well-known RNN-based approach for handling aspect-level SA task. In this model, Ma, Y. et al. proposed a novel stacked attention mechanism with different sentence-level and aspect-level representation learning strategies within an LSTM-based architecture. Through experiments, the Sentic-LSTM model demonstrated remarkable improvement in aspect-based SA task within benchmark datasets.
- **BERT** [13]: is the general BERT model which can be fine-tuned for different NLP's tasks, proposed by Devlin, J. et al. The BERT recently is considered as the most well-known transformer architecture which can be utilized for different task-driven training objectives. There are several pre-trained BERT versions which have been trained in large-scale textual corpora for different languages. For our experiments in this paper, we reused the pre-trained BERT (large/uncase) version which is officially released by Google and fine-tuned for handling sentiment classification task in different datasets. As a rich-contextual text representation learning framework, BERT can be capable for handling aspect-level SA task.
- **BERT4ABSA** [14]: is recently proposed by Xu, H. et al. for sentiment-driven review reading comprehension and multiple-level aspect-based SA tasks. The BERT4ABSA is a modified version of the original BERT with custom sentimental aspect-labelling mechanism to effectively fine-tune for multiple downstream sub-tasks of SA.
- **ABSA-BERT-pair** [15]: similar to the BERT4ABSA model, in the ABSA-BERT-pair model, Sun, C. et al. proposed a modified BERT version with the facilitation of auxiliary information from sentence-pairs to efficiently fine-tuned for aspect-level SA task. With different implementations, the ABSA-BERT-pair can be adopted for both sentence/document-level and aspect-level SA tasks.
- **SentiLARE** [16]: is considered as the recent BERT-based approach for SA task which is also our main competitor in this paper. Recently proposed by Ke, P. et al. [16], the SentiLARE model utilizes the external resource such as SentiWordNet as the extra linguistic knowledge for assisting the rich contextual representation learning mechanism of pre-trained BERT model for handling multiple downstream sub-tasks in SA. Different from previous BERT-based SA models, like as: BERT4ABSA and ABSA-BERT-pair, the SentiLARE model is considered as more powerful to properly model and capture sentimental aspects of input documents due to the capability of integrating with external linguistic knowledge.

For the configurations of listed above comparative baselines, we kept the same configurations as described in their original works in which these models achieved the highest performances in different downstream SA tasks. For common configurations which are similar to our proposed TopFuzz4SA model, we set them as the same values as listed in Table 1.

4.2. Experimental results & discussions

In this section, we present experimental outputs of different SA models for both sentence/document-level and aspect-level SA tasks in benchmark datasets.

4.2.1. Experiments on sentence/document-based SA task

For sentence/document-level sentiment classification task, all models are implemented to learn from the training set and predict the sentiment polarity of documents in testing set as the classical text classification problem. Table 2 show the experimental results in terms of F-1 evaluation metric for different SA baselines in standard datasets which demonstrate the outperformances of our proposed TopFuzz4SA model in comparing with recent state-of-the-art baselines.

Table 2. Experimental outputs for sentence/document-level SA task with different baselines in terms of F-1 evaluation metrics in benchmark datasets

	SST	AR	MR	Yelp-5	IMDb
SR-LSTM	0.56021	0.63812	0.81821	0.58921	0.87921
Sentic-LSTM	0.57812	0.62013	0.86082	0.62812	0.91692
BERT	0.56213	0.60921	0.84021	0.63089	0.89213
BERT4ABSA	0.59892	0.63092	0.87321	0.65982	0.95921
ABSA-BERT-pair	0.59821	0.62012	0.88921	0.68721	0.96821
SentiLARE	0.60921	0.69092	0.91081	0.69821	0.97181
TopFuzz4SA	0.61231	0.72921	0.93021	0.71023	0.98921

In general, as shown from the experimental outputs, we can recognize that most of transformer-based SA models (BERT, BERT4ABSA, ABSA-BERT-pair, SentiLARE and our proposed TopFuzz4SA) achieved better performance than previous LSTM-based model (SR-LSTM and Sentic-LSTM) about 6.48% in all datasets. It proves the fact that the use of transformer-based textual representation learning technique can support to obtain richer contextual information from texts for leveraging the performance of SA task in comparing with previous RNN-based models. Specifically, in comparing with previous LSTM-based models, our proposed TopFuzz4SA model significantly improves the accuracy performance in term F-1 metric about 13.95% (SR-LSTM) and 10.18% (Sentic-LSTM) for all benchmark datasets. Similar to that with BERT-based SA models, our proposed TopFuzz4SA model also slightly gains better results in the sentence/document-level SA tasks about: 12.35% (general BERT), 6.69% (BERT4ABSA) and 5.53% (ABSA-BERT-pair). For our main competitor in this paper, the SentiLARE model, our proposed TopFuzz4SA model also outperforms approximately 2.32% in the accuracy performance of sentence/document-level SA task in all datasets.

4.2.2. Experiments on aspect-based SA task

In this section, we conducted extensive experiments to compare the performance of BERT-based models, including: BERT, BERT4ABSA, ABSA-BERT-pair, SentiLARE and our proposed TopFuzz4SA on the aspect-level sentiment classification task. Similar to previous empirical studies of Ke, P. et al. in the SentiLARE [16] model, all models are evaluated the accuracy performance in terms of Accuracy and F-1 metrics in handling the aspect-based SA task at two levels: aspect term extraction and aspect term sentiment classification. Table 3 and Table 4 show the experimental outputs for aspect-based SA task in aspect term extraction and aspect term sentiment classification levels by using different models in the standard SemEval-2014 dataset.

Table 3. Experimental on aspect-based SA task (aspect term extraction) with different BERT-based baselines in terms of Accuracy and F-1 evaluation metrics

Model	Laptop				Restaurant			
	Accuracy		F-1		Acc		F-1	
BERT	0.91082		0.78921		0.89023		0.84821	
BERT4ABSA	0.94092	↑0.033	0.80821	↑0.024	0.92082	↑0.034	0.86092	↑0.015
ABSA-BERT-pair	0.93912	↑0.031	0.79821	↑0.011	0.92821	↑0.042	0.89278	↑0.052
SentiLARE	0.94892	↑0.041	0.82733	↑0.048	0.95911	↑0.077	0.90821	↑0.070
TopFuzz4SA	0.96271	↑0.057	0.83582	↑0.059	0.96881	↑0.088	0.91082	↑0.073

Table 4. Experimental on aspect-based SA task (aspect term sentiment classification) with different BERT-based baselines in terms of Accuracy and F-1 evaluation metrics

Model	Laptop				Restaurant			
	Accuracy		F-1		Acc		F-1	
BERT	0.82931		0.72901		0.82892		0.73921	
BERT4ABSA	0.84971	↑0.024	0.72911	↑0.000	0.81081	↓0.021	0.77681	↑0.050
ABSA-BERT-pair	0.84821	↑0.022	0.72912	↑0.000	0.82971	↑0.001	0.78612	↑0.063
SentiLARE	0.86721	↑0.045	0.75781	↑0.039	0.83721	↑0.010	0.80813	↑0.093
TopFuzz4SA	0.87072	↑0.049	0.78713	↑0.079	0.85892	↑0.036	0.81721	↑0.105

Experimental outputs in Table 3 and Table 4 demonstrate the outperformances of our proposed TopFuzz4SA model in comparing with recent BERT-based models in both aspect-based SA task in aspect term extraction and aspect term sentiment classification levels. In more details, for the aspect term extraction level aspect-based SA task, our proposed TopFuzz4SA model achieves better performance than BERT-based models (BERT, BERT4ABSA and ABSA-BERT-pair) averagely 4.81% and 4.86% in terms of Accuracy and F-1 metrics, respectively. The TopFuzz4SA model also outperforms our main competitor SentiLARE model about 1.2% and 0.64% in terms of Accuracy and F-1 metrics, respectively for this task. Similar to that with the aspect term sentiment classification level, our proposed TopFuzz4SA model also significantly achieves better performances approximately 3.85% and 7.23% in terms of Accuracy and F-1 metrics in comparing with BERT, BERT4ABSA and ABSA-BERT-pair. For the SentiLARE model, our model also slightly improves the performance about 1.47% and 2.45% in terms of Accuracy and F-1 evaluation metrics. To sum up, through experimental outputs in both sentence/document-level and aspect-level SA tasks, we prove the effectiveness of our proposed ideas in this paper which is a combination of topic-oriented auto-encoding mechanism with the integration of fuzzy neural representation learning for feature noise and ambiguity reduction.

4.3. Ablation studies on proposed TopFuzz4SA model

In this section, we conduct extensive empirical studies on the parameter sensitivity of our proposed TopFuzz4SA model, includes: dimensionality of word embedding vector (d^w), hidden state of the RNN-based architecture (h^{RNN}) in our topic-driven auto-encoding mechanism and the number of layers which are used for the topic-driven auto-encoding ($k^{enc-dec}$) and FDNN-based (k^{FDNN}) architectures.

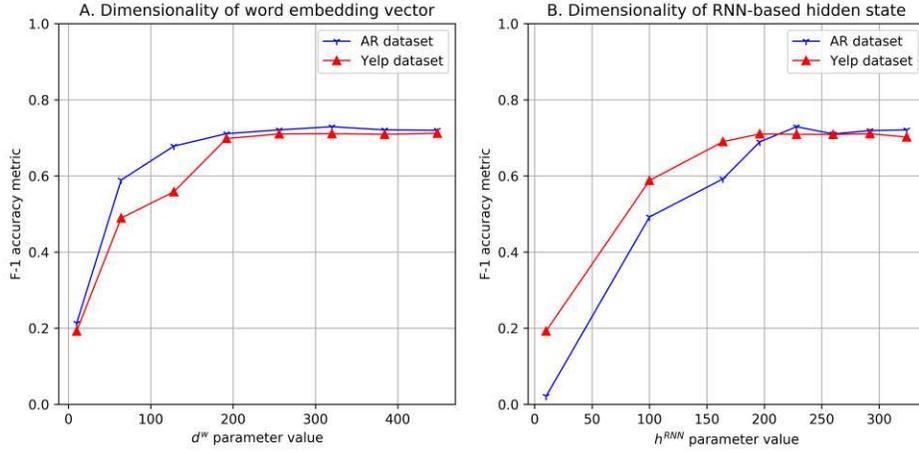


Figure 2. Experimental studies on the sensitivity of (d^w) and (h^{RNN}) parameters on the accuracy performance of our proposed TopFuzz4SA model in AR and Yelp datasets

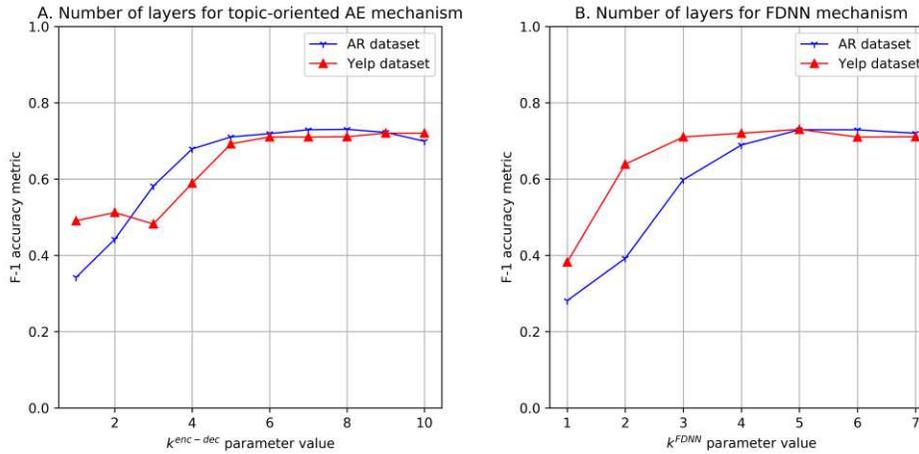


Figure 3. Experimental studies on the sensitivity of ($k^{enc-dec}$) and (k^{FDNN}) parameters on the accuracy performance of our proposed TopFuzz4SA model in AR and Yelp datasets

To do this, we varied the values of (d^w) and (h^{RNN}) parameters in range of [10, 448] and [10, 324], then reported the changes on accuracy performance of our proposed TopFuzz4SA for the sentence-level SA task in AR and Yelp datasets. Figure 2 shows the experimental outputs for studies on the influences of these two parameters upon the accuracy performance of our TopFuzz4SA model. The experimental outputs show that our model is quite insensitive with these parameters in which it reaches the high performance with value of $d^w > 200$ and $h^{RNN} > 220$. Similar to previous studies on the (d^w) and (h^{RNN}) parameters, to evaluate the effects of ($k^{enc-dec}$) and (k^{FDNN}) parameters on the overall performance of our model, we conducted extensive empirical studies on the same AR and Yelp datasets with different values of these two parameters within range [1,10] for $k^{enc-dec}$ and [1, 7] for k^{FDNN} parameters. Experimental outputs (in Figure 3) show that our proposed model reach the stability in performance with values of $k^{enc-dec} \geq 6$ and $k^{FDNN} \geq 5$ for both AR and Yelp datasets.

4.4. Studies on fuzzy vs. non-fuzzy approaches in TopFuzz4SA model

To evaluate the effectiveness of applying fuzzy learning concept on reducing the feature noise and ambiguity for learnt textual representations which effectively support for the improvement on multiple downstream SA tasks, we implemented two versions of our proposed TopFuzz4SA model. The first version

is the original implementation of TopFuzz4SA with the support of fuzzy neural learning mechanism in the FDNN-based architecture, named as: TopFuzz4SA-Fuzzy. The second version is TopFuzz4SA model without the setup of fuzzy neural layers in the FDNN-based architecture, named as: TopFuzz4SA-DL. Then, we utilized these two versions of TopFuzz4SA model to handle the sentence-based SA task in the AR and Yelp datasets with different training set size (%) and reported the accuracy outputs of each version in terms of F-1 evaluation metric.

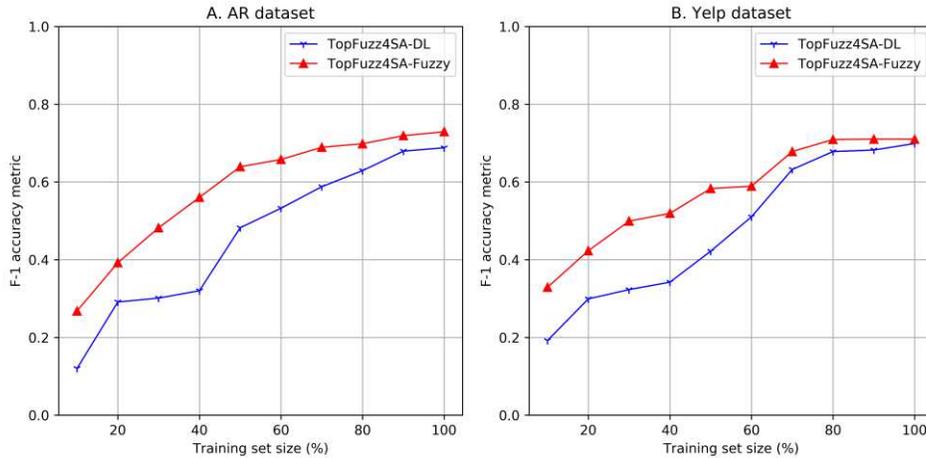


Figure 4. Experimental studies on the influence of applying fuzzy learning concept for feature noise and ambiguity reducing on our proposed TopFuzz4SA model in AR and Yelp dataset

Experimental outputs in Figure 4 show that the TopFuzz4SA-Fuzzy version remarkably achieves better performance than the TopFuzz4SA-DL version in which the accuracy performance of the fuzzy-based version stably increases with different training set size (%) and is higher than the only-deep-learning based version in the after all. This extensive empirical studies show the usefulness of applying fuzzy neural learning concept on alleviating the feature noise and ambiguity from input texts in which significantly improve the performance of several primitive SA tasks as the result.

5. CONCLUSIONS & FUTURE WORKS

In this paper, we propose a novel approach of an integrated fuzzy neural learning concept with the topic-driven auto-encoding mechanism for handling multiple downstream sentiment analysis (SA) tasks, called as TopFuzz4SA. In our proposed TopFuzz4SA model, we apply the neural topic modelling approach to model and learn the distributed latent topic over the text corpus to facilitate the topic-driven attention-based mechanism in our textual auto-encoding mechanism for SA task. Then the achieved textual representation by the topic-driven encoder-decoder architecture is fed to a fused fuzzy deep neural network (FDNN) based architecture to eliminate the feature noise and ambiguity which can effectively support to leverage the accuracy performance of sentiment classification task in the after all. Extensive experiments in benchmark datasets demonstrate the effectiveness of our proposed TopFuzz4SA model in comparing with baselines for SA task. For our future work, we intend to extend our proposed TopFuzz4SA model to handle the dynamic sentiment polarity classification task upon the real-time textual chats or QA based conversions.

DECLARATIONS

This study was funded by Thu Dau Mot University, Binh Duong, Vietnam.

ACKNOWLEDGEMENT

This research is funded by Thu Dau Mot University, Binh Duong, Vietnam.

CONFLICT OF INTEREST

This research is funded by Thu Dau Mot University, Binh Duong, Vietnam.

REFERENCES

- [1] Zhang, L., Wang, S., & Liu, B., "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1253, 2018.
- [2] Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A., "Deep learning for aspect-based sentiment analysis: a comparative review," *Expert Systems with Applications*, vol. 118, pp. 272-299, 2019.
- [3] Chen, T., Xu, R., He, Y., & Wang, X., "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications*, vol. 72, pp. 221-230, 2017.
- [4] Cheng, J., Zhao, S., Zhang, J., King, I., Zhang, X., & Wang, H., "Aspect-level sentiment classification with HEAT (hierarchical attention) network," in *Proceedings of the 2017 ACM on conference on information and knowledge management*, 2017.
- [5] Rao, G., Huang, W., Feng, Z., & Cong, Q., "LSTM with sentence representations for document-level sentiment classification," *Neurocomputing*, vol. 308, pp. 49-57, 2018.
- [6] Huang, Q., Chen, R., Zheng, X., & Dong, Z., "Deep sentiment representation based on CNN and LSTM," in *2017 International Conference on Green Informatics (ICGI)*, 2017.
- [7] Sutskever, I., Vinyals, O., & Le, Q. V., "Sequence to Sequence Learning with Neural Networks," in *Advances in Neural Information Processing Systems*, 2014.
- [8] Bahdanau, D., Cho, K., & Bengio, Y., "Neural machine translation by jointly learning to align and translate," *3rd International Conference on Learning Representations, ICLR*, 2015.
- [9] Vaswani, Ashish, et al., "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.
- [10] Al-Smadi, M., Talafha, B., Al-Ayyoub, M., & Jararweh, Y., "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 8, pp. 2163-2175, 2019.
- [11] Ma, Y., Peng, H., Khan, T., Cambria, E., & Hussain, A., "Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis," *Cognitive Computation*, vol. 10, no. 4, pp. 639-650, 2018.
- [12] Manshu, T., & Bing, W., "Adding prior knowledge in hierarchical attention neural network for cross domain sentiment classification," *IEEE Access*, vol. 7, pp. 32578-32588, 2019.
- [13] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L., "Deep contextualized word representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018.
- [14] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I., "Improving language understanding by generative pre-training," *OpenAI*, 2018.
- [15] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019.
- [16] Xu, H., Liu, B., Shu, L., & Yu, P., "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*, 2019.

- [17] Sun, C., Huang, L., & Qiu, X., "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*, 2019.
- [18] Ke, P., Ji, H., Liu, S., Zhu, X., & Huang, M., "SentiLARE: Sentiment-Aware Language Representation Learning with Linguistic Knowledge," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.
- [19] Deng, Y., Ren, Z., Kong, Y., Bao, F., & Dai, Q., "A hierarchical fused fuzzy deep neural network for data classification," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 1006-1012, 2016.
- [20] Nguyen, T. L., Kavuri, S., & Lee, M., "A fuzzy convolutional neural network for text sentiment analysis," *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 6, pp. 6025-6034, 2018.
- [21] Miao, Y., Grefenstette, E., & Blunsom, P., "Discovering discrete latent topics with neural variational inference," in *International Conference on Machine Learning (PMLR)*, 2017.
- [22] Srivastava, A., & Sutton, C., "Autoencoding variational inference for topic models," in *5th International Conference on Learning Representations (ICLR)*, 2017.
- [23] Ma, D., Li, S., Zhang, X., & Wang, H., "Interactive attention networks for aspect-level sentiment classification," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017.
- [24] Yoon, J., & Kim, H., "Multi-channel lexicon integrated CNN-BiLSTM models for sentiment analysis," in *Proceedings of the 29th conference on computational linguistics and speech processing (ROCLING 2017)*, 2017.
- [25] Zheng, C., Zhang, K., Wang, H. J., & Fan, L., "Topic-Aware Abstractive Text Summarization," in *arXiv preprint arXiv:2010.10323*, 2020.
- [26] Pontiki, Maria, et al., "SemEval 2014 Task 4 - Aspect Based Sentiment Analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation*, 2014.
- [27] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D., "The Stanford CoreNLP natural language processing toolkit," in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014.
- [28] Baccianella, S., Esuli, A., & Sebastiani, F., "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 2010.