Int J Syst Assur Eng Manag (June 2021) 12(3):542–552
https://doi.org/10.1007/s13198-021-01113-y

ORIGINAL ARTICLE

Universally domain adaptive algorithm for sentiment classification using transfer learning approach

B. Vamshi Krishna[1] · Ajeet Kumar Pandey[2] · A. P. Siva Kumar[3]

Received: 20 July 2019 / Revised: 16 April 2021 / Accepted: 20 April 2021 / Published online: 8 May 2021

Abstract Huge amount of unstructured data is posted on the cloud from various sources for the purpose of feedback and reviews. These review needs require classification for many a reasons and sentiment classification is one of them. Sentiment classification of these reviews quite difficult as they are arriving from many sources. A robust classifier is needed to deal with different data distributions. Traditional supervised machine learning approaches not works well as they require retraining when domain is changed. Deep learning techniques perform well to handle these situations, but they are more data hungry and computationally expensive.

Transfer learning is a feature in the cross-domain sentiment classification where features are transferred from one domain to another without any training. Moreover, transfer learning allows the domains, tasks, and distributions used in training and testing to be different. Therefor transfer learning mechanism is required to transfer the sentiment features across the domains.

This paper presents a transfer learning approach using pretrained language model, ELMO which helps in transferring sentiment features across domains. This model has been tested on text reviews posted on twitter data set and compared with deep learning methods with and without pretraining process, also our model delivers promising results. This model permits flexibility to plug and play parameters with target models with easier domain adaptivity and transfer sentiment features. Also, model enables sentiment classifiers by using the transferred features from an already trained domain and reuse the sentiment features by saving the time and training cost.

Keywords Sentiment classification · Domain adaptation · Deep learning · Transfer learning · Word embeddings

1 Introduction

As the usage of online opinions, recommendations, ratings and obtaining feedback is playing a major role in the decision making process. Also, volume of online textual reviews generated in social media websites and blogs is growing exponentially. Reviews can source from different domains and it is difficult to prepare labeled training data and there will be no related data to train. The cost of labeling training data for a larger number of domains involves huge cost. Also, it is difficult to design a robust classifier to deal with different data distributions covering from different domains.

Deep learning techniques can be used to extract high level features from online reviews in an unsupervised fashion. Large amounts of unlabeled data across all domains to learn the intermediate representations. Neural network based sentiment classification models are used to learn low-dimensional text features without any feature engineering.

Deep learning models perform well in domain adaptation process using neural networks for sentiment classification, but they are more data hungry and computationally expensive. Also, lot of parameters need to set and it is hard to train a network that generalizes well with limited data.

✉ B. Vamshi Krishna
vamshikrishna.bs@gmail.com

1 Sapiens Technologies Pvt Ltd., Bangalore, India

2 Bombardier Transportation Ltd., Hyderabad, India

3 JNTUA University, Ananthapur, India

This is the reason for need of transfer learning to be researched to transfer features from one model to another without or minimal training process.

Transfer learning can be defined as, given a source domain D_S and learning task T_S, a target domain D_T and learning task T_T, transfer learning aims to help improve the learning of the target predictive function $f_T(.)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$ or $T_S \neq T_T$ Pan and Yang (2010).

Transfer learning is a feature in the cross domain sentiment classification in which the features are transferred from an existing model to a new model without starting training from scratch. training and testing data are selected from different domains. Transfer learning allows the domains, tasks, and distributions used in training and testing to be different.

This model gives flexibility for sentiment classifiers in domain adaptation by transferring features on labeled reviews from one source and deployed on another domain.

2 Related works

2.1 Basis for domain adaptation solutions

Structural corresponding algorithm was proposed for domain adaptation and a measure of domain similarity was identified by selecting pivot features which links source and target domains based on their common frequency and mutual information by using source labels Blitzer et al. (2007).

Spectral feature alignment algorithm was proposed by Pan et al. (2010) with the help of bipartite graph by discovering the relationship between domain specific and independent words from different domains for learning meaningful text representations.

Relative adaptive bootstrapping algorithm was proposed in a domain adaptation framework for sentiment and topic extraction in a target domain without any labeled data by expanding seeds in target domain Fangtao Li et al. (2012).

A joint sentiment topic model was proposed by Yulan He et al. (2011) for extracting polarity bearing topics and words from different domains can be grouped under the same polarity bearing topics.

A Joint approach was proposed by Xia et al. (2015) for feature ensemble and sample selection by using Principal component analysis to handle label and instance adaptation.

An approach for cross domain sentiment classification was proposed by creating a thesaurus which is sensitive to the sentiment of words expressed in various domains Bollegala et al. (2011).

A novel inter corpus statistical approach was proposed by Hai et al. (2014) with the help of Domain Relevance measure and used for opinion feature extraction between different corpus.

2.2 Deep learning and neural networks for domain adaptation

Deep learning approach was used for high level feature extraction by using a stack of denoising and auto encoders from the text reviews of all the available domains Glorot et al. (2011).

Document level sentiment classification model was proposed by Zhaopeng Tu et al. (2012) using convolution kernels with the help of high impact substructures of sentences guided by a polarity lexicon.

Neural sentiment classification model was proposed by Chen et al. (2016) building a hierarchical long short-term memory (LSTM) using information of both global user and product to generate sentence and document level representation jointly.

Document level sentiment classification model was proposed by Jiacheng Xu et al. (2016) with a recurrent neural based architecture by using cached LSTM neural networks for capturing the overall semantic information in long texts.

Deep unordered model for sentence and document level of classification was proposed by Iyyer et al. (2015) on pretrained embeddings by using a deep averaging network.

Model was proposed with a combination of recursive neural tensor networks and the stanford sentiment tree bank which is a corpus for the sentiment analysis task Socher et al. (2013).

A hybrid model was proposed by Xiao and Cho (2016) for character level sentiment classification based on neural network with combination of convolution and recurrent layer.

Aspect level classification attention model was proposed by Tang et al. (2016) using deep memory network to compute the representation of a sentence with regard to an aspect.

Feature embedding model was proposed by Yang and Eisenstein (2015) for domain adaptation by using representational learning which are dense representations of individual features.

For sentence level classification task, a convolutional neural network (CNN) which is trained on top of pretrained word vectors was used Yoon Kim (2014).

Neural network model was used for the target dependent sentiment classification by relating the content word with target word in Twitter data set Tang et al. (2016).

Sentiment classification model was proposed by Dahou et al. (2016) for Arabic reviews by building CNN which is pretrained using word embeddings.

2.3 Transfer learning for domain adaptation

A regularization algorithm was used for learning word representations for various domains and for predicting words for the target domain by using a skip gram model Yang et al. (2017).

Aligned recurrent transfer algorithm was used in pretraining framework by learning cross domain word collocations for a target domain using recurrent neural network Cui et al. (2019).

A simple and effective single step auxiliary loss transfer learning approach with the help of pretraining language model was introduced by transferring its weights to classifier (Chronopoulou et al. (2019)) Chronopoulou et al. (2019).

A few shot learning technique was used by hierarchical pooling method over pretrained word embedding for text classification Pan et al. (2019).

Bidirectional encoder representations from transformers architecture was designed to pretrain deep bidirectional representations from unlabeled text by jointly conditioning on both left and right content in all layers Devlin et al. (2019).

A deep contextualized model with word embeddings from language models which are learned functions of the internal state of deep bidirectional language model Peters et al. (2018).

A deep LSTM encoder pretrained on context vectors was used for transferring knowledge in machine translation task which are used in various natural language processing (NLP) tasks McCann et al. (2017).

A semi supervised method for adding pretrained content embeddings using from bidirectional language to many NLP systems for sequence labeling tasks Peters et al. (2017).

Universal language model fine tuning (ULMFiT) was introduced for effective transfer learning which can be applied in any NLP task McCann et al. (2018).

2.4 Research motivation

As the supervised learning classifier trained on a domain must be retrained if the domain shifts. From the literature it is found neural models for sentiment analysis are used to learn low-dimensional text features without any feature engineering. This paper integrates transfer learning techniques with neural network to capture and transfer sentiment features across domains.

3 Research background

3.1 Word embedding models

Word embeddings is a kind of distributed learning for natural language words and are learned in an unsupervised manner from a large collection of text. Able to capture syntactic and sematic level information of the associated with words.

Word2vec model takes word vector as input and produce word embeddings which are mappings of the words with same sematic structure. This model is built from Continuous bag of words (CBOW) and Skip-gram models which are basically shallow neural networks.

Global vectors for word representations (Glove) is another unsupervised model for obtaining vector representations of words.

Word embedding models are pretrained on text corpus as shown in Fig. 1, and easily plug and play parameters with target models in transferring features and predicting of surrounding content words.

3.2 Pretrained language models

Pretrained models have the ability to transfer features of existing model to a new model. They are mostly unsupervised models Unsupervised learning can be used in each level of hierarchy, for representing features which are based on the features discovered at the previous level. This ability will help to perform transfer features across domains.

Embeddings from language models (ELMO) is a pretrained model trained on bi-directional LSTM with language models. ELMO word vectors trains separate language model in the both directions of text to scan the words for sentiment analysis.

Bidirectional encoder representations from transformers (BERT) is another pretrained model with additional layers with bi- directional language models and use for unlabeled text.

Universal language model fine-tuning for text Classification (ULMOFIT) is a language model that learns during pretraining on most general domain.

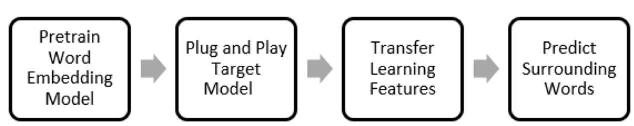

Fig. 1 Pretrained model

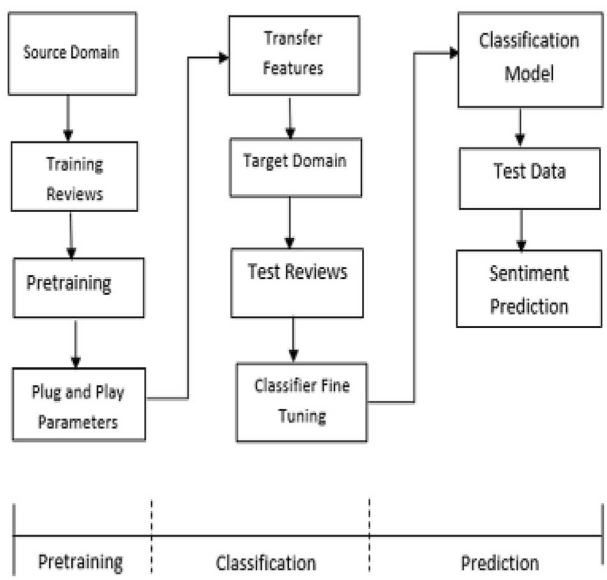

Fig. 2 Proposed model architecture

4 Proposed work

The model architecture is shown in the Fig. 2, which uses transfer learning techniques for transferring sentiment features across domains. The proposed model structure is divided into three main parts, Pretraining, Classification and Prediction. Pretraining models are used to transfer the sentiment features across various domains from one model to another easily with plug and play parameters. Classification model of new domain use these transferred features from an already trained domain and reuse the sentiment features by saving the time and training cost. Classifier is also fine-tuned for predicting the sentiment of test data from target domain.

The transfer algorithm steps are shown in the algorithm mentioned in the section IV-E, for transferring the sentiment features across domains. Pretrained models and neural network are used for sentiment classification purpose which is implemented by using Python libraries Keras and Tensorflow. Glove and ELMO models are used for pretraining and convolutional neural network and few variants of recurrent neural network are used in building the neural models. This pretrained model is scalable and can adapt test data from new domains from scratch without any training.

4.1 Data-set Used

Text reviews of twitter data set is used as input to this model. Data set is split into training, testing and validation.

Table 1 CNN-ELMO model

Model parameters	Settings
Embedding	ELMO
Embed. Size	1024
Epochs	12
Batch Size	128
Optimization	rmsprop

Table 2 CNN MODEL

Model parameters	Settings
Embedding Size	10,000
Convolution Filters	128
Epochs	12
Batch Size	128
Optimization	adam

Table 3 RNN LSTM MODEL

Model parameters	Settings
Embedding Size	10,000
LSTM	100
dropout	0.2
Epochs	12
Batch Size	128
Optimization	adam

Table 4 RNN-GRU MODEL

Model parameters	Settings
Embedding Size	10,000
GRU	100
dropout	0.2
Epochs	12
Batch Size	128
Optimization	adam

Table 5 Reviews

Total reviews	Training set	Test set	Validation set
5000	3000	1000	1000

4.2 Data Preprocessing

Twitter reviews after filtering retweets are preprocessed by removing stop words, punctations, digits and other special symbols. These cleaned text reviews are split into tokens and converted into word vectors.

Table 6 Accuracy

Model	Training accuracy	Testing accuracy
CNN	1.0	0.6406
CNN—PRETRAIN	0.9815	0.6644
LSTM	0.9403	0.6206
LSTM-PRETRAIN	0.7695	0.6278
GRU	0.8007	0.6326
GRU PRETRAIN	0.7480	0.6404
ELMO PRETRAIN	0.81	0.7790

4.3 Transfer Processing

Glove and ELMO models are used to convert word vectors into a continuous vector space word embeddings to predict sentiment of surrounding words. Then features are transferred in other domains as well without training from scratch.

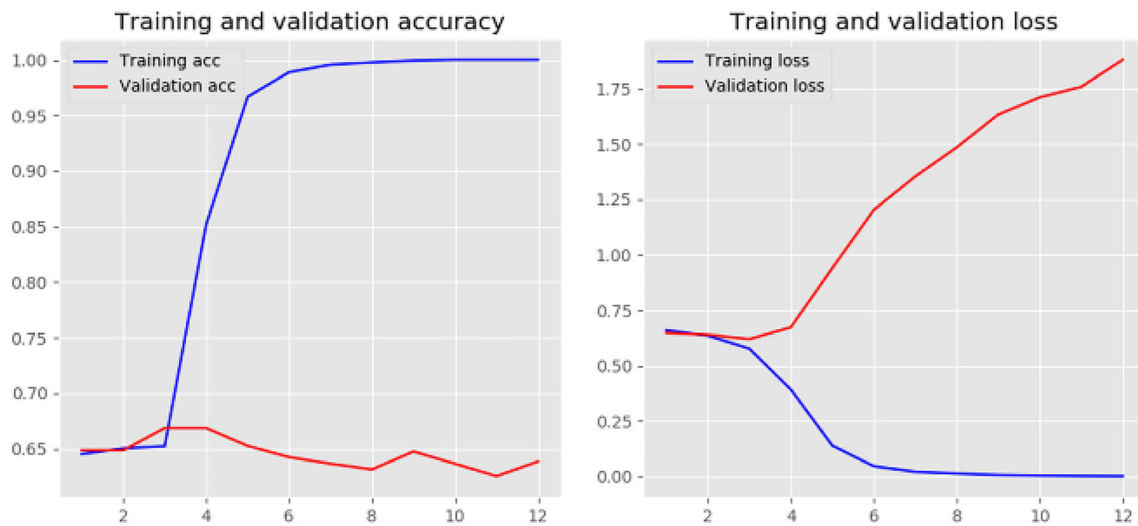

Fig. 3 CNN—Accuracy and Loss

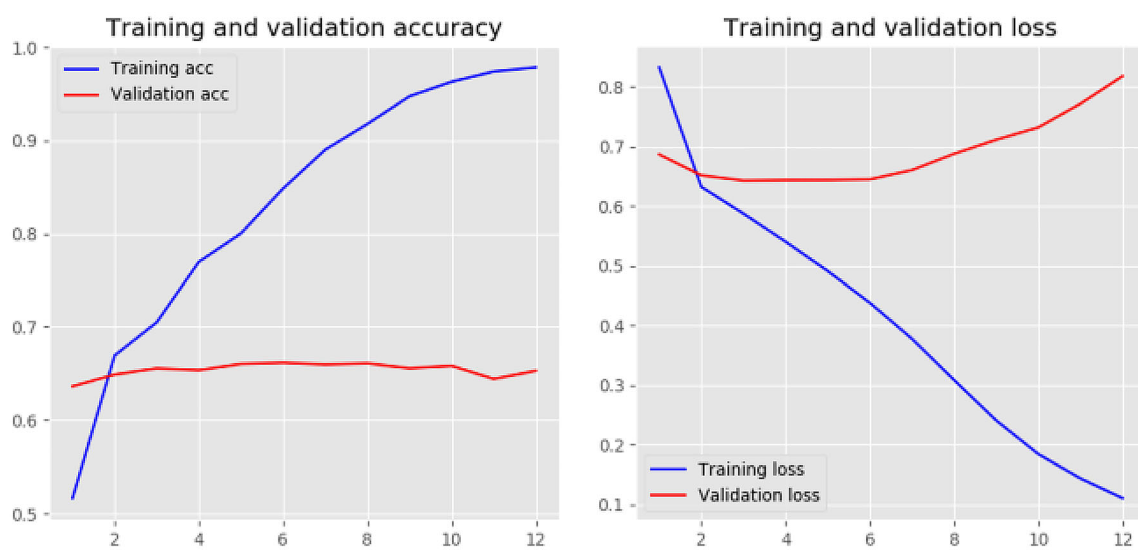

Fig. 4 CNN Pretrain—Accuracy and Loss

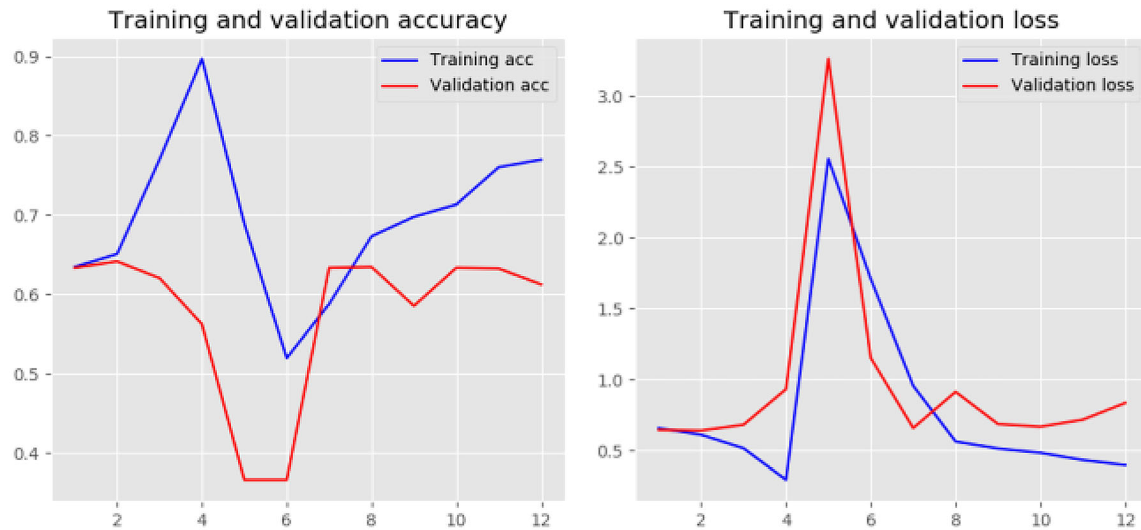

Fig. 5 RNN LSTM—Accuracy and Loss

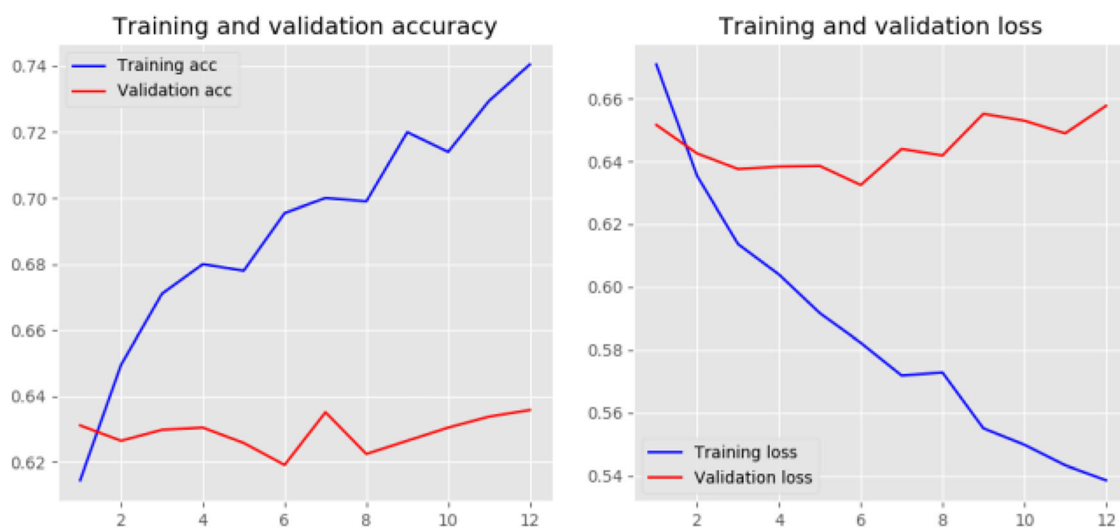

Fig. 6 RNN LSTM Pretrain- Accuracy and Loss

4.4 Pretraining model used

Glove embeddings with 27B and 100-dimensional feature vector are used to convert the text reviews into word vectors and are embedded in embedding layer of neural network. ELMO model is used for pretraining and transfer of features across dense layers. Table 1 shows the model parameters and settings used in pretraining of ELMO model.

4.5 Proposed algorithm

Below is the transfer learning procedure used for sentiment classification purpose:

Each review is represented as a sequence of vector T (x1, x2, …., xT).

Let features F (f1,f2, …., fn) are to be transferred from source domain D_S to target domain $D_{T.}$

Pretrain the source domain features using word embedding model.

Plug and play parameters for target data set of target domain D_T.

Transfer feature set F and fine tune to classification model for the target domain $D_{T.}$

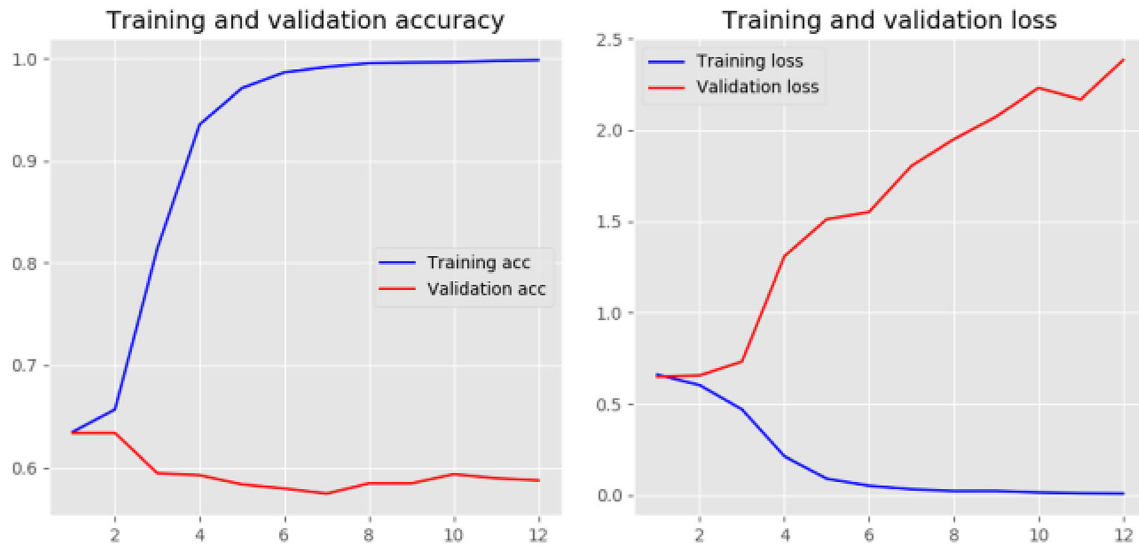

Fig. 7 RNN GRU—Accuracy and Loss

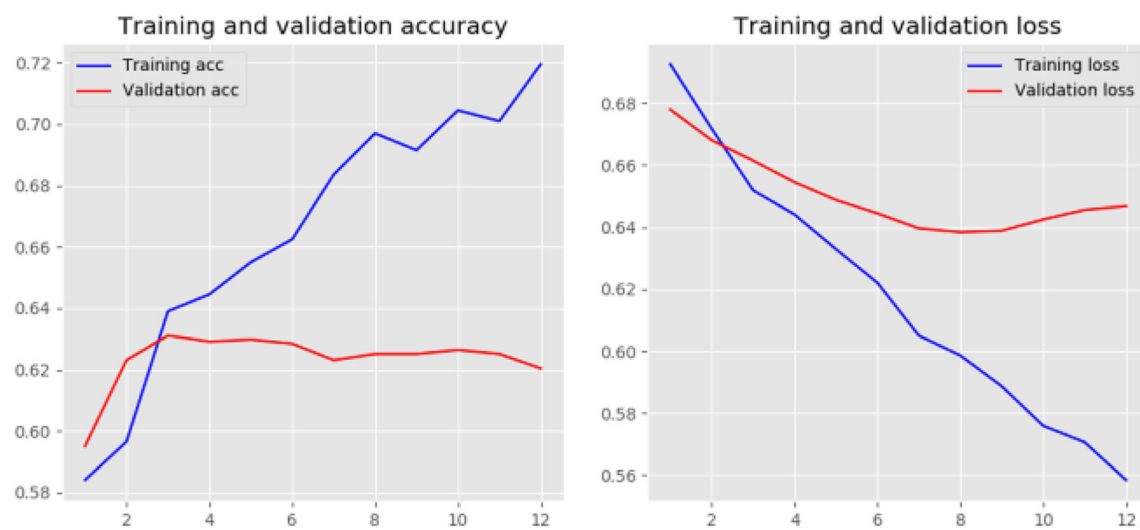

Fig. 8 RNN GRU Pretrain—Accuracy and Loss

5 Classification layer predict the sentiment categories

5.1 Model parameter and settings

Neural models are trained with Adam optimizer and binary Cross-Entropy loss function using a batch size of size 128, epoch 12.

Table 2 shows the model parameters and settings used in convolutional neural network (CNN) model. Five convolutional layers with rectified linear units as an activation function, 128 filters and max pooling size is set to 2.

Tables 3 and 4 shows the model parameters and settings used in recurrent neural network (RNN) with LSTM and GRU models respectively. 100 LSTM encoders or gated units with dropout rate and recurrent dropout as 0.2 with tanh activation function are used.

5.2 Classifier fine tuning and Sentiment prediction

Convolutional neural network model is trained on these transferred feature vectors derived as an output from pre-trained model and fine-tuned for classifying the sentiment of test data from target domain.

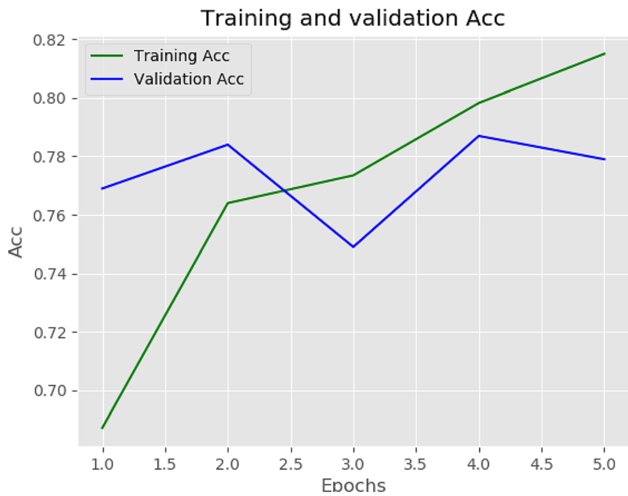

Fig. 9 ELMO Pretrain- Accuracy

CNN performs dual task of feature learning and used as sentiment classification with the transferred features in embedding layer. Dense layers are used with activation as 'softmax', loss as 'binary crossentrpy' and optimizer as 'adam'. For the outcome of sentiment, softmax classification layer is used.

6 Results and discussion

6.1 Comparison of classification accuracies

The Classification accuracies of the proposed model which use ELMO for transfer learning is compared with other existing methods CNN, RNN-LSTM, RNN-GRU with and without pretraining process.

Table 8 Reviews

Total reviews	Training set	Test set	Validation set
10,000	8000	1000	1000

Table 9 Accuracy

Model	Training accuracy	Testing accuracy
CNN	0.999	0.6817
CNN PRETRAIN	0.9815	0.6644
LSTM	0.9393	0.5185
LSTM PRETRAIN	0.8096	0.6507
GRU	0.9980	0.6036
GRU PRETRAIN	0.7752	0.6727
ELMO PRETRAIN	0.8145	0.8275

Table 5 shows the data set details used. Table 6 shows the results of prediction accuracies from different trained and pretrained models and among all of them ELMO pretrained model has high testing accuracy when compared to other existing methods.

Figure 3 and 4 shows the training and validation accuracies and loss values for CNN without pretraining and pretrained models respectively.

Figure 5 and 6 shows the training and validation accuracies and loss values for RNN-LSTM without pretraining and pretrained models respectively.

Figure 7 and 8 shows the training and validation accuracies and loss values for RNN-GRU without pretraining and pretrained models respectively.

Table 7 Sensitivity analysis

Neural model	*Training data*	*testing data*	*Training accuracy*	*Testing accuracy*
CNN	Large	Low	1.0000	0.6406
CNN	Low	Large	0.9920	0.6163
CNN	Medium	Medium	1.0000	0.6318
LSTM	Large	Low	0.7509	0.6696
LSTM	Low	Large	0.7270	0.6183
LSTM	Medium	Medium	0.7692	0.6037
GRU	Large	Low	0.7466	0.6708
GRU	Low	Large	0.7020	0.6533
GRU	Medium	Medium	0.7432	0.6269
ELMO	Large	Low	0.8007	0.7718
ELMO	Low	Large	0.7870	0.7209
ELMO	Medium	Medium	0.7980	0.7887

Above models are pretrained

Table 10 Sensitivity Analysis

Neural model	Training data	Testing data	Training accuracy	Testing accuracy
CNN	Large	Low	0.9996	0.6797
CNN	Low	Large	1.0000	0.6169
CNN	Medium	Medium	1.0000	0.5930
LSTM	Large	Low	0.8281	0.6236
LSTM	Low	Large	0.7625	0.6179
LSTM	Medium	Medium	0.7580	0.6194
GRU	Large	Low	0.7816	0.6507
GRU	Low	Large	0.7475	0.6389
GRU	Medium	Medium	0.7616	0.6170
ELMO	Large	Low	0.7762	0.7964
ELMO	Low	Large	0.7395	0.7735
ELMO	Medium	Medium	0.7718	0.8038

Above models are Pretrained

Table 11 Reviews

Total reviews	Pretraining set	Test set
6940	6248	692

Table 12 Accuracy

Metrics	Song et al. (2020)	Proposed model
Accuracy	73.35%	78.79%

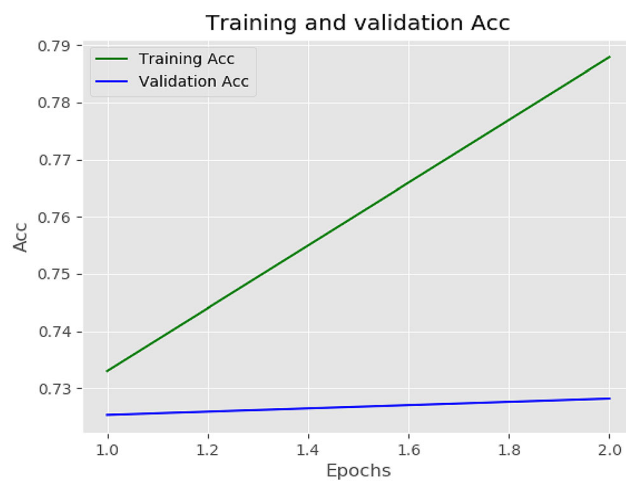

Fig. 10 ELMO Pretrain- Accuracy

Figure 9 shows the training and validation accuracies of pretrained ELMO model.

Table 5 shows the data set details used. Table 7 shows the results of sensitivity analysis when training, validation and testing data sets are considered as large, medium and low volumes of data respectively by using GLOVE and pretraining approach implemented by using models CNN, RNN-LSTM, RNN-GRU and ELMO-CNN respectively.

Table 8 shows the extended data set details used. Table 9 shows the results of prediction accuracies and Table 10 shows the results of sensitivity analysis of data when training, validation and testing data sets are considered as large, medium and low volumes of data respectively. Results are consistent even in case of increase of the data set size.

For the graphs from 1 to 8 x-axis and y-axis indicates the number of epochs and training, validation accuracies or loss values respectively for various models with and without pretraining.

6.2 Comparison with existing work

The performance of the proposed model is compared with earlier work Song et al. (2002) and found the results are promising. The data set used is twitter reviews from different domains by searching with the keywords ‘bill gates’, ‘taylor shift, ‘xbox’, ‘windows7′ and ‘google’. The dataset as shown in Table 11 is split into pretraining data of 6,248 tweets and testing data of 692 tweets and experiments are performed using proposed algorithm. In results comparison, it is observed that proposed model performs better results than the earlier model in terms of accuracy as shown in Table 12. In addition, the training and validation accuracies and loss values by the proposed classifier is shown in Fig. 10

Below are the main contributions of this paper:

1. We show that transfer learning from language models can achieve competitive results when neural models are integrating with ELMO
2. We address the problem of manual data labelling and training cost with the help of pretraining model
3. Compared transferred learning models with deep learning models and also with earlier models.

7 Conclusion

A transfer learning model for cross domain opining mining and sentiment classification has been presented in this paper. The proposed model is based on transferring features to classifier of another domain. Also, the model is useful to analyze textual reviews which are unstructured and ungrammatical in nature and enhances hidden sentiment prediction.

Model has been applied and tested on twitter data set with 10,000 recent reviews and the average training and testing accuracies are 81% and 78% respectively. Sensitivity analysis is performed with different volumes of data set by considering training, validation and testing data as large, medium and low and the results are consistent even with the extended data set. Result shows that the proposed approach using ELMO is efficient when compared to existing approaches. In future work we may move from twitter to more generic domains and apply multi task learning.

References

Blitzer J, Dredze M and Pereira F, (2007) Biographies, bollywood, boom-boxes and blenders, In: association of computational linguistics, 440–447

Bollegala D, Weir D and Carroll J (2011) Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification," In: association of computational linguistics, pages 132–141, June 2011

Chen H, Sun M, Tu C, Lin Y and Liu Z, (2016) Neural sentiment classification with user and product attention, In: association of computational linguistics, 1650–1659, Nov 2016

Alexandra Chronopoulou, Christos Baziotis, Alexandros Potamianos, Wei Wang (2019) An embarrassingly simple approach for transfer learning from pretrained language models. In: Proceedings of 2019 association for computational linguistics, p 2089–2095, Jun 2019

Wanyun Cui, Guangyu Zheng, Zhiqiang Shen, Sihang Jiang, Wei Wang (2019) Transfer learning for sequences via learning to collocate. In: Proceedings of the 2019 conference paper at ICLR, Feb 2019

Dahou A, Xiong S, Zhou J, Houcine Haddoud M, Duan P (2016) Word embeddings and convolutional neural network for arabic sentiment classification. In: COLING 2016, the 26th international conference on computational linguistics, 2418–2427, Dec 2016

Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of Oreilly AI conference, Apr 2019.

Glorot X, Bordes A and Bengio Y, (2011) Domain adaptation for large-scale sentiment classification: a deep learning approach, in proceedings of the 28th international conference on machine learning, 2011, 513–520

Hai Z, Chang K, Kim J-J, Yang CC (2014) Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance. IEEE Transactions On Knowledge And Data Engineering 26(3):623–634

He Y, Liny C and Alani H, (2011) Automatically extracting polarity-bearing topics for cross-domain sentiment classification, In: association of computational linguistics, 123–131, June 2011

Howard J, Ruder S, "Universal Language Model Fine-tuning for Text Classification," Proceedings of conference of ACL, Vol 1, page 328–339, Jul 2018.

Iyyer M, Manjunatha V, Boyd-Graber J, Daume H (2015) Deep unordered composition rivals syntactic methods for text classification. In: association of computational linguistics, 1681–1691, July 2015

Y Kim (2014) Convolutional neural networks for sentence classification. In: association of computational linguistics, 1746–1751, Oct 2014

Li F, Pan SJ, Jin O, Yang Q and Zhu X, (2012) Cross-domain co-extraction of sentiment and topic lexicons, In: association of computational linguistics, 410–419, July 2012

McCann B, Bradbury J, Xiong C, Socher R (2017) Learned in translation: contextualized word vectors. In: Proceedings of neural information processing systems, Vol 1, Aug 2017

Pan SJ, Yang Q (2010a) A Survey on Transfer Learning. IEEE Trans Knowl Data Eng 22(10):1345–1359

Pan S J, Ni X, Sun J-T, Yang Q and Chen Z, Cross-domain sentiment classification via spectral feature alignment, In: ACM 2010, 751–760

Pan C, Huang J, Gong J, Yuan X (2019) Few-shot transfer learning for text classification with lightweight word embedding based models. IEEE J 7:53296–53304

ME. Peters, Ammar W, Bhagavatula C, Power R (2017) Semi-supervised sequence tagging with bidirectional language models. In: Proceedings of conference of ACL, Apr 2017

Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. In: Proceedings of conference of acl: human language technologies, vol 1, pp 2227–2237

Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng AY, Potts C (2013) Recursive deep models for semantic compositionality over a sentiment Treebank. In: association of computational linguistics, 1631–1642, Oct 2013

Song Y, Wang J, Liang Z, Liu Z, and Jiang T, "Utilizing BERT intermediate layers for aspect based sentiment analysis and natural language inference," arXiv preprint, 2020.

Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. In: EMNLP 2016

D Tang, B Qin, X Feng, T Liu (2016) Effective LSTMs for target-dependent sentiment classification. In: COLING 2016, the 26th international conference on computational linguistics, 3298–3307, Dec 2016

Tu Z, He Y, Foster J, Genabith J, Liu Q and Lin S, (2012) Identifying high-impact sub-structures for convolution kernels in document-level sentiment classification, In: association of computational linguistics, 338–343, July 2012.

Xia R, Zong C, Hu X and Cambria E, Feature ensemble plus sample selection: domain adaptation for sentiment classification, Proceedings of the twenty-fourth international joint conference on artificial intelligence (IJCAI 2015), 4229–4233.

Xiao Y, Kyunghyun Cho (2016) Efficient character-level document classification by combining convolution and recurrent layers. In: computational language, Feb 2016

Xu J, Chen D, Qiu X and Huang X, (2016) Cached long short-term memory neural networks for document-level sentiment classification," In: association of computational linguistics, 1660–1669, Nov 2016

Yang W, Lu W, Zheng V (2017) A simple regularization-based algorithm for learning cross-domain word embeddings. In: Proceedings of the 2017 conference on empirical methods in natural language processing, 2898–2904, Sep 2017

Yi Yang, J Eisenstein (2015) Unsupervised multi-domain adaptation with feature embeddings. In: association of computational linguistics, 672–682, June 2015