

Improving Word Embedding Coverage in Less-Resourced Languages Through Multi-Linguality and Cross-Linguality: A Case Study with Aspect-Based Sentiment Analysis

MD SHAD AKHTAR, Department of Computer Science & Engineering,
Indian Institute of Technology Patna

PALAASH SAWANT, Department of Computer Science & Technology, Goa University
SUKANTA SEN, ASIF EKBAL, and PUSHPAK BHATTACHARYYA,
Department of Computer Science & Engineering, Indian Institute of Technology Patna

In the era of deep learning-based systems, efficient input representation is one of the primary requisites in solving various problems related to Natural Language Processing (NLP), data mining, text mining, and the like. Absence of adequate representation for an input introduces the problem of data sparsity, and it poses a great challenge to solve the underlying problem. The problem is more intensified with resource-poor languages due to the absence of a sufficiently large corpus required to train a word embedding model. In this work, we propose an effective method to improve the word embedding coverage in less-resourced languages by leveraging bilingual word embeddings learned from different corpora. We train and evaluate deep Long Short Term Memory (LSTM)-based architecture and show the effectiveness of the proposed approach for two aspect-level sentiment analysis tasks (i.e., aspect term extraction and sentiment classification). The neural network architecture is further assisted by hand-crafted features for prediction. We apply the proposed model in two experimental setups: multi-lingual and cross-lingual. Experimental results show the effectiveness of the proposed approach against the state-of-the-art methods.

CCS Concepts: • **Computing methodologies** → **Discourse, dialogue and pragmatics; Supervised learning;**

Additional Key Words and Phrases: Sentiment analysis, Aspect-Based Sentiment Analysis (ABSA), cross-lingual sentiment analysis, deep learning, Long Short Term Memory (LSTM), bilingual word embeddings, data sparsity, low-resourced languages, Indian languages

ACM Reference format:

Md Shad Akhtar, Palaash Sawant, Sukanta Sen, Asif Ekbal, and Pushpak Bhattacharyya. 2018. Improving Word Embedding Coverage in Less-Resourced Languages Through Multi-Linguality and Cross-Linguality: A Case Study with Aspect-Based Sentiment Analysis. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 18, 2, Article 15 (December 2018), 22 pages.

<https://doi.org/10.1145/3273931>

Palaash Sawant's work was carried out while he was intern at IIT Patna.

Authors' addresses: M. S. Akhtar, S. Sen, A. Ekbal, and P. Bhattacharyya, Department of Computer Science and Engineering, Indian Institute of Technology Patna, Bihar, India, 801106; emails: {shad.pcs15, sukanta.pcs15, asif, pb}@iitp.ac.in; P. Sawant, Department of Computer Science & Technology, Goa University, Goa, India, 403206; email: palaash77@gmail.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Association for Computing Machinery.

2375-4699/2018/12-ART15 \$15.00

<https://doi.org/10.1145/3273931>

1 INTRODUCTION

Sentiment analysis (Pang and Lee 2005; Turney 2002) is a well-established and an important field of study in Natural Language Processing (NLP). It aims to extract the subjective information in a piece of user-written text and classify it into one of the predefined set of classes (e.g., *positive*, *negative*, *conflict*, or *neutral*). The application ranges from an individual's learning from other users' experiences to a feedback system for organizations (e.g., a user review highlighting the cons of a product or service can cause severe damage to the reputation of the product or service and also to the organization that offers such a product or service). Sentiment analysis performed on a coarser level (i.e., document or sentence level) does not reveal crucial information to a user who is sensitive to the finer details, such as the *display quality* of a laptop or the *ambience* of a restaurant). Aspect-Based Sentiment Analysis (ABSA) (Hu and Liu 2004; Pontiki et al. 2014) is a relatively new dimension of sentiment analysis which analyzes the text at a much finer granularity level. It first identifies various aspects (or features or attributes) of the product or service in a text and then assigns a sentiment class to each of these. For example, in the following review, the user has expressed different sentiments (i.e., *positive* and *negative*) toward two aspects (i.e., *food* and *service*) of a restaurant. Though the user is happy with the quality of food being served, she was not amused with the service.

Great food but the service was dreadful!

On a sentence-level, the example conveys a *conflict* sentiment toward the restaurant, without revealing the details of what the user liked and disliked. In contrast, ABSA identifies all the aspects (i.e., *food* and *service*) that have been discussed in the review and then assigns the *positive* sentiment to *food* and the *negative* sentiment to *service*, thus revealing the finer details of the restaurant review. Identification of aspect terms is known as *aspect term extraction* or *opinion target extraction*, whereas assigning polarity information to the aspect terms is known as *sentiment classification*. In this work, we focus on both aspect term extraction and aspect sentiment classification tasks.

Literature survey evidences a wide range of research on sentiment analysis (either at the sentence level or at the document level) being carried out in recent years (Chernyshevich 2014; Gupta et al. 2015; Jagtap and Pawar 2013; Kaljahi and Foster 2016; Kim and Hovy 2004; Mukherjee and Liu 2012; Poria et al. 2016; Toh and Wang 2014; Turney 2002; Wagner et al. 2014; Zhuang et al. 2006). However, most of these research efforts are focused on resource-rich languages, predominantly English. Like many other NLP problems, research on sentiment analysis involving Indian languages (e.g., Hindi, Bengali, etc.) is limited (Bakliwal et al. 2012; Balamurali et al. 2012; Joshi et al. 2010; Kumar et al. 2015; Singhal and Bhattacharyya 2016). Due to the scarcity of various qualitative resources and/or tools in such languages, the problems have become more nontrivial and challenging to solve.

The research on ABSA involving Indian languages has started only recently (e.g., Akhtar et al. (2016a, 2016b)). In one of our earlier research efforts, we created a benchmark setup for ABSA in Hindi (2016a). For evaluation purpose, we trained Conditional Random Field (CRF) and a Support Vector Machine (SVM) for aspect term extraction and sentiment classification, respectively. We proposed a hybrid Convolutional Neural Network (CNN)-based model for aspect-level sentiment classification in another recent work proposed in Akhtar et al. (2016b). We showed the effectiveness of our proposed approach in multiple domains, namely product reviews, movie reviews, and Tweets.

2 MOTIVATION AND PROBLEM DEFINITION

As discussed earlier, Indian languages are resource-constrained in nature (i.e., there is a lack of readily availability of different resources and tools such as Part-of-Speech (PoS) tagger, Named

Entity Recognizer (NER), Parser, Morphological Analyzer and the like in the required measure. As a side-effect, it is more challenging to build a robust system for any particular application.

It is well-established that a good amount of annotated corpus is the foremost requirement for any supervised machine learning system to achieve an acceptable performance level. Insufficient training samples for building the model have been a huge bottleneck in solving the problems in resource-constrained languages (e.g., Indian languages such as Hindi, Bengali, etc.). Traditionally, researchers project the problem into a common space in a cross-lingual setup and aim to leverage the availability of various resources/tools in the resource-rich languages (Balamurali et al. 2012; Barnes et al. 2016; Singhal and Bhattacharyya 2016; Zhou et al. 2016). Bilingual dictionaries and/or machine translation had been the preferred choice for this projection.

Recently, Deep Learning (DL)-based techniques have been established as the benchmark in solving several NLP problems due to its capability of extracting relevant set of features on its own (i.e., during training) from the underlying word embedding representations. Thus, it minimizes dependency on the extracted features from any external resource/tool. A qualitative word representation (word embeddings or word vector) plays an important role in any (deep) neural network-based system. It is a nontrivial task for any DL architecture to learn and extract the relevant hidden features (semantic, syntactic, or lexical) without qualitative word representations, hence it hampers the overall performance of the system. In practice, state-of-the-art distributed representation models such as GloVe (Pennington et al. 2014) or Skip-Gram with Negative Sampling (SGNS) (Mikolov et al. 2013) are common choices to preserve the quality of word representation. However, these systems require a large corpus for training and building the models (e.g., the pretrained Google News Word2Vec model was trained on 100 billion words, pretrained common-crawl GloVe model was trained on 840 billion words). Unfortunately, the lack of corpus in such a range limits the quality of word representation, and many languages are not at par with English or other resource-rich languages.

Another great challenge in the learning process is the missing representations of input words (Out-of-Vocabulary (OOV) words) in a word representation model. This gives rise to the problem of data sparsity with regard to word representation. The trivial solution, as reported in the literature, is to use either a random vector (Dhingra et al. 2017) or a zero vector (Bahdanau et al. 2017) for such words. Though both these solutions are easy to use, they do not provide relevant and contextual information to the learning algorithm, in general. Similarly in a cross-lingual scenario, representation of a word in a source language does not have any correlation with the representation of the translated word in the target language, hence, these are not ideal for use directly for training and/or testing.

In this article, we aim to improve the coverage of word representation in a resource-constrained language scenario (here, for Hindi word embeddings) by leveraging the information of resource-rich language (here, English word embeddings). In its original form, embedding of a word in one language (say, Hindi) and embedding of the same word (translated) in another language (say, English) does not pose any association between them. Hence, word embeddings of one language can not be directly used for other language(s). Therefore, we utilize bilingual word embeddings (Luong et al. 2015) trained on an English-Hindi parallel corpus to bridge language divergence in the vector space. The proposed method is based on a DL architecture, the Long Short Term Memory (LSTM) network (Hochreiter and Schmidhuber 1997). For benchmarking, we evaluate the proposed method for aspect-level sentiment analysis in both multi-lingual and cross-lingual setups. In this work, we address two subtasks of aspect-level sentiment analysis; namely *aspect term extraction* and *aspect sentiment classification*. For evaluation, we use the dataset that we created ourselves (Akhtar et al. 2016a). It consists of 5,417 review sentences in Hindi. In the cross-lingual setup, we

use an English dataset from SemEval 2014 shared task on ABSA (Pontiki et al. 2014) for training, while for testing we use Hindi datasets from Akhtar et al. (2016a). Another reason for utilizing bilingual word embeddings the fact that the target dataset (Akhtar et al. 2016a) (which is primarily in Hindi Devanagari script) also contains a few English words (both in transliterated form and Roman script) for which no representation is available in Hindi embeddings (e.g., DVD, कॉम्बिनेशन | (combination), इंस्टालेशन | (installation), यूजर इंटरफ़ेस | user interface, etc).

Major contributions and/or features of our proposed approach are as follows: (i) We train and use bilingual embeddings on an Amazon product review corpus consisting of *English–Hindi* parallel corpora, which assists in bridging the diversity of two languages; (ii) we propose to improve the word representation coverage in a low-resource language by utilizing the resource-rich language word embeddings; (iii) we leverage the semantic richness of various lexicons of the target (English) language for enhancing the performance of the system; (iv) we study three competitive bilingual/cross-lingual word representation approaches for ABSA; and (v) we provide a detailed comparative and error analysis of the obtained results.

As already mentioned, research on ABSA involving Indian languages is very limited. Some recent works include our previous efforts, such as Akhtar et al. (2016a, 2016b), related to ABSA. We proposed feature-driven supervised approaches in Akhtar et al. (2016a) for aspect term extraction and aspect sentiment classification. (We trained CRF and SVM as classifiers on top of various language-independent features for aspect term extraction and aspect sentiment classification, respectively.) In a CNN-based hybrid model proposed in Akhtar et al. (2016b), we successfully showed cascading of CNN and SVM for a wide variety of problems, including aspect-level sentiment classification. In addition, we also utilized an optimized set of features obtained through a multi-objective genetic algorithm-based optimization technique for enhancing the performance. We evaluated the efficacy of this proposed approach in multiple domains and languages. A multi-lingual CNN-based sentiment analysis (not ABSA) model has been proposed in Singhal and Bhattacharyya (2016). The core idea of the work was to project all words in a resource-poor language into a resource-rich language via machine translation. In addition, the authors also modified a training dataset by augmenting all the polar words along with their polarities as training instances. Barnes et al. (2016) employed bilingual word embeddings for sentiment classification in a cross-lingual setup. In comparison, our proposed system utilizes bilingual embeddings to reduce the effect of data sparsity in both cross-lingual and multi-lingual scenarios. We summarize the key differences between our proposed approach and these existing systems in Table 1.

Our current research is an extension to our previously proposed work on sentiment analysis (Akhtar et al. 2018). However, in comparison to the earlier work, our current work differs on the following points: (i) The existing work (Akhtar et al. 2018) studies only the aspect sentiment classification problem, whereas in the current work we address both *aspect term extraction* and *aspect sentiment classification*. It should be noted that both of these problems refer to the two different paradigms of supervised classification (i.e., aspect term extraction is a sequence-labeling problem, whereas the aspect sentiment classification is a classification problem). We have empirically shown that our proposed approach is effective for solving both these problems, viz. classification (aspect term classification) as well as sequence-labeling (aspect term extraction). (ii) We study and analyze the behavior of three different forms of bilingual/cross-lingual word embeddings in our current work. In contrast, in our previous work, we employed only one bilingual word embedding for the study. (iii) In our current work, we provide detailed descriptions of the various modules (e.g., bilingual word embeddings, *English→Hindi* Statistical Machine Translation (SMT) systems, etc.). (iv) We present here a detailed analysis for comparison with existing systems and the errors that we encounter.

Table 1. Key Differences of the Proposed Approach and the Existing Systems

	Proposed Approach	Barnes et al. (2016)	Singhal and Bhattacharyya (2016)	Akhtar et al. (2016b)	Akhtar et al. (2016a)
Setup	Multi-lingual & Cross-lingual	Cross-lingual	Multi-lingual	Mono-lingual	Mono-lingual
Approach	Deep Neural Network (LSTM)	Word embedding (SVM)	Deep Neural Network (CNN)	Deep Neural Network (CNN)	Feature-driven (CRF & SVM)
Problem	Aspect term extraction & Aspect sentiment classification	Aspect sentiment classification	Sentence-level sentiment classification	Aspect sentiment classification	Aspect term extraction & Aspect sentiment classification
Word Embeddings	Shared vector-space bilingual embeddings	Shared vector-space bilingual embeddings	Projected source language dataset to target language and utilizes target-side pre-computed embeddings.	Mono-lingual embeddings	-
Data Sparsity	Minimize the effect of data sparsity through bilingual embeddings.	-	Minimized the effect of data sparsity through $src \rightarrow tgt$ projection.	-	-
	Replace the OOV words with translated forms, which usually happens to be its closest neighbour in the shared vector space. Hence, the semantic closeness is preserved to an extent.	-	Translated each word of the source language into target language which may introduce loss of sentiment in target language as a side-effect (Mohammad et al. 2016).	-	-
Hand-crafted Features	Richer set of lexicon based features.	-	Augmented the polar words in the training instances.	Optimized feature set from multi-objective genetic algorithm.	Basic features

2.1 Problem Definition

As discussed earlier, *data sparsity* often poses a significant challenge to machine learning (and neural network learning). The prime motivation of this work is to minimize the effect of *data sparsity*, thereby enabling any DL framework to effectively learn its hidden features. For this, we propose to use bilingual embeddings computed from a parallel corpus (approx. 7.2M English-Hindi parallel sentences). We hypothesize that addressing data sparsity in an intelligent manner will yield increased performance. We try to establish our hypothesis through experiments on two ABSA tasks, namely *aspect term extraction* and *aspect classification* in both multi-lingual and cross-lingual scenarios. Next, we present a very brief description for each these two tasks.

Table 2. Example of Aspect-Based Sentiment Analysis and Its Corresponding BIO Encoding Scheme

Review Text	
Devanagari	इसकी ऑडियो क्वालिटी शानदार है।
Transliterated	Isakee AWDiyo kvaaliTee shaanadaara hai.
Translated	Its audio quality is superb.

	इसकी	ऑडियो	क्वालिटी	शानदार	है	
BIO encoding	O	B-ASP	I-ASP	O	O	O
Modified encoding	O	I-ASP	I-ASP	O	O	O

Aspect term	Aspect sentiment classification
ऑडियो क्वालिटी AWDiyo kvaaliTee audio quality	positive

2.1.1 Aspect Term Extraction. The task of aspect term extraction is to predict the boundaries of all aspect terms present in a sentence. To tackle multi-word aspect terms (e.g., *battery life*), we follow the BIO notation scheme to mark each token as either **B-ASP** (begin of an aspect term), **I-ASP** (inside an aspect term), or **O** (outside of aspect term). This projects the problem as a sequence labeling task, where the current prediction depends on the current input as well as on the previous output. During the experiment, we observe that the distribution of training instances belonging to these three categories (**B-ASP**, **I-ASP**, and **O**) is highly imbalanced in nature (i.e., approximately 99% of the tokens belong to the class '**O**'). We try to address this issue by projecting the three-class scheme (i.e., **B-ASP**, **I-ASP**, and **O**) into a two-class encoding scheme (i.e., **I-ASP** and **O**) by merging **B-ASP** and **I-ASP** together. This encoding scheme was employed in the CoNLL 2003 shared task on Named Entity Recognition (Tjong Kim Sang and De Meulder 2003). Consequently, we observe improved performance in the prediction. An example scenario is depicted in Table 2.

2.1.2 Aspect Sentiment Classification. Aspect-based sentiment classification deals with assigning sentiment polarity (i.e., *positive*, *negative*, *neutral*, or *conflict*) to the aspect terms. We define a context window of $\pm 5^1$ words around the aspect term as our training and classification instances. The reason for adapting such an arrangement is to ensure that the sentiment-bearing words associated with one aspect term do not intervene with the classification of the other aspect terms present in the same sentence. Also, it was observed during analysis of the dataset that sentiment-bearing words often appear closer to the target aspect terms (with a few exceptions).

3 PROPOSED METHODOLOGY

In this section, we describe our proposed methodology that we adapt for *aspect term exaction* and *sentiment classification* in Hindi. We propose to use an LSTM architecture on top of bilingual word embeddings for prediction. LSTM is a special kind of Recurrent Neural Network (RNN) which efficiently learns long-term dependencies. Bidirectional LSTM is an extended version of LSTM which

¹The value of 5 was set empirically by varying the context window size from 2 to 6.

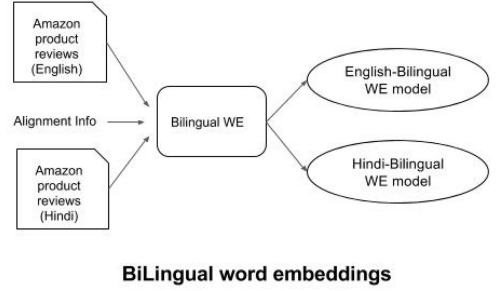
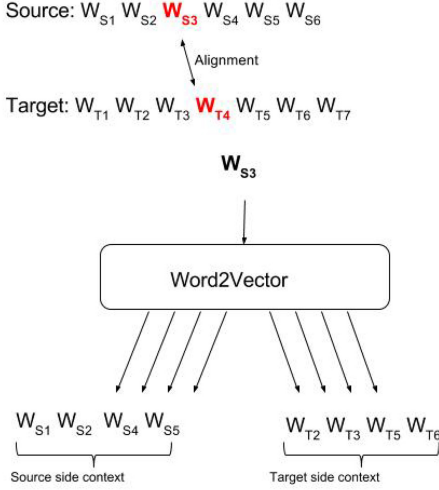
BiLingual Skip-gram model

Fig. 1. Training scenario for skip-gram bilingual word embeddings.

takes both forward and backward sequences into account. Our model consists of two bidirectional LSTM layers followed by two fully connected layers and one output layer.

3.1 Bilingual Word Embedding

We employ bilingual word embeddings (Luong et al. 2015) trained on a parallel English-Hindi corpus. We generate a parallel corpus for an Amazon product review datasets² (consisting of approx. 7.2M sentences) using an in-house product-review-domain based *English*→*Hindi* Statistical Machine Translation (SMT) system.³

The parallel corpus, along with the alignment information, is used to train two (English and Hindi) Skip-Gram with Negative Sampling (SGNS) Word2vec (Mikolov et al. 2013) models which share the common vector space. If a word W_S is aligned to word W_T then the context information C_T of the target word W_T is also used as the context of the source word W_S along with its own context information C_S for computing the word vectors. The underlying idea follows a famous quotation by English linguist Firth that “*You shall know a word by the company it keeps*” (Firth 1957). By utilizing the context information of both the source and target sides, resultant word embeddings of W_S and W_T come semantically closer to each other in the vector space. An example is shown in Figure 1.

A bilingual skip-gram model creates two separate word embeddings, one each for source (Hindi) and target language (English). First, we extract word representations for all the words in a sentence from the Hindi monolingual word embeddings. Subsequently, at the second step, we translate all OOV words (words whose representations are missing in Hindi embeddings) into English and then perform another lookup in the corresponding English word embeddings. For instance, if embedding of a word “अच्छा|achcha” is unknown, we translate it into the English word “good” and use its word embedding in place of the source word “अच्छा|achcha.” Thus, the missing

²<http://snap.stanford.edu/data/other.html>.

³39.5 BLEU score.

representation of the OOV word is replaced by its translated target side representation. Since both English and Hindi word embeddings share a common vector space, this replacement strategy proves to be an effective technique. In our case, we observe a reduction of approximately 65% (i.e., 243 OOVs remaining out of total 698 OOVs) in OOV words by the proposed replacement strategy. Consequently, an increase in the F-measure/accuracy value is also observed during evaluation.

3.1.1 English→Hindi SMT System. The English→Hindi MT system is based on a standard phrase-based SMT (Koehn et al. 2003) system. We employ a widely used and standard machine translation toolkit *Moses* (Koehn et al. 2007) for training the system. The alignment information is obtained from the *mosesdecoder* (Koehn et al. 2007) during translation of the reviews. We use the following set of parameters in *Moses*: *grow-diag-final-and* heuristics for word alignment using GIZA++ (Och and Ney 2003), *msd-bidirectional-fe* for reordering the model, and a *4-gram language model* with modified *Kneser-Ney smoothing* (Kneser and Ney 1995) using KenLM (Heafield 2011). We use the *minimum error rate training* (Och 2003) for tuning the system. The SMT system was trained on an in-house product domain parallel corpus. After preprocessing, the numbers of parallel sentences for training, test, and development sets are 112,469, 5,640, and 602, respectively.

3.2 Features

Finally, we employ various hand-crafted features to assist the network. In addition to the features targeted for Hindi, we also try to leverage the effectiveness of English side resources by translating a word into English and then extracting its feature representation. We extract and implement the following set of features for use in our tasks. It should be noted that we do not include any lexical or syntactic features during training as these features are automatically learned from the data itself.

3.2.1 Aspect Term Extraction.

(1) Hindi:

- a. **IndoWordNet** (Bhattacharyya 2010) **synset**: WordNet is lexical database which groups a set of words based on their senses, called *synsets*. In Akhtar et al. (2017) it has been shown that synset-based feature provides crucial information for predicting unseen examples. Following a similar approach, we extract the top 3 most frequent words from the synset and use them as a feature value.
- b. **Frequent aspect terms**: From training dataset, we compile a list of frequent aspect terms which occur at least 5 times in the corpus. We then define a binary-valued feature that fires if the current token is present in the list. This feature is included with the assumption that infrequent words have more chances of belong to the aspect terms.

Considering aspect term extraction as a sequence labeling task, we did not include English resources for identifying aspect terms because, after translation, word order of the source language is not preserved in the target language. Hence, the sequence information gets distorted, which, in turn, will confuse the system rather than assist it.

3.2.2 Aspect Sentiment Classification.

(1) Hindi:

- a. **SentiWordNet for Indian languages** (Das and Bandyopadhyay 2010): We define two features that mark the sum of positive scores and sum of negative scores of all the words in a sentence. We assign a score of +1 and -1, respectively, to each positive and negative word in the sentence.
- b. **Semantic Orientation (SO)** (Hatzivassiloglou and McKeown 1997) **score**: Semantic orientation defines the association of a word with regard to its positivity and negativity.

It can be defined as

$$SO(w) = PMI(w, pos) - PMI(w, neg),$$

where $PMI(w, pos)$ and $PMI(w, neg)$ are the point-wise mutual information of word w in positive and negative reviews, respectively. We compute the SO score of each word in the context window of size ± 5 and take the cumulative SO score as a feature value for training the system.

(2) **English:**

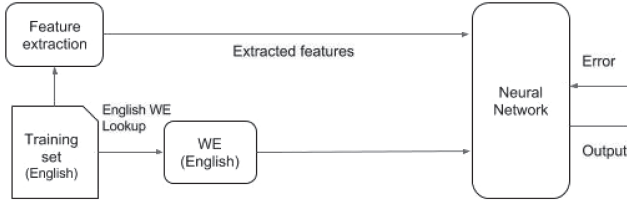
- a. **Bing Liu** (Ding et al. 2008) **lexicon**: We define two features in the same way as we did with Hindi SentiWordNet.
- b. **MPQA** (Wiebe and Mihalcea 2006) **lexicon**: We extract two features following the same approach as above.
- c. **SentiWordNet** (Baccianella et al. 2010): Three features are extracted for every word, denoting its positivity (posScore), negativity (negScore), and objectivity ($1 - [\text{posScore} + \text{negScore}]$) score, respectively.
- d. **Semantic Orientation (SO)** (Hatzivassiloglou and McKeown 1997) score: This feature is defined in the same way as mentioned earlier. Only difference is we utilize an English corpus for calculating SO scores.

An overall schema of the proposed methodology is depicted in Figure 2. Figure 2(a) and (b) show the training architectures for the cross-lingual and multi-lingual scenarios, respectively. Since our test datasets for both variants are in Hindi, the testing scenario for cross-lingual and multi-lingual setups are also the same as represented in Figure 2(c).

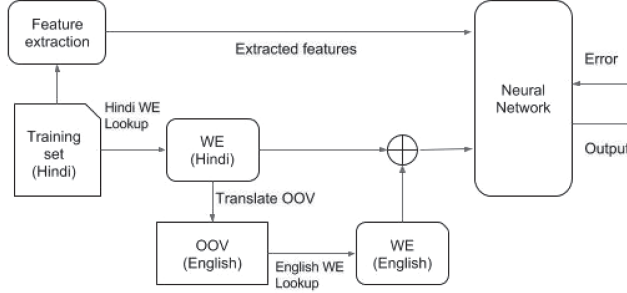
3.2.3 Deep Neural Network Architecture. For an effective combination of word embeddings and extracted features, we try three different architectures as depicted in Figure 3. In the first architecture (A1, Figure 3(a)), we combine the extracted features with word representations and pass it through an LSTM network followed by dense and output layers. In the second architecture (A2, Figure 3(b)), we do not combine features and word representations. Rather, we learn sentence embeddings through an LSTM network and then concatenate it with the extracted features before feeding it to the dense layer. Finally, in the third architecture (A3, Figure 3(c)), we train separate LSTMs for the extracted features and word embeddings. Subsequently, we merge their representations at the dense layer. The choice of separate LSTMs for the hand-crafted features in architecture A3 is driven by the fact that the dimension of a word embedding is usually very high as compared to its corresponding hand-crafted features. If trained together, as in architecture A1, extracted features of low dimension usually get overshadowed by the high-dimensional word embeddings, thus making it nontrivial for the network to learn from the extracted features. Further, to exploit the sequence information of words in a sentence, we pass the hand-crafted features of each word through a separate LSTM layer. For example, in the following review sentence there are two positive words (“liking” and “recommending”) and only one negative word (“far”). In a model that takes into account only the simple polar word score, the sentence would have high relevance toward the positive sentiment. However, the sequence information of the phrase “*far from liking and recommending*” dictates the negative sentiment of the sentence.

“I’m far from liking and recommending this phone to anyone.”

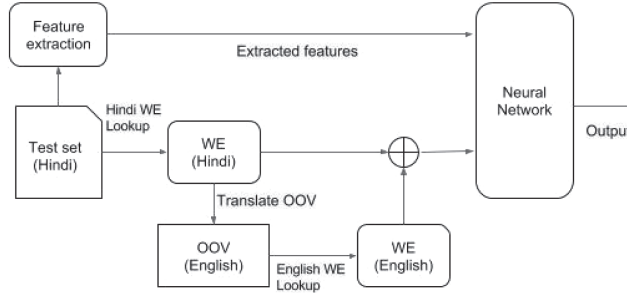
In contrast to A3, architecture A2 does not rely on the sequence information of the extracted features and lets the network learn on its own.



(a) Training scenario in cross-lingual setup.



(b) Training scenario in multi-lingual setup.



(c) Testing scenario in cross-lingual and multi-lingual setup.

Fig. 2. Proposed schema.

4 EXPERIMENTAL RESULTS

In subsequent subsections, we describe the datasets, experimental setup, and evaluation results and provide the necessary analysis.

4.1 Datasets

We use an ABSA dataset⁴ in Hindi that we developed ourselves (Akhtar et al. 2016a) for evaluation purposes. A total of 5,417 review sentences across 12 domains are present, along with 4,509 aspect terms. Each aspect term belongs to one of the four classes: *positive*, *negative*, *neutral*, and *conflict*. We split the dataset into 70%, 10%, and 20% as training, development, and test data, respectively, for the experiment.

In the cross-lingual setup, we utilize the English dataset of SemEval-2014 shared task on ABSA (Pontiki et al. 2014) for training and the Hindi ABSA dataset (Akhtar et al. 2016a) for testing. The

⁴<http://www.iitp.ac.in/~ai-nlp-ml/resources.html>.

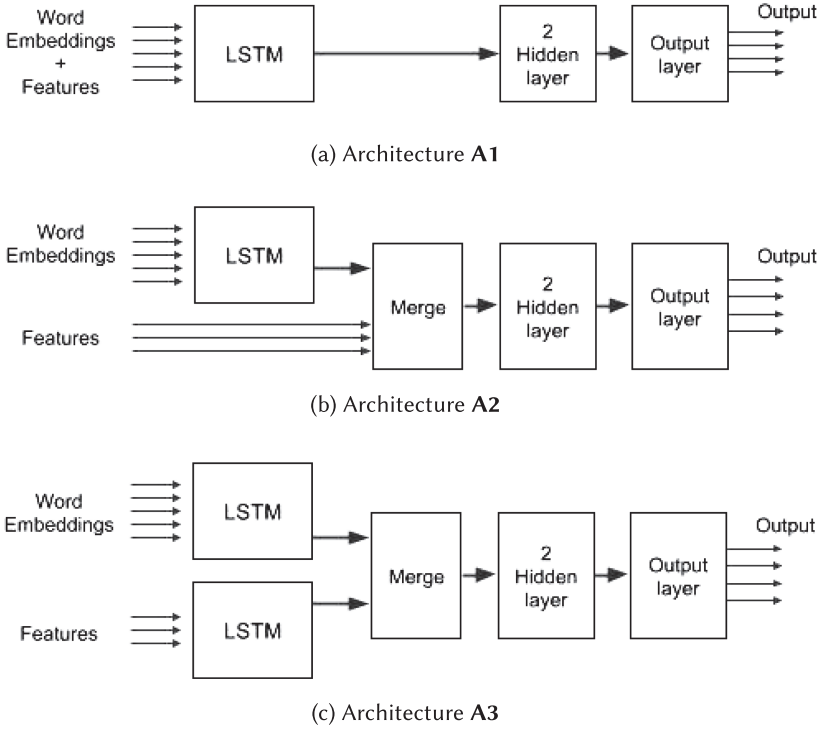


Fig. 3. Neural network architectures.

Table 3. Dataset Statistics: Pos: Positive, Neg: Negative, Neu: Neutral, and Con: Conflict

Datasets	Sentiment				#Aspect Terms	#Reviews
	#Pos	#Neg	#Neu	#Con		
Hindi (Akhtar et al. 2016a)	1,986	569	1,914	40	4,509	5,417
English (Pontiki et al. 2014)	1,328	994	629	61	3,012	3,845

English dataset comprises product reviews in two domains (i.e., restaurant and laptop). However, we only employ the laptop domain dataset, as most of the reviews in the Hindi ABSA datasets belong to the electronics domain. For training in a cross-lingual setup, we combine the training and gold test datasets. In total, there are 3,845 review sentences comprising of 3,012 aspect terms.

Brief statistics of both datasets are presented in Table 3.

4.2 Experiments

We use the Python-based neural network library Keras⁵ for implementation. For evaluation, we use the standard F-measure and accuracy for aspect term extraction and aspect sentiment classification, respectively. We train the network for 100 epochs with the early stopping criteria turned on (i.e., preserving the best learned parameter at each epoch). As an activation function, we utilize “*tanh*” at the intermediate layers, while for classification, we use “*softmax*” at the output layer. To prevent the network from overfitting, we incorporate an efficient regularization

⁵<http://keras.io>.

Table 4. **Aspect Term Extraction in Multi-lingual Setup:** Baseline and Experimental Results

Models	F-measure (%)
Baseline (Monolingual WE)	50.03
Bilingual	48.01
Bilingual + Embedding (OOV)	53.40

OOV: Out-of-vocabulary words.

Table 5. **Aspect Term Extraction in Cross-lingual Setup:** Baseline and Experimental Results

Models	F-measure (%)
Baseline (Monolingual WE)	06.31
Bilingual	25.27
Bilingual + Embedding (OOV)	33.39

OOV: Out-of-vocabulary words.

technique called *Dropout* (Srivastava et al. 2014). At each layer of training, dropout skips a few neurons randomly. We fix the dropout rate at 45% during training, while for optimization we use the *Adam* optimizer (Kingma and Ba 2014).

4.2.1 Aspect Term Extraction. Experimental results for aspect term extraction in the multi-lingual setup are reported in Table 4. The first row represents an LSTM-based baseline system that utilizes monolingual word embeddings (i.e., an SGNS model trained only on 7.2M Amazon review translated Hindi sentences) for the predictions. The next two rows represent usage of bilingual word embeddings. The difference between the second row (48.01%) and baseline system (50.03%) of Table 4 is the usage of bilingual embeddings in place of monolingual embeddings. We observe a performance loss of approximately 2 percentage points with bilingual embeddings. However, when OOV words are translated and corresponding word embeddings are computed, the same LSTM network produces an F-measure of 53.4%, thus yielding a performance increase of approximately 5.4 points. This improvement evidences that the richness of target language (English) word embeddings helps the system to efficiently solve the problems encountered in the resource-poor source language. Finally, we introduce the extracted features of Section 3.2.1 to the network. However, performance of the system does not improve. A possible reason for such result would be the presence of a high-dimensional sparse feature vector (i.e., one-hot encoding for IndoWordNet synset).

The results of aspect term extraction in the cross-lingual setup, where we train the network utilizing the English dataset (Pontiki et al. 2014) and evaluate the model on the Hindi dataset (Akhtar et al. 2016a), are reported in Table 5. The baseline model in Table 5 employs monolingual word embeddings of English and Hindi for training and testing, respectively. Since the vector space of the two difflanguages is completely unrelated, it is no surprise that the baseline system achieves a mere 06.31% F-measure. Using only the bilingual word embeddings, the system achieves a 25.27% F-measure. By increasing the coverage of input word embeddings using machine translation, the proposed system obtains an increased F-measure of 33.39%. This improvement in F-measure justifies the use of translated words for obtaining word embeddings. Similar to the multi-lingual case, augmenting handcrafted features does not have good effect on overall system performance.

Table 6. **Aspect Classification in Multi-lingual Setup: Baseline and Experimental Results**

Models		Accuracy (%)
Baseline (Monolingual WE)		63.64
Bilingual WE		62.51
Bilingual WE + Embedding (OOV)		64.83
Bilingual WE + Embedding (OOV) + Features (Hin)	A1	71.52
	A2	71.58
	A3	73.50
Bilingual WE + Embedding (OOV) + Features (Eng)	A1	71.32
	A2	73.50
	A3	76.29

A1: Word embeddings and extracted feature are combined and fed into single LSTM network. **A2:** Extracted features are directly merged with LSTM output. **A3:** One LSTM network each for word embeddings and extracted features. **OOV:** Out-of-vocabulary words.

4.2.2 Aspect Sentiment Classification. In Table 6, we report the results for aspect sentiment classification in the multi-lingual setup. Similar to the aspect extraction task, monolingual word embeddings (63.64%) work better than bilingual word embeddings (62.51%). However, after addressing the problem of data sparsity, the performance of the system improves to 64.83%. Thus, it establishes the effectiveness of our proposed approach in addressing the data sparsity problem where resource-rich target language (i.e., English) word embeddings come to the rescue of resource-poor languages by minimizing the problem to a greater extent.

Further, we utilize in-language lexicons (SentiWordNet for Indian languages and Semantic score) to see the effect of hand-crafted features. As already mentioned, we experiment with three architectures (i.e., A1, A2, A3) for both multi-lingual and cross-lingual setups: A1: we concatenate resultant word embeddings with the extracted features and learn a single LSTM network for the prediction; A2: we concatenate the LSTM learned sentence embeddings with the extracted features at the dense layer for classification; and A3: we train two LSTM networks separately, one each for word embeddings and extracted features, and then merge the learned features together for classification.

First architecture (A1) reports an achieved accuracy of 71.52% while the second (A2) and third architectures (A3) obtain accuracies of 71.58% and 73.50%, respectively. Since the number of lexicons in Hindi are limited, we try to leverage the relatively high-quality lexicons of English in our system. Similar to the previous case, we experiment with all three architectures. We obtain accuracies of 71.32%, 73.50%, and 76.29% for A1, A2, and A3, respectively.

Table 7 reports the experimental results for aspect classification in cross-lingual setup. Similar to the aspect term extraction cross-lingual setup, the baseline model reports a mere 16.29% accuracy by utilizing monolingual word embeddings of English and Hindi for training and testing, respectively. However, the system achieves an accuracy of 48.94% when we introduce bilingual word embeddings. With the increased word embedding coverage (i.e., after OOV translation), the proposed system reports an increased accuracy of 50.79%. Furthermore, with the inclusion of target-side lexicon-based features our proposed system reports a significant performance improvement of approximately 6–10 points for all the three architectures.

We observe three phenomena from these results: (i) qualitative lexicons of a resource-rich language can assist in solving the problems of a resource-poor language; (ii) use of lexicon-based features is the driving force for predicting sentiment in aspect classification task; and (iii) use of

Table 7. **Aspect Classification in Cross-lingual Setup:** Baseline and Experimental Results

Models		Accuracy (%)
Baseline (Monolingual WE)		16.29
Bilingual WE		48.94
Bilingual WE + Embedding (OOV)		50.79
Bilingual WE + Embedding (OOV) + Features (Eng)	A1	56.68
	A2	56.90
	A3	60.39

A1: Word Embeddings and Extracted Feature are Combined and Fed into Single LSTM Network. **A2:** Extracted features are directly merged with LSTM output. **A3:** One LSTM network each for word embeddings and extracted features. **OOV:** Out-of-vocabulary words.

Table 8. **Aspect Term Extraction in Multi-lingual Setup:** Comparative Systems

Models	Description	F-measure (%)
Baseline (Monolingual WE)		50.03
System (Akhtar et al. 2016a)	Features - CRF (3 class i.e. <i>B_ASP</i> , <i>I_ASP</i> & <i>O</i>)	41.07
	Features - CRF (2 class i.e. <i>I_ASP</i> & <i>O</i>)	43.07
Proposed	Bilingual WE + Embedding (OOV) - LSTM	53.40

Table 9. **Aspect Term Extraction in Cross-lingual Setup:** Comparative Systems

Models	Description	F-measure (%)
Baseline (Monolingual WE)		06.31
Proposed	Bilingual WE + Embedding (OOV) - LSTM	33.39

Baseline (Monolingual WE) \Rightarrow : Trained on english monolingual WE and Evaluated on Hindi monolingual WE.

separate LSTMs (one for word embeddings and the other for features) helps the network to efficiently extract relevant features for prediction without interfering with each other.

4.3 Comparative Analysis

We present the comparative results in Tables 8, 9, 10, and 11. We observe that our proposed system clearly outperforms the baseline model in both aspect term extraction and sentiment classification.

We further compare our system with existing state-of-the-art systems to establish the success of our approach. For aspect term extraction in multi-lingual setup (Table 8), we compare our proposed approach with Akhtar et al. (2016a), which is a feature-driven CRF-based model. In our previous system (Akhtar et al. 2016a), we obtained an F-measure value of 41.07% utilizing various lexical and syntactic features, while our current proposed system yields an F-measure of 53.4%, an improvement of more than 12 points.

We do not compare our system for aspect term extraction in a cross-lingual setup as we are not aware of any state-of-the-art system concerning language pairs. However, as depicted in Table 9, our proposed system performs better than the baseline system, with an improvement of more than 25 points, which clearly shows the effectiveness of bilingual word embeddings.

Table 10. **Aspect Classification in Multi-lingual Setup:** Comparison with the Baseline and State-of-the-art Methods

Models	Description	Accuracy (%)
Baseline (Monolingual WE)		63.64
System (Akhtar et al. 2016a)	Feature - SVM	54.05
System (Akhtar et al. 2016b)	Monolingual WE + Lexicons - CNN	65.96
System (Singhal and Bhattacharyya 2016)	Translate (All words) + Monolingual WE + Lexicon - CNN	68.31
Proposed system	Bilingual WE + Embedding (OOV) + Lexicons - LSTM	76.29

Table 11. **Aspect Classification in Cross-lingual Setup:** Comparison with the Baseline and State-of-the-art Methods

Models	Description	Accuracy (%)
Baseline (Monolingual WE)		16.29
System (Barnes et al. 2016)	Bilingual WE - SMO	39.47
System (Singhal and Bhattacharyya 2016)	Translate (All words) + Monolingual WE + Lexicon - CNN	56.22
Proposed system	Bilingual WE + Embedding (OOV) + Lexicons - LSTM	60.39

Baseline (Monolingual WE) \Rightarrow : Trained on English monolingual WE and evaluated on Hindi monolingual WE.

In the multi-lingual setup for aspect classification task, we compare the proposed model against three state-of-the-art systems (Akhtar et al. 2016a, 2016b; Singhal and Bhattacharyya 2016). In our earlier attempt (Akhtar et al. 2016a), we developed a feature-based SVM system for aspect classification and obtained an accuracy of 54.05%. Application of CNN-based sentiment classification was reported in both (Akhtar et al. 2016b) and in Singhal and Bhattacharyya (2016). Our previous system (Akhtar et al. 2016b) is a monolingual approach which learns from the in-language word embeddings and optimized set of features, while the system in Singhal and Bhattacharyya (2016) is a multilingual approach and tries to leverage the resources of a resource-rich language. An accuracy of 65.96% was reported in Akhtar et al. (2016b), while the system reported in Singhal and Bhattacharyya (2016) obtained an accuracy value of 68.31%. However, in our currently proposed system, we achieve an accuracy of 76.29%, which is approximately 10% and 8% higher compared to the other systems (Akhtar et al. (2016b) and Singhal and Bhattacharyya (2016)), respectively.

For the cross-lingual setup, we compare our proposed method with the state-of-the-art system proposed in Barnes et al. (2016) and Singhal and Bhattacharyya (2016). Barnes et al. (2016) used bilingual word embeddings to train a Sequential Minimization Optimization (SMO) classifier. On the same dataset, their systems reportedly achieved accuracies of 39.47% and 56.22% as compared to 60.39% for our proposed system.

Statistical significance tests (*T-test*) show that performance improvements in the proposed model are statistically significant (95% confidence level) with p -value = 0.03786 and p -value = 0.01361 in multi-lingual and cross-lingual setups for sentiment classification problem, respectively. Similarly, for aspect term extraction, the obtained results are significant, with p -value = 0.04451 and p -value = 0.03166 for multi-lingual and cross-lingual setups, respectively.

Table 12. Performance of the Proposed System with Different Bilingual/Cross-lingual Embeddings

Models	Multi-lingual			Cross-lingual		
	Embeddings			Embeddings		
	E1	E2	E3	E1	E2	E3
Bilingual	48.01	37.72	42.18	25.27	27.00	36.64
Bilingual + Embeddings (OOV)	53.40	40.30	45.67	33.39	29.59	37.81

a. Aspect Term Extraction

Models		Multi-lingual			Cross-lingual		
		Embeddings			Embeddings		
		E1	E2	E3	E1	E2	E3
Bilingual		62.51	63.24	55.09	48.94	43.11	46.64
Bilingual + Embeddings (OOV)		64.83	63.57	56.22	50.79	44.03	48.80
Bilingual + Embeddings (OOV) + Features (Hindi)	A1	71.52	70.33	66.22	-		
	A2	71.58	69.80	64.76			
	A3	73.50	71.05	66.15			
Bilingual + Embeddings (OOV) + Features (English)	A1	71.32	73.64	71.58	56.68	52.64	58.27
	A2	73.50	73.24	69.93	56.90	58.07	60.52
	A3	76.29	73.77	71.92	60.39	62.03	64.58

b. Sentiment Classification

Embeddings **E1**: Utilizes aligned parallel corpus (Luong et al. 2015); **E2**: Utilizes canonical correlation analysis (Faruqui and Dyer 2014) and **E3**: Utilizes aligned-documents (Vulić and Moens 2015).

4.4 Comparison with Other Bilingual/Cross-lingual Embeddings

In this subsection, we present the comparative analysis of our proposed method utilizing other bilingual/cross-lingual embeddings. For the analysis we employ three different techniques to compute the bilingual/cross-lingual embeddings (i.e., embeddings *E1* (Luong et al. 2015), *E2* (Faruqui and Dyer 2014), and *E3* (Vulić and Moens 2015)). Embedding *E1* (Luong et al. 2015) (c.f. description in Section 3.1) employs a parallel corpus and alignment information for computing embedding model. If a word W_S is aligned to word W_T , then the context information C_T of target word W_T is also used as context of the source word W_S along with its own context information C_S for computing word vectors. By utilizing the context information of both source and target sides, resultant word embeddings of W_S and W_T are made semantically closer to each other in the vector space. In comparison, embedding *E2* (Faruqui and Dyer 2014) requires two monolingual (Languages $L1$ & $L2$) embeddings and a dictionary containing $L1 \leftrightarrow L2$ mapping. Utilizing the mapping information, it performs Canonical Correlation Analysis (CCA) on two embeddings and projects these into a shared vector space where they are maximally correlated. Embedding *E3* (Vulić and Moens 2015) is an SGNS model computed on aligned documents. It merges the aligned documents together before removing the sentence marker from the merged documents and then performs random shuffling. Finally, an SGNS model is trained on the shuffled data to compute cross-lingual embeddings.

We present the comparative analysis of our proposed method utilizing other bilingual/cross-lingual embeddings in Table 12. Results for the cross-lingual scenario suggest that all three embeddings are quite competitive with each other; however, in a multi-lingual scenario, embedding *E1* reports better performance in almost all cases. It should be noted that our main objective in this work is to reduce the effect of data sparsity, and we observe from the reported results in Table 12 that all three different embeddings validate our hypothesis (c.f. Rows 1 and 2 of Table 12(a)

Table 13. Comparative Analysis of Monolingual Embeddings and Bilingual Embeddings for Aspect Classification in Multi-lingual Setup

Models	Bilingual	Models	Monolingual	
	Size = 7.2M		Size = 7.2M	Size = 53M
Bilingual	62.51	Monolingual	63.64	68.74
Bilingual + Embedding (OOV)	64.83			
Bilingual + Embedding (OOV) + Features (Eng)	76.29	Monolingual + Features (Eng)	70.86	77.74

and Table 12(b)) in each of the problem-scenario pairs (i.e., *aspect-term extraction (multi-lingual & cross-lingual)* and *sentiment classification (multi-lingual & cross-lingual)*). This suggests that handling OOV words in a better way can produce better performance as well.

4.5 Comparison with Regard to Varying Corpus Size for Bilingual Word Embedding Computation

As stated earlier, our prime motivation for this work is to minimize the effect of *data sparsity* while learning through any deep neural network architecture. For this, we propose to use bilingual embeddings computed from a parallel corpus (approximately 7.2M English-Hindi parallel sentences), which is created utilizing an SMT system. Since the SMT system is not fully accurate, some errors are introduced while translating. Also, 7.2M is not a considerably large number in terms of word embedding computation. However, the underlying method performs considerably better compared to state-of-the-art systems, even with all these limitations.

To show the effectiveness of bilingual embeddings in minimizing data sparsity, we also experiment with monolingual Hindi embeddings computed on 53M sentences. Following the proposed approach (except computing embeddings for OOV words), we obtain an accuracy of 77.74% for the aspect classification task. Table 13 shows a comparison of monolingual and multi-lingual approaches for classification. Despite the limitations discussed above (i.e., SMT error and corpus size), the proposed method with bilingual embeddings (76.29%) performs considerably at par against the monolingual embeddings created from a very large corpus of 53M (77.74%). However, the monolingual WE computed using the same amount of corpus (i.e., 7.2M sentences) produces an accuracy of only 63.64%. Furthermore, with the help of lexicon-based features, accuracy of this system increases to 70.86% (compared to 76.29% in our proposed model). It can also be observed that system performance is improved by just including representations of the OOV words. The performance of the proposed system would have been much better without the previously mentioned limitations.

4.6 Error Analysis

We performed detailed error analysis on the results that we obtained.

4.6.1 Aspect Term Extraction. Confusion matrices for multi-lingual and cross lingual setups are depicted in Table 14(a) and 14(b), respectively. We observe that the precision of the *I-ASP* class of multi-lingual (61.2%) and cross-lingual (32.65%) is the main difference between the two setups. In a cross-lingual scenario, our system incorrectly predicts the *I-ASP* class more than twice the number of correct predictions. We also preform qualitative analysis of the predictions as presented next.

Table 14. Aspect Term Extraction: Confusion Matrix

		Predicted	
		I-ASP	O
Actual	I-ASP	616	692
	O	389	84943

a. Multi-lingual setup.

		Predicted	
		I-ASP	O
Actual	I-ASP	447	861
	O	922	84410

b. Cross-lingual setup.

- (1) **Presence of preposition:** We observe that presence of a preposition within an aspect term makes it challenging for the proposed system to make a correct identification of aspect terms. For example, in the following review “इंस्टालेशन|iMsTAleshana” (installation), “यूजर इंटरफ़ेस|yUjara iMTaraFesa” (user interface), and “आवाज़ का स्तर|AvAZ kA stara” (level of voice) are the aspect terms. Our proposed system correctly tags “इंस्टालेशन|iMsTAleshana” (installation) and “यूजर इंटरफ़ेस|yUjara iMTaraFesa” (user interface) as aspect terms. However, for “आवाज़ का स्तर|AvAZ kA stara” (level of voice), it predicts “आवाज़|AvAZ” (voice) as an aspect term but fails to tag the next two tokens “का|kA” (of) and “स्तर|stara” (level), possibly due to the presence of a preposition “का|kA” (of).

Devanagari: इसकी विशेषताओं में आसान इंस्टालेशन, साफ़ सुथरा यूजर इंटरफ़ेस, और अच्छी आवाज़ का स्तर शामिल हैं।

Transliteration: isakI visheShatAoM meM AsAna iMsTAleshana, sAF sutharA yUjara iMTaraFesa, aura achChI AvAZ kA stara shAmila haiM.

Translation: Its features include easy installation, clean user interface, and good level of voice.

	...	अच्छी	आवाज़	का	स्तर	शामिल	हैं
Actual	...	O	I-ASP	I-ASP	I-ASP	O	O
Predicted	...	O	I-ASP	O	O	O	O

4.6.2 Aspect Sentiment Classification. Quantitatively, *neutral* is the most problematic class in both multi-lingual and cross-lingual setups. It mainly confuses with *positive* class. Approximately, 20% and 40% of the *neutral* instances are tagged as *positive*, respectively, in multi-lingual and cross-lingual setups. In comparison, false positives of *neutral* with regard to *positive* also account for approximately 20% in both the setups. Our system does not predict the *conflict* class at all, possibly due to the insufficient number of training instances. To improve the performance of conflict class, we also perform *conflict-vs-all* training by balancing the two classes (i.e., conflict and others). However, only 1 conflict instance is correctly identified with a F-measure value of 1.2%. Table 15 and Table 16 depict the confusion matrices for multi-lingual and cross-lingual setups, respectively.

Qualitatively, the following are the few problematic cases where our system continuously performs below par.

Table 15. Aspect Classification in Multi-lingual Setup:
Confusion Matrix

		Predicted			
		Positive	Negative	Neutral	Conflict
Actual	Positive	328	10	82	0
	Negative	11	72	16	0
	Neutral	70	22	288	0
	Conflict	1	1	1	0

Table 16. Aspect Classification in Cross-lingual Setup:
Confusion Matrix

		Predicted			
		Positive	Negative	Neutral	Conflict
Actual	Positive	303	37	80	0
	Negative	4	85	10	0
	Neutral	152	72	156	0
	Conflict	0	2	1	0

- (1) **Lack of polar information inside context.** Our system finds it challenging to classify the sentiment of those aspect terms whose polar information lies outside the context window. In the following sentence, the aspect term is “वज़न|weight” and the actual sentiment toward it is positive. The polar information “तुलना में लगभग आधा|about half as compared” and “हल्का|lighter” are far from the aspect term, hence, not captured within the context window.

Devanagari: इसका वज़न नए आईपैड की तुलना में लगभग आधा है और यह अन्य उपलब्ध 7-इंच टेबलेट्स से भी हल्का है।

Transliteration: isakA vaZana nae AIpaiDa kI tulana meM lagabhaga AdhA hai aura yaha anya upalabdha 7-iMcha TebaleTsa se bhi halkA hai.

Translation: Its weight is about half as compared to the new iPad and it is lighter than other available 7-inch tablets.

- (2) **Implicit sentiment:** The presence of implicit sentiment is not correctly classified in our proposed system. The following review contains “बनावट|built” as an aspect term and its negative sentiment is derived from the phrase “प्लास्टिक फील|plastic feel.”

Devanagari: इस टेबलेट की बनावट काफी प्लास्टिक फील देता है।

Transliteration: isa TebaleTa kI banAvaTa kAphi plAsTika phila detA hai.

Translation: The built of this tablet gives a fairly plastic feel.

5 CONCLUSION

ABSA in resource-poor languages (e.g., Hindi) is not much matured in comparison to resource-rich languages such as English due to a lack of sufficient resources. In our study, we proposed a novel approach to increase the footprint of word representations for two subtasks of ABSA (i.e., aspect term extraction [opinion target extraction] and aspect sentiment classification). Specifically, we addressed the representation of OOV words using bilingual word embeddings. Evaluation results suggest that the handling of OOV words in a systematic manner offers improvement over the baselines and various state-of-the-art systems in two different setups, cross-lingual and multi-lingual.

Finally, we conclude with the following observations: (i) A systematic approach to leverage the qualitative resources of the resource-rich language can indeed improve the performance of the system in resource-constrained language; (ii) our results verify that the driving force for a sentiment analyzer are the sentiment lexicons; (iii) the sequence of hand-crafted word-level features can be better learned through a separate LSTM rather than just concatenation; and (iv) late fusion of hand-crafted features in the neural network is comparatively better than early fusion. Our future work focuses on evaluating the proposed approach with other languages and investigating methods for exploiting the benefits of DL as well as traditional supervised methods for ABSA.

ACKNOWLEDGMENTS

Asif Ekbal acknowledges Young Faculty Research Fellowship (YFRF), supported by Visvesvaraya PhD scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia).

REFERENCES

- Md Shad Akhtar, Deepak, Asif Ekbal, and Pushpak Bhattacharyya. 2017. Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. *Knowledge-Based Systems* 125 (2017), 116–135.
- Md Shad Akhtar, Asif Ekbal, and Pushpak Bhattacharyya. 2016a. Aspect based sentiment analysis in Hindi: Resource creation and evaluation. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, May 23–28, 2016. European Language Resources Association (ELRA), Portoro, Slovenia, 2703–2709.
- Md Shad Akhtar, Ayush Kumar, Asif Ekbal, and Pushpak Bhattacharyya. 2016b. A hybrid deep learning architecture for sentiment analysis. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016): Technical Papers*, December 11–16, 2016. Osaka, Japan, 482–493.
- Md Shad Akhtar, Palaash Sawant, Sukanta Sen, Asif Ekbal, and Pushpak Bhattacharyya. 2018. Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, 572–582. <http://aclweb.org/anthology/N18-1053>.
- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, May 17–23, 2010 (17–23). European Language Resources Association (ELRA), Valletta, Malta, 2200–2204.
- Dzmitry Bahdanau, Tom Bosc, Stanislaw Jastrzebski, Edward Grefenstette, Pascal Vincent, and Yoshua Bengio. 2017. Learning to compute word embeddings on the fly. *CoRR* abs/1706.00286 (2017). arxiv:1706.00286 <http://arxiv.org/abs/1706.00286>.
- Akshat Bakliwal, Piyush Arora, and Vasudeva Varma. 2012. Hindi subjective lexicon: A lexical resource for Hindi polarity classification. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, May 21–27, 2012. Istanbul, Turkey, 1189–1196.
- A. R. Balamurali, Aditya Joshi, and Pushpak Bhattacharyya. 2012. Cross-lingual sentiment analysis for Indian languages using linked wordnets. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING): Posters*, 8–15 December 2012. Mumbai, India, 73–82.
- Jeremy Barnes, Patrik Lambert, and Toni Badia. 2016. Exploring distributional representations and machine translation for aspect-based cross-lingual sentiment classification. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016): Technical Papers*, December 11–16, 2016. Osaka, Japan, 1613–1623.
- Pushpak Bhattacharyya. 2010. IndoWordnet. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*. Valletta, Malta, 3785–3792.
- Maryna Chernyshevich. 2014. IHS R&D belarus: Cross-domain extraction of product features using conditional random fields. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, August 23–24, 2014. Dublin, Ireland, 309–313.
- Amitava Das and Sivaji Bandyopadhyay. 2010. SentiWordNet for indian languages. In *Proceedings of the 8th Workshop on Asian Federation for Natural Language Processing*, August 2010. Beijing, China, 56–63.
- Bhuwan Dhingra, Hanxiao Liu, Ruslan Salakhutdinov, and William W. Cohen. 2017. A comparative study of word embeddings for reading comprehension. *CoRR* abs/1703.00993 (2017). arxiv:1703.00993 <http://arxiv.org/abs/1703.00993>.
- Xiaowen Ding, Bing Liu, and Philip S. Yu. 2008. A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM'08)*. ACM, New York, 231–240.

- Manaal Faruqui and Chris Dyer. 2014. Improving vector space word representations using multilingual correlation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Gothenburg, Sweden, 462–471. <http://www.aclweb.org/anthology/E14-1049>.
- J. R. Firth. 1957. A synopsis of linguistic theory 1930–55. *Studies in Linguistic Analysis (special volume of the Philological Society)* 1952–59 (1957), 1–32.
- Deepak Kumar Gupta, Kandula Srikanth Reddy, Asif Ekbal, et al. 2015. PSO-ASent: Feature selection using particle swarm optimization for aspect based sentiment analysis. In *Natural Language Processing and Information Systems (NLDB 2015)*, June 17–19 2015. Springer, Passau, Germany, 220–233.
- Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Madrid, Spain, 174–181.
- Kenneth Heafield. 2011. KenLM: Faster and smaller language model queries. In *Proceedings of the 6th Workshop on Statistical Machine Translation*. Association for Computational Linguistics, 187–197.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 168–177.
- V. S. Jagtap and Karishma Pawar. 2013. Analysis of different approaches to sentence-level sentiment classification. *International Journal of Scientific Engineering and Technology (ISSN: 2277-1581)* 2 (2013), 164–170.
- Aditya Joshi, A. R. Balamurali, and Pushpak Bhattacharyya. 2010. A fall-back strategy for sentiment analysis in Hindi: A case study. In *Proceedings of the 8th International Conference on Natural Language Processing (ICON 2010)*. Kharagpur, India.
- Rasoul Kaljahi and Jennifer Foster. 2016. Detecting opinion polarities using kernel methods. In *Proceedings of the Workshop on Computational Modelling of People’s Opinions, Personality, and Emotions in Social Media*. Osaka, Japan, 60–69.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics*. Association for Computational Linguistics, 1367.
- Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR* abs/1412.6980 (2014). <http://arxiv.org/abs/1412.6980>.
- Reinhard Kneser and Hermann Ney. 1995. Improved backing-off for m-gram language modeling. In *Proceedings of the 1995 International Conference on Acoustics, Speech, and Signal Processing (ICASSP’95)*, Vol. 1. IEEE, 181–184.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*. Association for Computational Linguistics, 177–180.
- Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 48–54.
- Ayush Kumar, Sarah Kohail, Asif Ekbal, and Chris Biemann. 2015. IIT-TUDA: System for sentiment analysis in Indian languages using lexical acquisition. In *Mining Intelligence and Knowledge Exploration*. Springer, 684–693.
- Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Bilingual word representations with monolingual quality in mind. In *NAACL Workshop on Vector Space Modeling for NLP*. Denver, United States, 151–159.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv Preprint arXiv:1301.3781* (2013).
- Saif M. Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2016. How translation alters sentiment. *Journal of Artificial Intelligence Research* 55, 1 (Jan. 2016), 95–130.
- Arjun Mukherjee and Bing Liu. 2012. Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1 (ACL’12)*. 339–348.
- Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, 160–167.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics* 29, 1 (2003), 19–51.
- Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL’05)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 115–124.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar, 1532–1543.

- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland, 27–35.
- Soujanya Poria, Erik Cambria, and Alexander Gelbukh. 2016. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems* 108 (2016), 42–49.
- Prerana Singhal and Pushpak Bhattacharyya. 2016. Borrow a little from your rich cousin: Using embeddings and polarities of English words for multilingual sentiment classification. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016): Technical Papers, December 11-16, 2016*. Osaka, Japan, 3053–3062.
- Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15 (2014), 1929–1958.
- Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the 7th Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4 (CONLL'03)*. 142–147.
- Zhiqiang Toh and Wenting Wang. 2014. DLIREC: Aspect term extraction and term polarity classification system. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland, 235–240.
- P. D. Turney. 2002. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Association for Computational Linguistics (ACL)*. Philadelphia, USA, 417–424.
- Ivan Vulić and Marie-Francine Moens. 2015. Bilingual word embeddings from non-parallel document-aligned data applied to bilingual lexicon induction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Association for Computational Linguistics, Beijing, China, 719–725. <http://www.aclweb.org/anthology/P15-2118>.
- Joachim Wagner, Piyush Arora, Santiago Cortes, Utsab Barman, Dasha Bogdanova, Jennifer Foster, and Lamia Tounsi. 2014. DCU: Aspect-based polarity classification for SemEval task 4. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland, 223–229.
- Janyce Wiebe and Rada Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics (COLING) and the 44th Annual Meeting of the Association for Computational Linguistics (ACL) (ACL-44)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 1065–1072.
- Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. 2016. Cross-lingual sentiment classification with bilingual document representation learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*. Berlin, Germany, 1403–1412.
- Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie review mining and summarization. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM'06)*. ACM, New York, 43–50.

Received July 2017; revised May 2018; accepted August 2018