# Leveraging Multilingual Resources for Language Invariant SentimentAnalysis

by

Allen Antony, Arghya Bhattacharya, Jaipal Singh Goud, Radhika Mamidi

Report No: IIIT/TR/2020/-1

Centre for Language Technologies Research Centre
International Institute of Information Technology
Hyderabad - 500 032, INDIA
November 2020

# Leveraging Multilingual Resources for Language Invariant Sentiment Analysis

**Allen J. Antony** **Arghya Bhattacharya** **Jaipal Singh Goud** **Radhika Mamidi**
Language Technologies Research Centre
International Institute of Information Technology - Hyderabad
Gachibowli, Hyderabad, Telangana-500032
{allen.antony@research., arghya.b@research.,
jaipal.singh@research.,radhika.mamidi@}iiit.ac.in

## Abstract

Sentiment analysis is a widely researched NLP problem with state-of-the-art solutions capable of attaining human-like accuracies for various languages. However, these methods rely heavily on large amounts of labelled data or sentiment weighted language specific lexical resources that are unavailable for low-resource languages. Our work attempts to tackle this data scarcity issue by introducing a neural architecture for language invariant sentiment analysis capable of leveraging various monolingual datasets for training without any kind of cross-lingual supervision. The proposed architecture attempts to learn language agnostic sentiment features via adversarial training on multiple resource-rich languages which can then be leveraged for inferring sentiment information at a sentence level on a low resource language. Our model outperforms the current state-of-the-art methods on the Multilingual Amazon Review Text Classification dataset (Prettenhofer and Stein, 2010) and achieves significant performance gains over prior work on the low resource Sentiraama corpus (Gangula and Mamidi, 2018). A detailed analysis of our research highlights the ability of our architecture to perform significantly well in the presence of minimal amounts of training data for low resource languages.



## 1 Introduction

Sentiment analysis refers to a series of methods, techniques, and tools aimed at extracting the intended sentiment from a written opinion. Traditional sentiment analysis techniques have relied on using supervised term weighting methods including terms' distribution of classes, word-level polarity scoring and using SVMs (Durant and Smith, 2006) and Naive Bayes classifiers (Prasad, 2010) for pattern extraction using hand-crafted features. The advent of deep learning techniques for sentiment analysis has now enabled the extraction of high quality sentiment data from written texts. One majorly overlooked factor in the performance of these neoteric approaches is their dependency on large annotated datasets compiled from multiple data sources related to or sourced from newspapers, tweets, photos and product reviews. (Socher et al., 2013; Kim, 2014; Tai et al., 2015; Iyyer et al., 2015; Wang et al., 2016).

Given global nature of the current information sharing infrastructure, most data generated belongs to one of the three languages : English, Mandarin or Spanish. This abundance of raw data aids and motivates the creation of annotated resources in these languages. Conversely, the paucity of annotated data in most languages makes it a challenging task to develop deep learning based solutions for them. Hence there is a pressing need to pay special attention to developing solutions capable of sentiment analysis in a low resource setting.

Some of the initial methods that attempt to tackle this problem of data scarcity using transfer learning (training a neural model on one language and applying the trained model on another language via weight sharing) do not perform well due to the limited overlap between the vocabularies of the different languages and difference in their syntactic structure (Chen et al., 2018b).

*Cross-lingual sentiment classification* (CLSC) methods try to alleviate this problem by leveraging labeled data from one language to improve the performance on another language (Bel et al., 2003). However, these methods typically rely on auxiliary cross-lingual resources such as a parallel corpora (Yarowsky et al., 2001; Xu and Yang, 2017), bilingual lexicons (Mihalcea et al., 2007) or the use of machine translation systems (Kanayama et al., 2004; Wan, 2009; Prettenhofer and Stein, 2010; Can et al., 2018). Unfortunately, the curation of such cross-lingual resources is both a time and a labour intensive task. Hence, there is a need for architectures that can perform well in the absence of such cross-lingual resources.

In this paper, we address this problem by presenting a neural *Language Invariant Sentiment Analyzer* (LISA) architecture that is capable of training on multiple monolingual sentiment labelled datasets to learn language agnostic sentiment features that can be transferred to perform sentiment analysis in low-resource languages **without leveraging any form of cross-lingual supervision**.

**Approach :** We formulate this problem as a *multi-lingual transfer learning* (MLTL) language adaptation task where we attempt to learn language agnostic sentiment features via adversarial training on labelled documents $(s_1, s_2...s_n)$ from multiple (source) languages to improve the performance on documents $(t_1, t_2...t_m)$ from a low resource (target) language. The key components of our approach include learning monolingual word embeddings from $s_1, s_2...s_n, t_1, t_2...t_m$ and projecting them to a shared multilingual semantic space. We employ an LSTM network to learn latent features $(z)$ from this multilingual space which is then used by a sentiment classifier $(\mathcal{S_C})$ to predict the sentiment polarity of a document $d \in \{s_1...s_n, t_1...t_m\}$. Concurrently, a language classifier $(\mathcal{C_L})$ is trained to predict the language of document $d$ based on $z$. During the adversarial training we try to minimize the binary cross-entropy loss of $\mathcal{C_S}$, while at the same time we maximize the cross-entropy loss of $\mathcal{C_L}$. This results in a setting where the LSTM learns to produce latent features $z$ that predicts the sentiment of document $d$ correctly independent of the language of document $d$. We hypothesize that in this setting, the latent features $(z)$ trained would contain sentiment features that are language agnostic.

In summary, the main contributions of this paper are :

- We introduce a language independent neural architecture for sentiment analysis without the use of language specific features or cross-lingual supervision.

- We provide extensive evaluations of the LISA architecture in two settings :
  (i) **Low-resource Setting** : Where labeled data in the target language is available in limited amounts.
  (ii) **No-resource Setting** : Where the is no labeled data available in the target language.

- Our experiments on the Multilingual Amazon Review Text Classification dataset and the Sentiraama dataset show that the proposed LISA architecture achieves better performance compared to prior work in the low-resource setting.

The paper is structured as follows : Section 2 highlights the related prior work in the field of CLSC. Section 3 introduces the datasets that are used in our experiments. Section 4 presents the methodology used to align multiple monolingual semantic spaces to a common multilingual semantic space. Section 5 describes in detail the various components of the LISA architecture. Section 6 explains the adversarial training methodology employed. Section 7 describes our experimental setup and provides a detailed comparison of our approach with prior work in both the low-resource and no-resource setting. Section 8 addresses the advantages and shortcomings of the proposed approach and state our concluding remarks.

## 2 Background and Related Work

**CLSC using Machine Translation Systems** : The most straightforward approach in CLSC involves using machine translation systems to translate sentences, words, phrases or documents in the target language to the source language and then learning a classifier in the source language to predict the sentiment (Kanayama et al., 2004; Wan, 2008; Wan, 2009; Banea et al., 2010; Lu et al., 2011; Can et al., 2018). The baseline CL-MT (Prettenhofer and Stein, 2010) method uses this technique by using Google Translate[1] to translate documents in the target language to the source language and learns a classifier in the source language using the bag-of-words features. Similarly, the **BiDRL** model (Zhou et al., 2016) used Google Translate and employed a joint learning approach to simultaneously learn both word and document representations in both source and target language which are then used for sentiment classification. However, these methods are overly reliant on the performance of the machine translation system utilized, which in many cases, are less than satisfactory.

**CLSC using cross-lingual resources** : Most popular methods in CLSC makes use of cross-lingual resources to bridge the language barrier and induce inter-language correspondence. Bel et al. (2003) used a bilingual dictionary to translate documents in the target language to the source language and trained a classifier in the source language for text classification. Mihalcea et al. (2007) used a bilingual lexicon to translate subjective words and phrases in the source language into the target language. Shi et al. (2010) utilizes a bilingual dictionary to translate the classification model from a source language to a target language rather than the documents themselves. Balamurali et al. (2012) used WordNet senses as features for CLSA in Indian languages (Hindi and Marathi). The **CLMM** model (Meng et al., 2012) treated the source language and the target language words in an unlabeled bilingual parallel dataset as generated simultaneously by a set of mixture components. The **CR-RL** approach (Xiao and Guo, 2013) learned word embeddings by using a set of bilingual word pairs where one part of the word vector contains language specific features and the other part contains language independent features. **CL-SCL** model (Prettenhofer and Stein, 2010) leveraged structural correspondence learning with the help of a bilingual dictionary to learn a source-target feature space. Pham et al. (2015) used a parallel corpus between the source language and the target language to learn bilingual paragraph vectors (**Bi-PV**). **UMM** (Xu and Wan, 2017) learned multilingual sentiment-aware word representations based on unlabeled parallel data and used pivot languages to transfer sentiment information in the absence of parallel data . The **CLDFA** approach (Xu and Yang, 2017) adopted cross-lingual distillation and adversarial techniques on parallel corpora for CLSC. Our work draws inspiration from the **ADAN-GRL** model (Chen et al., 2018b) which employed language adversarial training to learn language invariant features from bilingual word embeddings (BWE) which were created using a parallel corpus. In fact, our proposed model can be considered as a cross-lingually unsupervised variant of the **ADAN-GRL** model as we do not rely on parallel corpora to learn word representations. Furthermore, the **ADAN-GRL** model is limited by the BWE to only incorporate two language pairs (source and target) during training, whereas our LISA system is capable of leveraging multiple source languages and the target language for adversarial training.

**CLSC without cross-lingual supervision** Neoteric advances by Chen et al. (2018a) alleviates the need for cross-lingual resources by introducing a shared-private Mixture-of-Experts model (**MoE**) that learns both language specific features and language invariant features without cross-lingual supervision. Our work, although related to **MoE** in objective with respect to the lack of cross-lingual supervision, differs in the methodology. Direct comparison of our architecture against **MoE** (Table 4) proves that the (language invariant) features extracted by our architecture contains more sentiment related information than the (language specific + language invariant) features extracted by **MoE**.

## 3 Dataset Description

We conduct our experiments on two publicly available sentiment classification datasets :

**The Multilingual Amazon Review Text Classification dataset** (Prettenhofer and Stein, 2010) consists of sentiment labelled data in multiple languages. The vast amount of prior work on this dataset helps us to directly compare our results with the pre-existing state-of-the-art CLSC methods.

**The Sentiraama Corpus** (Gangula and Mamidi, 2018) is a real-world low resource sentiment corpus in Telugu (an agglutinating Indian language). We use this dataset to test the robustness of our system and evaluate our results in a truly low resource setting.

In the following subsections we describe both the corpora in detail.

[1] https://translate.google.com/

### 3.1 Multilingual Amazon Review Text Classification dataset

The Multilingual Amazon Review Dataset contains sentiment labeled product reviews in four languages (English, German, French and Japanese) across three domains (Books, Dvd and Music). The German, French and Japanese reviews were crawled from Amazon and the corpus was enhanced with English reviews from Blitzer et al. (2007). Each review contains a domain label, a review summary, a review text, and a rating from the set $\{1, 2, 4, 5\}$ where $\{1, 2\}$ denotes negative sentiment and $\{4, 5\}$ denotes positive sentiment. The reviews in each domain for each language are split into three disjoint balanced sets, namely, Train set, Test set and Unlabeled set. The dataset statistics are presented in Table 1.

| | | Train | Test | Unlabelled |
|---|---|---|---|---|
| English | Books | 2000 | 2000 | 50000 |
| | DVD | 2000 | 2000 | 30000 |
| | Music | 2000 | 2000 | 25220 |
| German | Books | 2000 | 2000 | 165470 |
| | DVD | 2000 | 2000 | 91516 |
| | Music | 2000 | 2000 | 60392 |
| French | Books | 2000 | 2000 | 32870 |
| | DVD | 2000 | 2000 | 9358 |
| | Music | 2000 | 2000 | 15940 |
| Japanese | Books | 2000 | 2000 | 169780 |
| | DVD | 2000 | 2000 | 68326 |
| | Music | 2000 | 2000 | 55892 |

**Table 1:** Multilingual Amazon Review Text Classification dataset statistics.

### 3.2 Sentiraama Dataset

The Sentiraama dataset consists of sentiment labelled documents in four domains : Books, Movies, Products and Song Lyrics. Each document is given a positive or a negative label. The corpus statistics are presented in Table 2.

| | Books | Movies | Products | Lyrics |
|---|---|---|---|---|
| Positive | 100 | 136 | 100 | 230 |
| Negative | 100 | 131 | 100 | 109 |
| Total | 200 | 267 | 200 | 339 |

**Table 2:** Sentiraama corpus statistics.

To avoid cross-domain discrepancies we restrict our experiments to the Books and Movies domain as it has similar counterparts in the Multilingual Amazon Review Dataset, i.e, Books and Dvd respectively. We divide the Books and Movie domains of the Sentiraama dataset to create a Train set and a Test set using an 80-20 train-test split. The statistics of the subset of the corpus that are used in our experiments are listed in Table 3.

| | Books | | Movies | |
|---|---|---|---|---|
| | +ve | -ve | +ve | -ve |
| Train | 80 | 80 | 108 | 105 |
| Test | 20 | 20 | 28 | 26 |

**Table 3:** Subset of the Sentiraama corpus used in our experiments.

## 4 Multilingual Word Representation

For our experiments, we train fastText embeddings (Bojanowski et al., 2017) to project each word to a monolingual semantic space for each language in the datasets described in Section 3. We then employ the unsupervised MUSE approach (Conneau et al., 2017) to align the monolingual spaces of each language in an adversarial manner to a common multilingual semantic space. While training MUSE we use English as the target semantic space and align all the other monolingual semantic spaces to this space. Let $\mathcal{X} = \{x_1, x_2, \dots x_a\}$ and $\mathcal{Y} = \{y_1, y_2, \dots y_b\}$ be the source and target fastText word embeddings respectively. Let $W$ be a linear mapping from $\mathcal{X}$ to $\mathcal{Y}$. A discriminator is trained to discriminate between elements randomly sampled from $W\mathcal{X}$ and $\mathcal{Y}$ while $W$ (which acts as the generator) is jointly trained to fool the discriminator. The discriminator loss function $\mathcal{L}_D(\theta_D|W)$ is formulated as:

$$\mathcal{L}_D(\theta_D|W) = -\frac{1}{a}\sum_{i=1}^{a} logP_{\theta_D}(source = 1|Wx_i) - \frac{1}{b}\sum_{i=1}^{b} logP_{\theta_D}(source = 0|y_i)$$

The Mapping objective function used to train $W$ is given by:

$$\mathcal{L}_W(W|\theta_D) = -\frac{1}{a}\sum_{i=1}^{a} logP_{\theta_D}(source = 0|Wx_i) - \frac{1}{b}\sum_{i=1}^{b} logP_{\theta_D}(source = 1|y_i)$$

Where $\theta_D$ denotes the discriminator parameters and $P_{\theta_D}(source = 1|z)$ is the probability that a

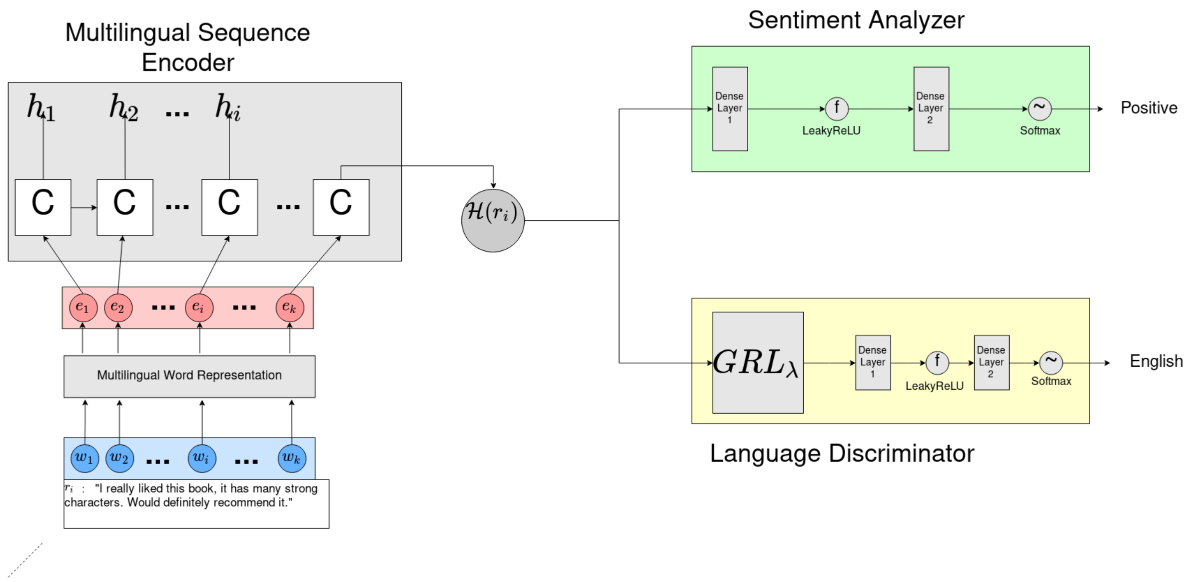


**Figure 1:** The LISA architecture.

vector $z$ is the mapping of a source embedding according to the discriminator.

Next, a synthetic parallel vocabulary consisting of the most frequent words and their mutual nearest neighbors are extracted from the resulting shared embedding space $W$ to fine-tune the mapping using the closed-form Procrustes solution (Schönemann, 1966) given by:

$$W^* = \underset{W \in O_d(\mathbb{R})}{argmin} \|WX - Y\|_F = UV^T$$
$$\text{with } U\Sigma V^T = \text{SVD}(YX^T)$$

Where $X$ and $Y$ are two aligned matrices containing the embeddings of the words in the trained space $W$, $d$ represents the dimension of the embeddings, $O_d(\mathbb{R})$ is the space of $d$ x $d$ matrices of real numbers with the orthogonality constraint and SVD($YX^T$) represents the singular value decomposition of $YX^T$.

# 5 LISA Architecture

The input to the LISA model is a review $r_i$ that is made up of a sequence of words $w_1, w_2, \ldots w_k$. Each review $r_i$ is associated with a language label $l_i \in L$ where $L = \{l_1, l_2, \ldots l_p\}$ is the set of all language labels used in training. Additionally, each review $r_i$ is also associated with a sentiment label $t_i \in \{positive, negative\}$ which denotes the sentiment polarity of the review. We project each word $w_i$ to the multilingual semantic space (from section 4) to obtain a sequence of $n$-dimensional word embeddings $e_1, e_2, \ldots e_k$ where $e_i \in \mathbb{R}^n$.

The following subsections describe in detail the individual components of the LISA architecture. Figure 1 shows the overall architecture of the proposed model.

## 5.1 Multilingual Sequence Encoder ($\mathcal{H}$)

The Multilingual Sequence Encoder ($\mathcal{H}$) processes the sequence of word embeddings $(e_1, e_2, \ldots e_k)$ and transforms it into an $m$-dimensional (hidden) vector $\mathcal{H}(r_i)$. To this end, the embeddings for all the words in review $r_i$ are passed sequentially through a Long Short-Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997). LSTMs are a variant of RNNs that learns features that model the long-term dependencies between the words. The LSTM network, at each time step outputs a hidden state $h_i$ for every input word embedding $e_i$, such that :

$$h_i = \text{LSTM}(e_i, h_{i-1}) \in \mathbb{R}^m$$

The final hidden state $\mathcal{H}(r_i) = h_k$ is then passed through a Language Discriminator ($\mathcal{C_L}$) and a Sentiment Analyzer ($\mathcal{C_S}$).

## 5.2 Language Discriminator ($\mathcal{C_L}$)

The goal of the Language Discriminator ($\mathcal{C_L}$) is to predict the language label $l_i$ based on $\mathcal{H}(r_i)$. In other words, $\mathcal{C_L}$ tries to predict the language from which the sequence of words $w_1, w_2, \ldots w_k$ come from. The $\mathcal{C_L}$ comprises of a Gradient Reversal Layer ($GRL_\lambda$), followed by two Dense Layers and an output Softmax Layer that applies the softmax function over all the languages used in training. During backpropagation, $GRL_\lambda$ multiplies the gradients by a factor of $-\lambda$ and during the forward pass it acts as the identity function. $\lambda$ is hyperparameter in the network.

### 5.3 Sentiment Analyzer ($\mathcal{C_S}$)

The Sentiment Analyzer ($\mathcal{C_S}$), as the name suggests, tries to predict the sentiment label $t_i$ of the input review $r_i$ based on $\mathcal{H}(r_i)$ . The $\mathcal{C_S}$ is made up of two Dense Layers followed by an output Softmax Layer that applies the the softmax function over the two sentiment polarities (positive and negative).

## 6 Adversarial Training

Inspired by recent works (Goodfellow et al., 2014; Ganin et al., 2016; Beutel et al., 2017), we train the LISA model using adversarial training on a set of labeled reviews $R = \{r_1, r_2, \ldots r_n\}$. The aim of the LISA model is to predict the sentiment label $t_i$ for a given review $r_i$ independent of the language label $l_i$.

We formulate the learning objective in a way that minimizes the sentiment classification loss from $\mathcal{C_S}$ and maximizes the language classification loss from $\mathcal{C_L}$. As a result, the LISA model tries to jointly optimize the below functions:

$$\arg\min_{\mathcal{H},\mathcal{C_S}} f(\mathcal{C_S}(\mathcal{H}(r_i)), t_i) - f(\mathcal{C_L}(\mathcal{H}(r_i)), l_i) \tag{1}$$

$$\arg\max_{\mathcal{C_L}} f(\mathcal{C_L}(\mathcal{H}(r_i)), l_i) \tag{2}$$

Where $f$ denotes the loss function used. This results in a setting where the $\mathcal{C_L}$ tries to predict $l_i$ based on a given $\mathcal{H}(r_i)$ and the encoder $\mathcal{H}$ tries to 'fool' the $\mathcal{C_L}$ by learning to create $\mathcal{H}(r_i)$ that is minimally influenced by the language label $l_i$ while at the same time, is maximally influenced by the $\mathcal{C_S}$ to predict he sentiment label $t_i$ correctly.

The M-LiST model (Goud et al., 2019) presents a similar setting for the task of open domain event detection that was trained using a Gradient Reversal Layer $GRL_\lambda$ (Ganin et al., 2016) between $\mathcal{H}$ and $C_L$. By using $GRL_\lambda$, the optimization functions (equations 1 and 2) can be simplified as :

$$\arg\min_{\mathcal{H},\mathcal{C_S},\mathcal{C_L}} f(\mathcal{C_S}(\mathcal{H}(r_i)), t_i) + f(\mathcal{C_L}(GRL_\lambda(\mathcal{H}(r_i))), l_i) \tag{3}$$

## 7 Experiments and Results

In this section we present an extensive set of experiments conducted on the Multilingual Amazon Review Text Classification dataset and the Telugu Sentiraama sentiment classification corpus. We evaluate our approach in the two settings described below :

**Low-resource setting** : We evaluate the performance of the LISA architecture in the low-resource setting (termed **LISA-LR**) by training it on the Train sets from multiple source languages and the limited Train set in the target language and then testing on the Test set of the target language.

**No-resource setting** : In the no-resource setting, we assume that the training data is not available for the target language. We train the LISA model (termed **LISA-NR**) on the Train sets of the source languages and evaluate the model on the target language Test set.

**LISA - No Language Discriminator** : To show the effectiveness of the Language Discriminator ($\mathcal{C_L}$), we conduct ablation experiments in the low-resource setting where we remove $\mathcal{C_L}$ from the LISA architecture. In this variant of the LISA model (termed **LISA-NoLD**), the Sentiment Analyzer only depends on the MUSE embeddings to learn $\mathcal{H}(r_i)$ to learn sentiment features. Our experiments show that **LISA-LR** performs significantly better in most cases than **LISA-NoLD**.

For the Multilingual Amazon Review Text Classification dataset in the low-resource setting, we train **LISA-LR** on the Train sets of all the four languages. We then test it on the Test set of the target language. In the no-resource setting, we train **LISA-NR** on the Train sets of three languages and test it on the Test set of the fourth language. We do this for each domain in the corpus independently. We compare our results against prior state-of-the-art methods that uses Machine Translation Systems (**CL-MT** and **BiDRL**), methods that leverage cross-lingual supervision (**UMM**, **Bi-PV**, **CR-RL** and **CL-SCL**) and the cross-lingually unsupervised **MAN-MoE** method of Chen et al. (2018a). The results are presented in Table 4.

For the Sentiraama Corpus in the low-resource setting, we train **LISA-LR** by leveraging the Train sets of all the languages in the Multilingual Amazon dataset along with the Sentiraama Train Set. We then test the system on the Sentiraama Test set.

| | German | | | | French | | | | Japanese | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Books | DVD | Music | | Books | DVD | Music | | Books | DVD | Music |
| CL-MT | 79.68 | 77.92 | 77.22 | | 80.76 | 78.83 | 75.78 | | 70.22 | 71.30 | 72.02 |
| BiDRL | 84.14 | 84.05 | 84.67 | | 84.39 | 83.60 | 82.52 | | 73.15 | 76.78 | 78.77 |
| UMM | 81.65 | 81.27 | 81.32 | | 80.27 | 80.27 | 79.41 | | 71.23 | 72.55 | 75.38 |
| Bi-PV | 79.51 | 78.60 | 82.45 | | 84.25 | 79.60 | 80.09 | | 71.75 | 75.40 | 75.45 |
| CR-RL | 79.89 | 77.14 | 77.27 | | 78.25 | 74.83 | 78.71 | | 71.11 | 73.12 | 74.38 |
| CL-SCL | 79.50 | 76.92 | 77.79 | | 78.49 | 78.80 | 77.92 | | 73.09 | 71.07 | 75.11 |
| MAN-MoE | 82.40 | 78.80 | 77.15 | | 81.10 | 84.25 | 80.90 | | 62.78 | 69.10 | 72.60 |
| LISA-LR | **85.45** | **84.90** | **86.55** | | **86.25** | **85.35** | **85.60** | | **79.20** | **83.30** | 80.892 |
| LISA-NR | 55.60 | 55.50 | 58.90 | | 68.95 | 70.65 | 64.30 | | 62.20 | 56.50 | 59.80 |
| LISA-NoLD | 81.20 | 77.70 | 80.75 | | 82.80 | 80.10 | 80.50 | | 79.05 | 83.15 | **82.542** |

**Table 4:** Results on the Multilingual Amazon Review Text Classification dataset. The numbers denote binary classification accuracies.

In the no-resource setting, **LISA-NR** only utilizes the Train set of all the languages in the Multilingual Amazon dataset and test the system on the Sentiraama Test set. We do this for the Books and Movies domain separately. We evaluate the results of **LISA-LR**, **LISA-NR** and **LISA-NoLD** against the Bernoulli Naive Bayes (Rish and others, 2001) and SVM (Joachims, 1998) baselines that use TF-IDF features which were set by Gangula and Mamidi (2018). The experimental results are given in Table 5

| | Books | Movies |
|---|---|---|
| SVM | 55 | 51.851 |
| Naive Bayes | 65 | 75.9 |
| LISA-LR | **72.5** | **85.185** |
| LISA-NR | 57.5 | 57.407 |
| LISA-NoLD | 67.5 | 68.51 |

**Table 5:** Results on the Sentiraama Dataset. The numbers denote binary classification accuracies. Note that the Naive Bayes and SVM accuracies presented in the table differ from the ones presented by Gangula and Mamidi (2018). We attribute this to the difference in the train/test splits and the the lack preprocessing guidelines which makes it hard to adequately replicate their results.

## 8 Analysis and Conclusion

**Analysis** : The results on the Multilingual Amazon Review Text Classification dataset proves our hypothesis that our model learns language invariant features that can be generalized across languages. The empirical results in Table 4 show that our model outperforms pre-existing state-of-the-art methods on this dataset. While our experiments on the Sentiraama dataset proves that our model can be applied in a real-world setting to enhance sentiment retrieval in a truly low resource language. The ablation experiments (LISA-NoLD vs LISA-LR) show that between language pairs that have similar syntactic structure (example : English, French and German), LISA-LR performs much better than LISA-NoLD. This shows the the performance gains over prior work are not just due to the use of MUSE embeddings. Rather, they are attributed to the adversarial training of the Language Discriminator and the Sentiment classifier that extracts language agnostic sentiment features from the MUSE semantic space. But for Japanese (which is dissimilar with respect to other languages in the corpus), the results show that LISA-LR does not have a significant boots over LISA-NoLD. This is because our language adversarial training will retain only features that are invariant across all four languages, which is restrictive such that the information learnt will be too sparse to be useful. Finally, the poor performance of LISA-NR shows that our approach cannot be used for Zero-Shot learning but will achieve state-of-the-art performance in the presence of limited amounts of data.

**Conclusions** : In this paper, we present the LISA model which focuses on exploiting language invariant features for multilingual sentiment analysis without any form of cross-lingual supervision. We back our claims by conducting a wide range of experiments over the Multilingual Amazon Review Text Classification dataset and the Sentiraama dataset which is a real-world low resource dataset. We show that our model outperforms not only the existing cross-lingually unsupervised methods but also methods that rely on strong cross-lingual supervision. Additionally, our model sets the new state-of-the-art accuracies for the Sentiraama corpus.

## References


Balamurali, AR, Aditya Joshi, and Pushpak Bhattacharyya. 2012. Cross-lingual sentiment analysis for indian languages using linked wordnets. *Proceedings of COLING 2012: Posters*, pages 73–82.

Banea, Carmen, Rada Mihalcea, and Janyce Wiebe. 2010. Multilingual subjectivity: Are more languages better? In *Proceedings of the 23rd international conference on computational linguistics*, pages 28–36. Association for Computational Linguistics.

Bel, Nuria, Cornelis HA Koster, and Marta Villegas. 2003. Cross-lingual text categorization. In *International Conference on Theory and Practice of Digital Libraries*, pages 126–139. Springer.

Beutel, Alex, Jilin Chen, Zhe Zhao, and Ed H Chi. 2017. Data decisions and theoretical implications when adversarially learning fair representations. *arXiv preprint arXiv:1707.00075*.

Blitzer, John, Mark Dredze, and Fernando Pereira. 2007. Domain adaptation for sentiment classification. In *45th Annv. Meeting of the Assoc. Computational Linguistics (ACL'07)*.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Can, Ethem F, Aysu Ezen-Can, and Fazli Can. 2018. Multilingual sentiment analysis: An rnn-based framework for limited data. *arXiv preprint arXiv:1806.04511*.

Chen, Xilun, Ahmed Hassan Awadallah, Hany Hassan, Wei Wang, and Claire Cardie. 2018a. Zero-resource multilingual model transfer: Learning what to share. *arXiv preprint arXiv:1810.03552*.

Chen, Xilun, Yu Sun, Ben Athiwaratkun, Claire Cardie, and Kilian Weinberger. 2018b. Adversarial deep averaging networks for cross-lingual sentiment classification. *Transactions of the Association for Computational Linguistics*, 6:557–570.

Conneau, Alexis, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2017. Word translation without parallel data. *arXiv preprint arXiv:1710.04087*.

Durant, Kathleen T and Michael D Smith. 2006. Mining sentiment classification from political web logs.

Gangula, Rama Rohit Reddy and Radhika Mamidi. 2018. Resource creation towards automated sentiment analysis in telugu (a low resource language) and integrating multiple domain sources to enhance sentiment prediction. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.

Ganin, Yaroslav, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Goud, Jaipal Singh, Pranav Goel, Allen J Antony, and Manish Shrivastava. 2019. Leveraging multilingual resources for open-domain event detection. In *Workshop on Interoperable Semantic Annotation (ISA-15)*, page 76.

Hochreiter, Sepp and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Iyyer, Mohit, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. 2015. Deep unordered composition rivals syntactic methods for text classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 1681–1691.

Joachims, Thorsten. 1998. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer.

Kanayama, Hiroshi, Tetsuya Nasukawa, and Hideo Watanabe. 2004. Deeper sentiment analysis using machine translation technology. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*.

Kim, Yoon. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Lu, Bin, Chenhao Tan, Claire Cardie, and Benjamin K Tsou. 2011. Joint bilingual sentiment classification with unlabeled parallel corpora. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 320–330. Association for Computational Linguistics.

Meng, Xinfan, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu, and Houfeng Wang. 2012. Cross-lingual mixture model for sentiment classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 572–581. Association for Computational Linguistics.

Mihalcea, Rada, Carmen Banea, and Janyce Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Proceedings of the 45th annual meeting of the association of computational linguistics*, pages 976–983.

Pham, Hieu, Thang Luong, and Christopher Manning. 2015. Learning distributed representations for multilingual text sequences. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 88–94.

Prasad, Suhaas. 2010. Micro-blogging sentiment analysis using bayesian classification methods. In *Technical Report*. Stanford University.

Prettenhofer, Peter and Benno Stein. 2010. Cross-language text classification using structural correspondence learning. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 1118–1127.

Rish, Irina et al. 2001. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46.

Schönemann, Peter H. 1966. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10.

Shi, Lei, Rada Mihalcea, and Mingjun Tian. 2010. Cross language text classification by model translation and semi-supervised learning. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1057–1067. Association for Computational Linguistics.

Socher, Richard, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.

Tai, Kai Sheng, Richard Socher, and Christopher D Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075*.

Wan, Xiaojun. 2008. Using bilingual knowledge and ensemble techniques for unsupervised chinese sentiment analysis. In *Proceedings of the conference on empirical methods in natural language processing*, pages 553–561. Association for Computational Linguistics.

Wan, Xiaojun. 2009. Co-training for cross-lingual sentiment classification. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-volume 1*, pages 235–243. Association for Computational Linguistics.

Wang, Yequan, Minlie Huang, Li Zhao, et al. 2016. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615.

Xiao, Min and Yuhong Guo. 2013. Semi-supervised representation learning for cross-lingual text classification. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1465–1475.

Xu, Kui and Xiaojun Wan. 2017. Towards a universal sentiment classifier in multiple languages. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 511–520.

Xu, Ruochen and Yiming Yang. 2017. Cross-lingual distillation for text classification. *arXiv preprint arXiv:1705.02073*.

Yarowsky, David, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the first international conference on Human language technology research*, pages 1–8. Association for Computational Linguistics.

Zhou, Xinjie, 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 (Volume 1: Long Papers)*, volume 1, pages 1403–1412.