International Journal of Artificial Intelligence and Machine Learning
Volume 10 • Issue 2 • July-December 2020

A Literature Review on Cross Domain Sentiment Analysis Using Machine learning

Nancy Kansal, Ajay Kumar Garg Engineering College Ghaziabad, India
Lipika Goel, Ajay Kumar Garg Engineering College Ghaziabad, India
Sonam Gupta, Ajay Kumar Garg Engineering College Ghaziabad, India

ABSTRACT

Sentiment analysis is the field of NLP which analyzes the sentiments of text written by users on online sites in the form of reviews. These reviews may be either in the form of a word, sentence, document, or ratings. These reviews are used as datasets when applied to train a classifier. These datasets are applied in the annotated form with the positive, negative or neutral labels as an input to train the classifier. This trained classifier is used to test other reviews, either in the same or different domains to know like or dislike of the user for the related field. Various researches have been done in single and cross domain sentiment analysis. The new methods proposed are overcoming the previous ones but according to this survey, no methods best suit the proposed work. In this article, the authors review the methods and techniques that are given by various researchers in cross domain sentiment analysis and how those are compared with the pre-existing methods for the related work.

KEYWORDS

Cross-Domain, Domain Adaptation, Machine Learning, Opinion Mining, Sentiment Analysis, Transfer Learning,

1. INTRODUCTION

Sentiment analysis is also called opinion mining or emotion AI. Sentiment Analysis is used to know about what people think about things they are consuming from watching a movie to purchasing an AC. On the basis of their thinking either positive or negative, producer or holders of service or product get to know whether the given service/product has the future scope or not. For example, there is a new movie released; various social networking sites are the source of thoughts of people who have seen the movie. On the basis of those reviews, publicity of the movie reaches to the makers of the movie and related trend goes on. Whenever there are elections to be held, election outcomes are predicted on the basis of the analysis of sentiments, opinion, and thoughts those are shared by public on various online portals or social networking sites or news. Sentiment Analysis is the study of attitudes of the holder of service towards the consumed service either in the form of love or hate; like or dislike; positive or negative (polarity). This attitude is analyzed from the text that is presented in the form of reviews in word form, sentence form or document form or in the form of ratings given by holder.

Sentiment Analysis can be done using various machine learning algorithms in which a model/classifier is trained using reviews that are annotated with the polarity positive or negative. These annotated reviews can be taken from any domain to train the classifier and the trained classifier is tested for the orientation of text or reviews in same or different domain.

DOI: 10.4018/IJAIML.2020070103

While doing sentiment analysis in single domain, classifier is trained in single domain and is tested for the same domain thus only problem is to annotate the dataset but in cross domain sentiment analysis problem arises that any word/feature having positive meaning in one domain may have negative meaning in other domain; or in one domain, positive sentiments are expressed by some words and in other domain those positive sentiments are expressed using different words(domain specific words in different domains). In the cross domain, the problem of dataset labeling may be a time-consuming and costly process as it is done manually. SentiWordNet can also be a solution as it is an opinion Lexicon derived from WordNet database which is having scores of positive and negative for attributes and hence can tell polarity of the document on the basis of overall polarity of words written in the document, but it is also having a limitation of words. Various methods and techniques have been proposed recently to overcome this problem of labeling of dataset. The aim of the study is to put some relevant studies together in this paper to help the researchers, by comparing the methods proposed in various studies and by also giving a performance comparison of techniques used in studies. This study aims to focus only on the most recent works published during the period 2010 to 2019.

The remainder of this paper is as follows: Section 2 defines some key Terminologies used to understand the study on cross domain sentiment analysis. Section 3 describes some challenges and issues related to sentiment analysis in the cross domain. Section 4 demonstrates the methodology for this study purpose followed by a brief discussion of methods used in researches those the authors have used in this study. Section 5 demonstrates various datasets used by different researches on cross domain sentiment analysis. Section 6 defines some of the baseline methods used to compare performances of their proposed methods by various scholars. Section 7 gives answers to the questions that are aimed at this survey study. It compares the performance of all the baselines proposed by various researches. Section 8 is about some discussion on methods given to the problem of sentiment analysis in the cross domain and the Section 9 concludes the paper.

2. KEY TERMINOLOGY

Here, the authors define some basic terminologies that are used for this review purpose.

2.1 Domain

With respect to this research, the domain is such collection where all the entities have similar characteristics like in electronics products, DVD is one domain and AC is a different domain. In social networking sites, Twitter is one domain and Facebook is a different domain.

2.2 Sentiment Analysis

Sentiment analysis is the field of natural language processing in which unstructured online public opinions about any product, social media, brand, news, or research and so on presented in the form of reviews, are transformed in structural information that is annotated dataset having the positive, negative, neutral or mixed polarity of sentiments.

Sentiments can be represented in word, sentence or document form and accordingly, analysis is applied on the stated.

2.2.1 Aspect Based Sentiment Analysis

When a researcher is interested in the particular feature/aspect along with positive, negative and neutral sentiments of product, to which users are interested in; it is called Aspect Based Sentiment Analysis. For Ex- if someone says “battery of the new phone is short term” then the negative sentiment is shown for the battery of the phone, not for the phone, hence here battery is one aspect of sentence-level sentiment.

2.3 Sentiment Analysis Methods

2.3.1 Rule/Lexicon-Based Approach

A lexicon is a collection of words associated with their individual polarity. In this approach, a lexicon is used to detect the polarity of sentiment documents. Like someof the popular lexicons are: AFINN-11, SentiWordNet, and SenticNet. The words of AFINN-11 are manually labeled by Finn Arup Neilsen in 2009-2010, SentiWordNet is augmented form of WordNet having sentiment information of each word, SenticNet provides orientataion associated with nearly 50,000 natural languages concepts.

2.3.2 Machine Learning-Based Approach

It has a machine learning classifier that is trained by first input the labeled features then polarity/label of unlabeled features are predicted, either in the same domain or in different domains. The output of this classifier is the polarity of sentiment features of the output domain.

2.3.3 Cross Domain Sentiment Analysis

While applying machine learning algorithms for sentiment analysis if model/classifier is trained using the dataset of one domain (called input domain) but is tested with the dataset for a different domain (called output domain), whether that dataset is labeled or unlabeled, then such analysis is called cross domain sentiment analysis. If the dataset is labeled overall sentiment polarity can be found easily but if the dataset is unlabeled, it is tough to predict the overall sentiment of a document. Sentiment analysis is a predominant task in every field that too when it is a smart era of the internet. But it is economically unreliable to do sentiment analysis in every domain, so cross domain sentiment analysis is performed in which classifier is trained in input domain using annotated dataset of that domain and is tested on output domain to annotate the sentiment polarity expressed by the sentiments presented in form of reviews (words, sentences or documents) or ratings as well.

3. CHALLENGES AND ISSUES IN CROSS DOMAIN SENTIMENT ANALYSIS

3.1 Feature Meagerness

It is the problem when feature those are expressed in the output domain is not found in the input domain. Due to which classifier trained in the input domain is not sufficiently trained for sentiment analysis in the output domain.

3.2 Polarity Deviation

When any word in one domain has either positive or negative polarity but in other domains, the same word has opposite polarity then it may cause bad results of the trained classifier as actual sentiments are opposed by the classifier.

3.3 Lexical Ambiguity

When a word/feature has different meanings due to different contexts of different domains (as a word has many different meanings based on the context they are being used to) hence classifier trained for input domain may not be accurate for testing in the output domain.

4. SYSTEMATIC LITERATURE REVIEW

The authors have performed a systematic literature review to survey current state-of-the-art around cross domain sentiment analysis and based on that work the authors tried to seek the answer to the following two questions.

Question 1: which method is widely used as a baseline to compare the performance of proposed methods by different authors?
Question 2: which baseline method gives the best comparison results among all the baselines on the basis of their compared performance analysis?

The authors performed this survey on cross domain sentiment analysis, for which the authors followed the following steps:

4.1 Searching Process

The authors started this survey by searching for relevant topics for cross domain sentiment analysis. The authors used the Google search engine for this searching process.

4.2 Sources

The authors search for digital libraries like IEEE. Google Scholar and ScienceDirect using keywords sentiment analysis cross domain sentiment analysis and cross domain sentiment analysis techniques, as these are keywords to this related review.

4.3 Study Inclusion Criteria

The authors have taken research papers mainly during the period 2010 to 2019, related to cross domain sentiment analysis. Table 1 shows a brief discussion of all the studies that the authors have taken for this review.

4.4 Research Focus

The authors have performed this research in order to give the answer to two above mentioned questions so that it might aid to researchers to enhance their research in respected field of cross domain sentiment analysis.

The authors have considered all the mentioned papers to give answers to the above questions.

5. DATASETS TAKEN BY STUDIES

Datasets collected for most studies (P1, P2, P3, P4, P5, P6, P7, P8, and P9) are in the English language. This one selected study has a Chinese dataset of reviews of restaurants and cameras from the Dianping website. Reviews from the camera are labeled by three experienced persons. The authors have not considered the part of the study related to that dataset for this review process.

Amazon product dataset is mostly used in research studies as this dataset is widely used to perform cross domain sentiment analysis. Table 2 shows a hetero domain dataset of Amazon product reviews.

Other studies are done on different datasets taken from various domains; those are presented in Table 3.

6. BASELINE METHODS TAKEN BY STUDIES FOR COMPARISON WITH THE PROPOSED

Various baseline methods are chosen by various research scholars for their studies which the authors have taken for this review process. Those baseline methods are discussed in Table 4. And some of the baseline methods are discussed below.

Table 1. Summary of selected studies related to Cross Domain Sentiment Analysis

Publication (Year)	Methodology/Finding	Proposed Classifiers	Performance/ Result	Natural Language Processing	Key
Cross domain sentiment classification via spectral feature alignment (2010)	SFA algorithm is proposed to reduce the gap between cross domain sentiment data. The co-occurrence matrix is used to gap between domain-specific words to domain-independent words. Features are represented in the form of a collection of words (Ngrams) that are labeled with +1(positive) and -1(negative) polarity based on all words in Ngrams. A bipartite graph is constructed to co-align domain-independent features to domain-specific features to find a new feature space. The spectral clustering algorithm is applied on feature bipartite graph to align domain-specific words if they have more common domain independent words and vice-versa. These clusters then represent a new dataset which is used to train sentiment classifier.	SFA, LSA, NoTransf, LSA,FALSA	Accuracy of SFA is compared with NoTransf, LSA, FALSA, SCL by 24 tasks on 2 datasets. the t-test is done on the comparison results of two datasets and SFA out-performs other methods with 0.95 confidence interval.	n-gram	P1
Cross domain sentiment classification using sentiment sensitive thesaurus (2013)	Sentiment sensitive thesaurus (SST) is created to align words having the same sentiments from different domains. SST is used to expand feature vector (training set) and using this L1 Logistic regression based binary classifier is trained which is used to predict the sentiment of the target domain.	L1 Regularized logistic Regression.	Performance varies with varying thesaurus size. Accuracy increases with an increase in thesaurus size. After saturation, it decreases with an increase in size. Trained Classifier is compared with SentiWordNet, and it performs better grouping of words that expresses similar sentiments.	Unigrams and bigrams (called lexicon elements), ratings (called sentiment elements).	P2
Cross domain sentiment classification using sensitive sentiment embeddings (2016)	The unsupervised classification method is used using spectral embeddings. Domain dependent features (pivots) are selected to map in embedded space as close as possible. Documents having the same polarity should be embedded close to each other than a document with different polarities.	Composite optimization model using OO matrices.	Performance is comparable to SCL and SFA.	pointwise mutual information (PMI) method is used for selecting pivots from the document.	P3
Cross-domain sentiment classification: An empirical investigation (2016)	Three datasets are used to compare performance using three different classifiers. Datasets are taken as the first dataset is created using sentiment 140 corpus, second is SemVal dataset and the third is dataset as three review domains. The performance of cross domain classification is determined by using these datasets by training the models. Supervised learning was applied as classifiers were tested on manually labeled tweets. 8 types of Emoticons were used to label tweets	SVM, NB, MNB	Best performance is gained using MNB trained by tweets dataset to determine sentiment in reviews. The best performance was gained using SVM with unigrams and ME with unigrams and bigrams.	Unigram, bigram, unigram and bigrams, unigrams with parts-of-speech (POS) bags.	P4
Cross-domain sentiment classification based on transfer learning and adversarial network (2018)	The shared knowledge Learning and transfer (SKLT) model is introduced based on Transfer Learning and adversarial Networks. Shared and Private models (bi-GRU) are used to learn shared sentiment knowledge and domain-specific knowledge.	SKLT, bi-GRU	Single bi-GRU, SKLT- frozen, SKLT-adaptation are the contrast models to compare with. And SKLT domain adaptation outperforms.	n-grams	P5
Hierarchical attention transfer network (HATN) for cross product sentiment classification (2018)	HATN automatically captures pivot and non-pivots elements. P-nets and NP-nets conduct Attention learning to find pivots and non-pivots elements. It provides a hierarchical attention transfer mechanism that automatically transfers the attention of emotions in both word and sentence levels across domains. HATNh is a proposed model that has hierarchical positional encoding.	NLTK used for tokenization, HATN, HATNh	Comparison is done with the baseline models like SFA, DANN, DAmSDA, CNN-aux, AMN, P-net, NP-net. And it is found that representation of P-net and NP-net are complementary. HATNh improves the performance of HATN by 0.41% on average.	Document-based features.	P6

continued on following page

Table 1. Continued

Publication (Year)	Methodology/Finding	Proposed Classifiers	Performance/ Result	Natural Language Processing	Key
Cross Domain sentiment classification by Capsule network with semantic rules (2018)	CapsuleDAR Model consist of two capsules is used. (Called Base Network and Rule Network). Rule Network to integrate semantic rule to capsule network to capture common knowledge of different domains. Base Network is having an embedding layer to convert word into a low dimensional vector representation, convolutional layer to extract n-gram features. Pivot Based Filter Initialization method is introduced. SCL is used to select pivot features. The K-means method is used to cluster the features. Incaps, Outcaps, and classcaps layers are used in Base Network. Rulecapes layer is used in Rule network. CORAL LOSS is used to minimize the feature difference between the source and target domain.	CapsuleDAR (capsule network in Domain Adaptation with semantic Rule)	The model outperforms various methods like SCL-MI,SS-FE,DANN, SVM, DACNN, DAmSDA, AE-SCL-SR, PBLM, and CapsuleNoDA. It gives a 7.9% improvement over its best competitor (AMN).	n-grams	P7
Adding prior knowledge In hierarchical attention neural network (HANP) for cross domain sentiment classification (2019)	Sentiment Dictionary Layer is used to identify all sentiment words in the context of pivots, non-pivots, and dis-pivots. 3-Layer CNN is used for contextual preservation from source Domain to target Domain. HANP is tested on various datasets for classification.	HANP	It is compared to HAN, CNN-aux, AMN, HATNh, HAN+CNN, HAN+CNN+pivots, HAN+CNN+pivots+non-pivots and gives a state-of-the-art performance with the max. average accuracy of 5.78% when compared with the CNN-aux.	n-grams	P8
CCHAN: An end-to-end model for cross domain sentiment classification (2019)	CTN + CTAN = CCHAN. Cloze Task Network (CTN) is used to obtain word embeddings and also matching is done between document and candidate answer.(to update word embeddings in the source as well as target domains). CTAN is used for sentiment classification.	CCHAN	Model is compared with HAN, CNN-aux, AMN, HATN*h*, CHAN, CCHAN-pivots, and it outperforms all the models.	n-gram	P9
Neural attentive network for cross-domain aspect level sentiment classification (2019)	It uses a weekly supervised Latent Dirichlet Allocation Model (Wilda) to learn Domain-specific Aspect and sentiment Lexicon representations. Aspect level sentiment classifier uses domain classification results and aspect document representation to classify aspect level sentiments in the target domain. LSTM is used to encode the input document. NAACL transforms document embeddings to domain-specific document embeddings.	Bi-directional LSTM	NAACL is superior to compared baseline methods in terms of classification accuracy and F1 score. And also it is shown that it can also find the words that are important to judge the polarity of the source text. Baseline methods are SVM, SVM feature, LSTM, TD-LSTM, JST, SFA, SDA-LSS, ATAE-LSTM, MemNET, RAM, IAM.	wsLDA (weakly supervised latent Drichilit Allocation) is used to finddomain-specific aspects from documents.	P10
Cross-domain co-extraction of sentiment and topic lexicons (2012)	A new bootstrapping-based method, Realtional Adaptive Bootstrapping (RAP) is proposed for expanding lexicon to retrain the classifier. Transfer Adaboost learning (TrAdaBoost) algorithm (Dai et al., 2007) is used for learning in RAP. They have used SVM as a base classifier in Tr-AdaBoost.	Relational Adaptive Bootstrapping(RAP), Tr-AdaBoost, SVM	The relational bootstrapping method(RAP) performs better than the TrAdaBoost and the cross-domain CRF algorithm, and achieves comparable results with the semi-supervised method.	POS tagging is used to represent previous, current and next words.	P11

6.1 NoTransf

Transfer learning is the process where a model is trained using a large amount of annotated dataset and this model is used as a baseline to train other data. In (Pan et al., 2010) pan et al. has used the NoTransf classifier that is trained only by training data of the source domain. And is used to test the target domain.

Table 2. Research studies based on Amazon Product dataset

Dataset	Domain	Key: Year	Author
Amazon product reviews	DVD, Kitchen, Books, Electronics	P1: 2010	Pan et al.
		P2: 2013	Bollegala et al.
		P3: 2016	Bollegala et al.
		P5: 2018	Xiaoyu Duan et al.
		P6: 2018	Li et al.
		P7:2018	Zhang et al.
		P8:2019	Tu Manshu and Wang Bing
		P9:2019	Tu Manshu and Zhao Xuemin

Table 3. Dataset taken by selected research studies

Key	Author(year)	Dataset	Domain	Description
P1	Pan et al.(2010)	Yelp and Citysearch reviews dataset Amazon product reviews dataset	Hotel Videogames Software Electronics	12000 reviews from www.yelp.com and www.citysearch.com 8000 review from each domain are taken from www.amazon.com
P4	Brian et al.(2016)	Sentiment140 corpus dataset SemEval dataset Reviews and rating dataset	Twitter tweets (emotions) Tweets (emotions) Hotels Doctors Restaurant	From 1.6 million tweets from www.twitter.com labeled with emoticons, 10000 were used for the study. Manually annotated tweets Were taken. 2836 annotated reviews were taken from.
P5	Xiaoyu Duan et al.(2018)	IMDB reviews dataset	Movies	Labeled English sentences are taken.
P10	Tang et al.(2019)	Semeval14, SemEval15 SemEval16 datasets	Restaurant Laptop	Five aspect based categories are used, those are price, food, service, ambiance and miscellaneous. Categorized on the basis of performance, price, quality, and appearance.

6.2 LSA

LSA is used to find the features having the same meaning in a review text document. Pan et al. (Pan et al., 2010) used LSA as a baseline method to train the classifier by applying LSA in domain-specific features.

6.3 FALSA

Pan et al. (Pan et al., 2010) used FALSA as the base method that works in the same way as LSA except that it applies LSA on the co-occurrence matrix of domain-specific and domain-independent features.

6.4 No Adapt

When a classifier is trained, feature expansion is done as preprocessing step to train the classifier but in No adapt baseline method feature extraction is not performed but binary classifier is trained only by using unigram and bigram features from annotated source domain and classifier is tested for the dataset of the target domain.

6.5 SFA

SFA aligns domain-specific features/words from different domains and forms a cluster of those aligned features. In (Bollegala et al., 2016; Li et al., 2018) authors have used SFA as a baseline method on their dataset to compare their proposed method's performances.

6.6 SVM

SVM is a discriminant also called hyperplane that separates the annotated features in two different classes in a multidimensional space. A-line/ discriminator is drawn between the two classes. Regularization parameter (c) is used to set the margin of the discriminator such that smaller c value, higher the margin and vice-versa. In (Heredia et al., 2016), Brian et al. have set c to 5.0 for their study and in (Zhang et al., 2018), Zhang et al. has used SVM with RBM kernel.

6.7 NAÏVE BAYES

In general, naïve Bayes classifier works on Bayes theorem which works on the relative probability of an outcome (p (x/E) means the probability of x while event E occurs). In sentiment analysis, it calculates the probability for features to belong in a particular class/polarity (positive and negative). It is called naïve as it assumes features/input words to be independent of each other. In (Heredia et al., 2016) Brian et al. used NB as a baseline as it can give good performance and shows dependencies of features on local as well as global level.

6.8 CNN-aux

It is CNN with two auxiliary tasks to aid sentence embedding. It is used by (Li et al., 2018; Manshu et al., 2019) as their baseline method for the same purpose. (Tu Manshu and Wang Bing, 2019) have also used it for sentiment classifiers.

6.9 AMN

(Li et al. 2017) proposed AMN that automatically captures pivots using an attention mechanism. It does not need a manual selection of pivots. It consists of two memory networks that were sharing parameter, one for sentiment classification and other for domain Classification. And both networks were jointly trained. Thus AMN was focused to learn pivots only (Li et al. 2017). AMN is used by (Tu and Wang, 2019) as the baseline for their study.

6.10 JST

JST is the extension of latent dirichlet allocation (LDA) that is used for document-level classification as it constructs an additional sentiment layer (Lin et al., 2012). JST is used by (Yang et al., 2019) with the same parameter as taken in its original paper.

6.11 HAN

It is used for document classification as it constructs a document vector. For this it first selects important words to form a sentence vector, the sentence vectors are aggregated to form the document vector. (Tu and Wang, 2019) have used HAN in their studies to compare the performance of their proposed models.

6.12 DANN

DANN can be applied to almost any feed-forward model by increasing a few standard layers and a gradient reversal layer and the resulting layer is trained (Ganin et al., 2016).

7. PERFORMANCE COMPARISON OF BASELINES PROPOSED IN STUDIES

Based on this comparison study of performance measures of various baselines it is found that all methods taken as baselines give accuracy according to the different parameters and datasets on which those are applied. The performance of methods varies with variation in parameter values and selection of source and target domain combinations. They also depend upon the different feature selection methods applied by different researches. Table 5. shows the values of performance matrices of different baseline methods that are calculated by researchers in their proposed papers. Performance comparison results pictured in Figure 1 show that although SFA is a widely used method in various studies SKLT gives the best accuracy in all the methods that the authors have studied for sentiment analysis in cross-domain.

8. DISCUSSION

Most of the techniques of cross domain sentiment analysis depend upon the similarity of source and target domains. During study an over belief is made upon the similarity of features of source and target domains however as there is meagerness between the features of source and target domains, the techniques give poor results with less accuracy. Furthermore, more accurate results can be found while using labeled datasets to train different models for classification. Labeling dataset manually is costly as well as time-consuming, hence various techniques are being applied by different researches in the last few years. The attention mechanism is introduced that automatically captures pivot features without human intervention. Based on that attentive network is proposed that selects important sentiment from the whole document dynamically and give higher word attention to only domain-specific and domain-independent or pivot whole-part relationships. Based on this review research the authors can classify cross domain sentiment analysis methods or techniques into two classes. The first method is based on the transfer of training data features to testing data features. Example studies of this class are feature-based and thesaurus based researches. The second class is the transfer of the complete document from the target domain to the source domain to work as a training dataset to train the model. An example under this class is active learning-based techniques.

There are different challenges that still need to overcome like polarity deviation and lexical ambiguity. Sentiments in different languages, mixed polarity sentiments, differences in contexts, etc are yet to be faced by the techniques introduced by different researches.

9. CONCLUSION

Sentiment analysis has gained a lot of attention from researchers as it is in demand with the increasing online sentiments of users on different topics as it gives the ability to extract insights from the opinions, sentiments, thoughts reviews and online response that is being given by users. **Cross domain sentiment analysis** is a relevant topic about the same application in which one topic can be used to predict certain decisions about other topics as it provides the facility to train and test sentiments behind heterogeneous topics that can be used to make decisions. Hence, sentiment analysis in the cross domain is widely used as their research topic by many researchers those all to give the solution to this problem by giving models for testing and training based on different methods and techniques to improve the accuracy of results.

The authors have performed this study on a systematic literature review on previous researches related to the same to help the researchers in their respective fields to build a model that gives better performance by knowing the pros and cons of previous related methods proposed. As per this study no technique or method yet proposed gives the perfect solution but later methods proposed are always better than former methods proposed in terms of accuracy. Performance of cross domain sentiment analysis depends on the proper selection of source domain to train the classifier to test the target

Table 4. Baseline methods used in studies

Baseline comparison methods	Elision	Key	Author (year)
No transfer	No-Transf	P1	Pan et al. (2010)
Latent semantic Analysis	LSA	P1	Thomas Hofmann (2001)
Featured latent semantic Analysis	FALSA	P1	Serafin and Di Eugenio (2004)
Negative adaptation	No adapt	P2,P3	
Spectral Feature Alignment	SFA	P3, P6, P10	Pen et al. (2010)
Structured Correspondence Learning	SCL	P1, P3	Blitzar et al. (2016)
Support Vector Machine	SVM	P4, P7, P10	Vladimir Vapnik and Hava Siegelmann (2001)
Naïve Bayes	NB	P4	(1960)
Discriminant Adaptive Nearest Neighbor	DANN	P6, P7	Ganin et al. (2016)
DANN+mSDA	DAmSD	P6,P7	Ganin et al. (2016)
Convolutional Neural Network auxiliary	CNN-aux	P6, P8, P9	Yu and Jiang (2016)
Adversarial Memory Network	AMN	P6, P8, P9	Li et al. (2017)
SCL Mutual Information	SCL-MI	P7	Blitzar et al. (2007)
Hierarchical attention Network	HAN	P8, P9	Yang et al. (2016)
Joint Sentiment Topic/ Model	JST	P10	Lin et al. (2012)
Long Short Term Memory	LSTM	P10	S. Hochreiter and J. Schmidhuber (1997)
Stacked Denoising Autoencoder with Domain and Sentiment Supervision	SDA-DSS	P10	Liu and Huang (2015)
Transfer Ada-Boost	Tr-AdaBoost	P11	Dai et al. (2007)

domain hence proper identification of source domain for a particular domain is most important for feature similarity of domains. The study should be done to select the proper source domain for the adaptation of the target domain.

Techniques should be chosen such as to face all the challenges for sentiment analysis in the cross domain. The most important in which is lexical ambiguity in which word/sentiment‘s meaning changes with context, hence proper domain selection is required to minimize this ambiguity.

Table 5. Performance comparisons of baselines proposed in different researches taken for this study

Baselines	Key	Accuracy (%)	F-Score	AUC-Score	Description	Average Accuracies
SFA(Spectral Feature Alignment)	P1 P2 P3 P6 P10	86.75 77.73 65.47 78.69 78.6	78.4		The highest accuracy achieved when comparing SFA with different datasets in the cross domain aspect. SFA when compared to the baseline in SST (Bollegala et al.2013). With the dimensionality set to 30, experimented on the Amazon dataset. Average acc. On Amazon reviews dataset.	77.44%
SCL(Structural Correspondence Learning)	P3	66.04			Average accuracy on target domain when different source domains are used.	66.04%
NoAdapt (Negative Adaptation)	P3	62.91			The average accuracy of the baseline when different source domains are used to compare the performance of the proposed methods.	62.91%
NB(Naïve Bayes)	P4	-		0.764	When using the sentiment140 Amazon reviews dataset.	
SVM(Support Vector Machine)	P4 P7 P10	- 80.2 72.7	70.9	0.780	Using sentiment140 corpus Amazon reviews dataset. Zhang et al. selected hyperparameter c between 10^-5 to 1.	76.45%
SKLT	P5	87.98			Average accuracy to compare performance in adversarial networks.(Duan et al.2019)	87.98%
DANN(Discriminant Adaptive Nearest Neighbor)	P6 P7	79.00 74.8			Average classification accuracy on Amazon review datasets. Encoded in 5000 dimension feature vector. Average accuracy When tested on 12 different sets of domains using the adaptation parameter between 0.001 and 1 with learning rate 0.001.	76.9%
DANN+mSDA	P6 P7	82.36 76.2			Average acc. Using Amazon review dataset with 5 output layers and a vector of 30000 dimensions. Average accuracy on 12 different domain sets with every instance encoded in a vector of 3000 dimensions.	79.28%
CNN-aux	P6 P8 P9	81.98 81.98 81.98			Average accuracy for Amazon reviews dataset to induce sentiment embeddings using two auxiliary tasks. Average accuracy when using 20 sets of different source and target domains. Average accuracy for 20 transfer pairs of Amazon review dataset.	81.98%
AMN(Adversarial Neural Network)	P6 P8 P9	82.79 82.79 82.79			Average classification accuracy for Amazon review dataset by learning domain shared representations. Average acc. When taken 20 different domain pairs for study. Average acc. For 20 transfer pairs on Amazon review dataset.	82.79%
HAN(Hierarchical Attention Network)	P8 P9	81.07 81.07			Average acc. When taken 20 different domain pairs for study. Average acc. For 20 transfer pairs on Amazon review dataset.	81.07%
JST	P10	79.3			The performance of this baseline is used for comparing performance of NAACL. Performance(given average accuracy) is measured by varying percentage of labeled data in target domain. Labeled data from source domain and labeled/unlabeled data from target domain are used as training set.	79.3%

continued on following page

Table 5. Continued

Baselines	Key	Accuracy (%)	F-Score	AUC-Score	Description	Average Accuracies
LSTM	P10	78.6			Average accuracy while taking the SemEval-14 S-res./Dianping D-res. as the source domains and use SemEval-14 S-laptop/Dianping D-camera as the target domains,by varying percentage of labeled data in target domain.	78.6%
SDA-DSS	P10	81.5			Average accuracy while taking the SemEval-14 S-res./Dianping D-res. as the source domains and use SemEval-14 S-laptop/Dianping D-camera as the target domains,by varying percentage of labeled data in target domain.	81.5%
Tr-AdaBoost	P11		0.51		Average F-Score while taking two tasks: Sentiment Lexicon extraction and Topic Lexicon extraction, on product and movie review datasets.	

Figure 1. Performance comparison of baselines using the line graph

REFERENCES

Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering. *Journal of Machine Learning Research*, *2*(December), 125–137.

Bollegala, D., Mu, T., & Goulermas, J. Y. (2016). Cross-Domain Sentiment Classification Using Sentiment Sensitive Embeddings. *IEEE Transactions on Knowledge and Data Engineering*, *28*(2), 398–410. doi:10.1109/TKDE.2015.2475761

Bollegala, D., Weir, D., & Carroll, J. (2013). Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus. *IEEE Transactions on Knowledge and Data Engineering*, *25*(8), 1719–1731. doi:10.1109/TKDE.2012.103

Dai, W., Yang, Q., Xue, G., & Yu, Y. (2007). Boosting for transfer learning. In *Proceedings of the 24th International Conference on Machine Learning* (pp. 193–200). ACM.

Duan, X., Zhou, Y., Jing, C., Zhang, L., & Chen, R. (2018). Cross-domain Sentiment Classification Based on Transfer Learning and Adversarial Network. In *Proceedings of the 2018 IEEE 4th International Conference on Computer and Communications (ICCC)* (pp. 2302-2306). IEEE Press. 10.1109/CompComm.2018.8780771

Ganin, Y., Hana, A. H. L., Laviolette, F., & Lempitsky, V. (2016). Domain-Adversarial Training of Neural Networks. *Journal of Machine Learning Research*, *17*, 1–35.

Heredia, B., Khoshgoftaar, T. M., Prusa, J., & Crawford, M. (2016). Cross-Domain Sentiment Analysis: An Empirical Investigation. In *Proceedings of the 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)* (pp. 160-165). IEEE Press. 10.1109/IRI.2016.28

Hochreiter, S., & Schmidhuber, J. (1997). Long Short Term Memory. *Journal Neural Computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Hofmann, T. (2001). Unsupervised Learning by Probabilistic Latent Semantic Analysis. *Machine Learning*, *42*(1-2), 177.

Li, F., Pan, S. J., Jin, O., Yang, Q., & Zhu, X. (2012, July). Cross-domain co-extraction of sentiment and topic lexicons. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 410-419). Association for Computational Linguistics.

Li, Z., Wei, Y., Zhang, Y., & Yang, Q. (2018, April). Hierarchical attention transfer network for cross-domain sentiment classification. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI Press.

Li, Z., Zhang, Y., Wei, Y., Wu, Y., & Yang, Q. (2017, August). End-to-End Adversarial Memory Network for Cross-domain Sentiment Classification. In IJCAI (pp. 2237-2243). Academic Press.

Lin, C., He, Y., Everson, R., & Ruger, S. (2012, June). Weakly Supervised Joint Sentiment-Topic Detection from Text. *IEEE Transactions on Knowledge and Data Engineering*, *24*(6), 1134–1145. doi:10.1109/TKDE.2011.48

Manshu, T., & Bing, W. (2019). Adding Prior Knowledge in Hierarchical Attention Neural Network for Cross Domain Sentiment Classification. *IEEE Access : Practical Innovations, Open Solutions*, *7*, 32578–32588. doi:10.1109/ACCESS.2019.2901929

Manshu, T., & Xuemin, Z. (2019). CCHAN: An End to End Model for Cross Domain Sentiment Classification. *IEEE Access : Practical Innovations, Open Solutions*, *7*, 50232–50239. doi:10.1109/ACCESS.2019.2910300

Pan, S. J., Ni, X., Sun, J. T., Yang, Q., & Chen, Z. (2010, April). Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th international conference on World wide web* (pp. 751-760). Academic Press.

Serafin, R., & Eugenio, B. D. (2004). FLSA: Extending Latent Semantic Analysis with Features for Dialogue Act Classification. ACL. 10.3115/1218955.1219043

Yang, M., Yin, W., Qu, Q., Tu, W., Shen, Y., & Chen, X. (2019). *Neural attentive network for cross-domain aspect-level sentiment classification*. IEEE Transactions on Affective Computing.

Zhang, B., Xu, X., Yang, M., Chen, X., & Ye, Y. (2018). Cross-Domain Sentiment Classification by Capsule Network with Semantic Rules. *IEEE Access : Practical Innovations, Open Solutions*, *6*, 58284–58294. doi:10.1109/ACCESS.2018.2874623

Nancy Kansal is a young and dynamic academician in the field of Computer Science. She received her B.Tech (Hons.) degree in Computer Science and Engineering from Radha Govind Group of Institutions in 2016. She completed her schooling with First class in Higher Secondary and in Secondary. Currently, she is pursuing her M.Tech in Computer Science and Engineering from Ajay Kumar Garg Engineering College (Ghaziabad) India which is affiliated to Dr. A. P. J. Abdul Kalam Technical University Uttar Pradesh, Lucknow, India. Her research interests include Machine Learning. She has a deep interest in research and innovations and their applications. She is a quick learner and dedicated to the task given to her. She is always passionate to learn new things. Her strength is her positive attitude and deep faith in God.

Lipika Goel has a B.Tech., an M.Tech. and persuing a PhD. She has over 10 years of experience in teaching. Currently, she is working as an Assistant Professor (CSE) in Ajay Kumar Garg Engineering College, Ghaziabad. She has published over 10 papers in various national/international journals. She is reviewer of many international journals/conferences. Her research interest include machine learning and data analytics.

Sonam Gupta has B.E., an Mtech, and a Ph.D. She has over 10 years of experience in teaching. Currently, she is working as an Associate Professor (CSE) in Ajay Kumar Garg Engineering College, Ghaziabad. She has published over 20 papers in various national/international journals. She is a reviewer of many international journals/ conferences. Her research interest includes software evolution, machine learning, and data analytics.