IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 18, 2020, accepted June 22, 2020, date of publication June 26, 2020, date of current version July 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005242

Unsupervised Genre-Based Multidomain Sentiment Lexicon Learning Using Corpus-Generated Polarity Seed Words

SWATI SANAGAR AND DEEPA GUPTA
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham Bengaluru Campus, Bengaluru 560035, India
Corresponding authors: Swati Sanagar (swatisanagar@yahoo.co.in) and Deepa Gupta (g_deepa@blr.amrita.edu)

ABSTRACT Sentiment lexicon learning is of paramount importance in sentiment analysis. One of the most considerable challenges in learning sentiment lexicons is their domain-specific behavior. Transferring knowledge acquired from a sentiment lexicon from one domain to another is an open research problem. In this study, we attempt to address this challenge by presenting a transfer learning approach that creates new learning insights for multiple domains of the same genre. We propose an unsupervised sentiment lexicon learning methodology scalable to new domains of the same genre. Incremental learning and the methodology learn polarity seed words from corpora of multiple automatically selected source domains. This process then transfers its genre-level knowledge of corpus-learned seed words to the target domains. The corpus-learned seed words are used for sentiment lexicon generation for multiple target domains of the same genre. The sentiment lexicon learning process is based on the latent semantic analysis technique and uses unlabeled training data from the source and target domains. The experiment was performed using 24 domains of the same genre, i.e., consumer product review. The proposed model displays the best results using standard evaluation measures compared with the competitive baselines. The proposed genre-based unsupervised approach achieves a maximum accuracy of 86% and outperforms methods recently presented in the literature.

INDEX TERMS Multiple domains, sentiment lexicon, corpus-learned seed words, transfer learning, latent semantic analysis.

I. INTRODUCTION

With the growth of the Internet, an enormous amount of information has become publicly available over the past decade. An increasing number of individuals share their experiences and opinions online, such as in consumer forums and on social media. The Web is evolving toward an era in which consumer opinions are predicted to dictate the final shape of products and services [1]. Sentiment analysis techniques process these messages to extract valuable information. Sentiment analysis is defined as the computational study of people's opinions, experiences, and emotions about an event, topic, object, individual, organization, etc. Sentiment analysis is valued and widely used by individuals and organizations worldwide [2].

In text data, sentiments and opinions are conveyed using opinion oriented words (OOWs), which are used to detect the underlying sentiment. Lexicon/knowledge-based and statistical/machine learning are the two primary approaches to sentiment analysis [1], [3], [4]. The combination of these two approaches forms the third approach called a hybrid approach. OOWs are an essential component of both statistical-learning and lexicon-based approaches. Lexicon-based approaches use OOWs from sentiment lexicons to perform tasks in sentiment analysis. Statistical/machine learning techniques use sentiment words as basic and deep features. Hybrid approaches use additional n-grams and part-of-speech (POS) patterns as features for sentiment classification [1]–[8]. Additionally, recent neural network approaches have used existing lexicons as resources to improve performance [9]. Therefore, building sentiment lexicons is an important task.

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

Commonly used general-purpose English language sentiment lexicons include SentiWordNet 3.0 [10], General Inquirer (GI) [11], the Multi-Perspective Question Answering (MPQA) lexicon [12], SenticNet [13], and Valence Aware Dictionary and Sentiment Reasoner (VADER) [14]. These sentiment lexicons are useful for general sentiment analysis; however, they may not be effective in a domain-specific context. In the following sentences taken from the Amazon consumer review dataset, the sentiment orientation of the opinion word ''big'' varies depending on the context:

A. These workout pants are too *big* for me.
B. This popcorn pops up *big* and in no time at all.
C. The earpieces are *big* for fitting.

Example A from the ''Clothe, Shoes, Jewel'' domain and example C from the ''Cell Phone'' domain express negative sentiments, whereas example B from the ''Grocery & Gourmet Food'' domain expresses a positive sentiment. The dominant score for the sentiment word ''big'' is positive according to SentiWordNet 3.0 if an approach that ignores the objective score is used. However, this score is not relevant to examples A and C. Examples from different domains thus demonstrate that general-purpose sentiment lexicons are less effective for domain-specific tasks because they represent a generic sentiment orientation.

As observed from the above examples, generic sentiment orientation is important but not applicable to all domains. Sentiment lexicons that are learned from a domain preserve the domain-based orientation of words. Moreover, they outperform general-purpose sentiment lexicons [15]. Very few domain-specific sentiment lexicons are publicly available, including sentiment lexicons from the National Research Council (NRC) Canada for Laptop, Restaurant, and other domains [16] and for the Finance domain from WordStat [17].

In summary, domain-specific sentiment lexicons provide greater accuracy for sentiment analysis tasks. However, owing to the lack of availability of domain-specific sentiment lexicons, it is essential to generate sentiment lexicons for multiple and diverse domains. This creates a need for an automatic sentiment lexicon generation model that can generate sentiment lexicons for different domains.

Corpus-based and hybrid sentiment lexicon generation creation approaches are either supervised or unsupervised/semi-supervised. Supervised approaches have been relatively well researched and can be used to learn a single domain or a small number of domains [18]. However, supervised learning is highly dependent on labeled data, which are often unavailable. Additionally, data labeling requires a significant amount of time and manual labor. Therefore, the scaling of supervised approaches to multiple domains is difficult. Labeled data scarcity is overcome by recent semi-supervised approaches that include combining labeled and unlabeled data [19], self-training [20], co-training [21], and sentiment topic-model [22] approaches. An unsupervised approach overcomes the absence of labeled data with a knowledge graph propagation approach using a sentiment lexicon [40]. An unsupervised approach with a small amount of seed information can provide significant assistance. Seed-word-based approaches have been used in single-domain sentiment lexicon learning, with the generated sentiment lexicons being able to extract domain-specific sentiment words [23]–[25]. Most seed-word-based approaches use standard seed words [25]. The usage of certain highly expressive words (e.g., bonus, garbage, relax, and crap) has significantly changed in the context of social media (e.g., word garbage has a negative connotation of being wasteful). Therefore, the use of only standard seed words or existing sentiment lexicons may be insufficient for creating a domain-based lexicon that can learn the domain-relevant polarity, domain-relevant polarity score, and domain-specific sentiment words. The generation of a sentiment lexicon that is scalable to multiple new domains remains an open research problem [26], [27].

In this work, we address these challenges by presenting a generic and unsupervised sentiment lexicon learning model for multiple domains of the same genre. The proposed model assumes that multiple domains belonging to the same genre have common characteristics in terms of the text used for expressing experiences and the style of expressing feelings, which is referred to in the literature as the *language domain* [28]. Our approach leverages the connection between domains of the same genre and presents a scalable mechanism for learning a sentiment lexicon using a hybrid approach.

Researchers have observed that knowledge transfer is more effective between similar domains than dissimilar domains [29], [30]. Accordingly, our novel approach first identifies similar domains, learns common knowledge, and then acquires transferrable knowledge. The two multidomain challenges reported in a study by Abdullah *et al.* [31] are addressed by our proposed model framework. The first challenge pertaining to information overload is addressed by a novel domain selection process and relevant information selection for optimized results. The second challenge pertaining to domain divergence is addressed by experimentation on 24 diverse domains of the same genre. The proposed approach segregates the source and target domains using a clustering mechanism. The source domains learn the polarity seed words in an unlabeled training process. The process starts with a pair of seven standard polarity seed words [25] and learns new seed words from the source domain corpora. The seed words that are learned from the corpora are further used to learn the target domain lexicon in the absence of labeled data.

The contributions of this study are as follows:

- An automated mechanism for categorizing domains using the affinity propagation algorithm that segregates the domains into source and target domains to avoid biased domain selection
- A sentiment lexicon generation mechanism for unlabeled data using latent semantic analysis (LSA) that relies on a truncated singular value decomposition (SVD)

- A corpus-learned polarity seed word generation model that combines knowledge learned from multiple source domains of the same genre
- Learning of a sentiment lexicon for multiple target domains using a polarity seed word generated on the basis of the proposed genre-based model
- Adaptability to domains, model performance analysis, and comparison with existing sentiment lexicons and other models

The remainder of this paper is organized as follows. Related work is discussed in Section II, while a description of the model, learning approach, and different stages are presented in Section III. Details of the corpora are provided in Section IV, while evaluation measures and baselines are discussed in Section V. The experimental setup is presented in Section VI, and the results of the comparative evaluation are provided in Section VII. An analysis of the experimental results, including a case study and a qualitative analysis, is presented in Section VIII. Conclusions and potential extensions of the proposed model are discussed in Section IX.

II. RELATED WORK

Automatic sentiment lexicon learning has been extensively studied using several techniques for over a decade and a half. Sentiment lexicon can be learned using two basic approaches: dictionary-based approaches, which expand or adapt existing sentiment dictionary bases, and corpus-based approaches, which learn new lexicons from corpora using different methods. We present a brief survey of these approaches in the following subsections.

A. DICTIONARY-BASED APPROACHES

Dictionary-based approaches start with initial inputs from linguistic resources (e.g., language dictionaries, thesauruses, and other popular lexicon resources) and learn new sentiment lexicons. Early studies used relation-based knowledge, such as synonyms and antonyms, to expand polarity lexicons. For example, a set of positive and negative adjectives was expanded using the WordNet synonym and antonym relations [23]. Initial approaches assigned a discrete polarity score to sentiment words, while later approaches used real value scores. SentiWordNet3.0 [10] was developed from WordNet glosses using synonyms and other relations. A word was assigned positive, negative, and objective scores based on subjectivity/objectivity relations. The semantic orientation calculator (SO-CAL) approach computed sentiment scores by incorporating intensification and negation [32]. This approach used GI and domain-based lexicons with associated sentiment word orientation and strength. Mutual information between SentiWordNet synsets and labeled corpora was explored to learn a sentiment lexicon [33]. In another approach, Thelwall lexicon and a two-dimensional transformed vector representation were used to calculate sentiment scores [34]. This method learned the sentiment scores and expanded the existing lexicon.

In another study, multiple existing sentiment lexicons were combined to collect common words having the same polarity [28]. These words were added to learn more words that expressed sentiments from a corpus using the point-wise mutual information (PMI) approach. A Chinese sentiment lexicon was learned [8] using labeled phrases from HowNet and the National Taiwan University Sentiment Dictionary (NTUSD). It determined the semantic fuzziness of these phrases at the character level. This approach calculated the semantic tendency of a character using its frequency in positive and negative phrases by applying the Gaussian density function. Because of the insufficiency of available lexicons, a new general-purpose lexicon was manually created based on a common word list from 12 dictionaries [35]. It contained approximately 30,000 words with discrete intensity scores. In another study, HowNet and NTUSD were used to manually identify domain-based synonymy for five consumer review domains [36].

Some methods for dictionary-based learning using graphs, rules, or a fuzzy approach have also been proposed. In one of these approaches, an active query strategy learned the most informative instances using syntactic rules [37]. General sentiment information was adapted to the target domain using domain-specific sentiment similarity. A rule-based n-gram sentiment lexicon creation approach used the general-purpose VADER lexicon [38]. The process used the SO-CAL method to calculate n-gram scores. A sequential learning domain lexicon adaptation approach used several heuristic rules and adapted the SenticNet lexicon to learn the target domain lexicon [39]. This cognitive supervision process tracked incorrectly predicted sentences and used them for supervision. This approach exploited the semantic similarity between words. A recent lexical-affective graph propagation approach was based on variants of the shortest path problem [40]. A graph connected the WordNet knowledge base with the WordNet-Affect hierarchy using semantic relations (e.g., hypernymy or meronymy) and hierarchical relations (e.g., "is a" or "has a"). The approach created a sentiment lexicon of affect scored concepts by computing the scores using semantic relations and WordNet hierarchical relations. Another approach proposed binned corpus polarity lexicon creation for stock market data [41], wherein price changes of particular stock were used as guiding polarity values. It featured fuzzy linguistic labeling, which served as an intermediate measure between polarity and discrete scores.

Recent advancements in neural network research reinvented the use of lexicons. A lexicon integration attention model integrated lexicon embedding into a Convolutional Neural Network (CNN) [42]. This model used three methods of integrating lexicons with word embeddings. The approach used NRC, Hu & Liu, and four other lexicons. A multilayer perceptron (MLP) approach used a positive and unlabeled learning approach to identify sentiment words from corpora [43]. This approach performed a double dictionary search in a Chinese dictionary and classified social media instances. Another study used a lexicon-enhanced network [9] by integrating SentiWordNet in the training process to obtain the sentiment embedding. The concatenated word embedding and its sentiment embedding were input to a Long Short Term Memory (LSTM) network. The model used a combination of multiple general-purpose sentiment lexicons. A recent attention-based model used the extended Affective Norms lexicon in addition to SentiWordNet [44]. The model added attention using a correlation matrix between a sentiment word and context word.

Dictionary-based approaches are highly dependent on existing lexicons. This problem can be overcome by corpus-based approaches, which are discussed in the following subsection.

B. CORPUS-BASED APPROACHES

Corpus-based sentiment lexicon learning approaches use corpus-based linguistic knowledge, linguistic relations, and linguistic dependencies. Corpus-based approaches can be generally categorized into in-domain learning, one-to-one domain learning, and multidomain learning.

1) IN-DOMAIN LEARNING

In-domain learning uses training and test data from the same domain. In an early study, additional sentiment words were identified from a corpus using the relation between conjunctions and modifiers, and polarity was assigned to new adjectives [45]. The semantic orientation approach identified the polarity of words using the PMI and LSA techniques [25]. The semantic orientation score of a word was calculated using its association with standard seed words. Sentiment word polarity disambiguation is handled using verb, noun, and adjectives as opinion words and other features that help in connecting two opinions such as conjunctions [46]. The polarity scores are computed using the Bayesian model. An information retrieval approach used adjectives, nouns, and verbs with specific dependency relations as context features [47]. The method disambiguated polarity of a word by forming positive context word vector from star-labeled positive corpus and similarly negative word vector from negative corpus. The polarity of a word is identified using a query vector similarity with respective positive and negative vectors. Another approach explored the relation between words, associated concept category, and polarity strength of labeled stars review [15]. The relationship between the frequency of words and categories was used to learn a concept quality lexicon using a vector-based method on hotel and movie review datasets. The variability of the sentiment polarity of words based on different domains was captured by a generative model [48]. This model used labeled data for training and experimented on five datasets, including movie review and SemEval Twitter datasets. The model learned a domain-specific sentiment lexicon by jointly considering sentiment words from domains using hierarchical supervision information of documents and words.

A rule-based approach was used to extract sentiment expressive words from a corpus [24]. Rules were designed based on subjectivity/objectivity dependency relations. Weighted lexical graph propagation and the k-nearest neighbor approach were used to learn sentiment words [49]. This approach used a small, manually selected set of seed words to propagate sentiment labels. An unsupervised approach used a few manually selected positive and negative words as seeds [50]. Sentiment propagation was performed using rule-based dependency graph propagation to learn new sentiment words. The experiments involved three datasets, including movie reviews and SemEval Twitter data.

A neural network MLP end-to-end approach [51] embedded scores in a neural network. This approach automatically generated domain adapted sentiment lexicons and used them in posterior sentiment analysis tasks. This neural network approach modified a sentiment lexicon using a distantly annotated text of a certain domain.

In the aforementioned approaches, sentiment lexicon learning is limited to a single domain. The learning scope and reusability of knowledge can be improved by knowledge-transfer approaches.

2) ONE-TO-ONE DOMAIN LEARNING

In one-to-one domain learning, one domain is the source and another domain is the target. The domain used for learning is called the source domain, and the learned knowledge is adapted in the target domain. An early study used structural correspondence learning [29], wherein frequently occurring features in the source and target domains were identified to build a correspondence between the domains. Experiments were conducted on four consumer review domains using labeled source data for all possible source and target combinations. Principal component analysis was used to select the samples [52]. A sample of labeled instances from the target and source domains were used to verify the closeness between the domains. This information helped adapt the scores of words from the source domain to the target domain for all source–target combinations of the four consumer review domains. Domain-independent words from the labeled source domain that were distributed most similarly to the target domain were selected as landmarks [53]. Subsequently, the geodesic flow kernel (GFK) [54] adaptation algorithm was applied to learn the discriminant features of landmarks with the help of their labels. The GFK constructed from source and target data projected a large subspace that lays on a geodesic flow curve. The geodesic flow curve represented the incremental difference in the geometric and statistical properties between the source and target domain spaces. The domain-invariant features of the geodesic flow curve were used as features for classification.

A deep learning stacked denoising autoencoder model used sparse rectifier units for domain adaptation [55]. This model performed text feature extraction for adaptation from labeled source data and other unlabeled data. This approach provided results on four consumer review domains for all possible source and target combinations. Moreover, an adversarial memory network was explored for domain knowledge transfer [56]. Data from the source and target domains were modeled together for different combinations of four consumer review domains. The model jointly trained two networks for sentiment and domain classification. In a supervised approach, a few target samples were added to source training for learning using neural networks and fuzzy inference systems [57] using wireless sensor data and Intel laboratory data.

3) MULTIDOMAIN LEARNING

In multidomain learning, multiple domains are involved in the learning process. A probabilistic generative model assigns each word a domain label, domain-dependent/independent label, and polarity [58]. Experiments were conducted, wherein the number of source and target domains was varied. The results demonstrated that an increase in the number of source domains improved performance, whereas an increase in the number of target domains decreased performance. A multitask learning approach was proposed to learn from multiple source domains [30]. Here, the process segregated the source and target domains. The domain similarity was determined using word distribution and sentiment graph similarity for the Home & Kitchen, Book, DVD, and Electronics domains. This approach explored sentiment-coherent and sentiment-opposite relations and extracted global knowledge from all three source domains, except the target domains. In addition, it extracted domain-specific knowledge from the most similar domains among the four domains, which was transferred to the target domain using the sentiment graphs of the words. This approach used all domains as the source, except the target domain.

Another study on the Hindi language translated consumer reviews from English to Hindi for four domains [59]. All domains except the target domain were considered source domains. A lexicon was learned from labeled source domain corpora and unlabeled target domain corpora. The approach was based on n-grams and PMI. A neural network learning approach explored learning from multiple source domains and a small amount of data from the target domain using Bidirectional Gated Recurrent Units (BiGRU) and a CNN [60]. This approach trained the model using different combinations of two and three source domains. The experiment included four Chinese and four English consumer domains. Another adversarial neural network study used all possible combinations of four consumer review domains with two source domains and one target domain [61]. The model separated domain-specific and common/shared features with two private encoders and one shared encoder. The adversarial training strategy ensured similar shared feature representations.

Furthermore, some multidomain studies focused on joint learning, in which knowledge is shared among domains. A hybrid approach used a genetic algorithm to learn a sentiment lexicon from corpora of different domains that was adaptable to multiple domains [62]. The work was based on Twitter data and used a seed lexicon in addition to corpus-based learning. In a recent neural-network-based approach, domain representation was used as attention to select the features most related to each domain. The process simultaneously extracted domain-specific and shared features. The model performed two tasks in a single network, which helped adapt the scores to different domains [63].

C. SUMMARY

Dictionary-based approaches use existing lexicons and provide quick and easy access to many sentiment words. However, these approaches rely on knowledge from an existing sentiment lexicon so that the advantages and disadvantages of an existing lexicon may be reflected in the learned sentiment lexicon. Different approaches use different general-purpose lexicons, which suggests that different lexicons may be effective for different approaches. In-domain sentiment lexicon learning using a corpus-based approach is limited to a single domain. These approaches accurately reflect domain-specific sentiments; however, the learning process does not share the learned knowledge across domains. To learn the sentiment lexicon of another domain, the same process must be repeated.

Most learning approaches depend on labeled corpora. In addition, research on one-to-one and multidomain learning uses labeled data from source domains, and in some cases, labeled data from the target domain. Herein, we overcome this drawback using an unsupervised learning model.

One-to-one domain learning has been extensively studied. Such learning approaches involve all combinations of source and target domain results without confirming a specific source domain. Multidomain learning studies have also been conducted with all possible source and target domain combinations using two or three source domains. However, no approach has proposed a strategy to select source domain(s) for knowledge transfer. We address this drawback using a genre-based learning approach.

Multiple-domain learning studies are generally based on few (typically four) domains. These studies involve multiple combinations of source domains and a single target domain. In contrast, our model is based on 24 consumer review domains, and the learning approach involves multiple source and target domains.

Our model uses an approach different from existing approaches. Thus, it is not directly comparable with the one-to-one domain-level knowledge transfer or supervised adaptation models. A preliminary study was presented in [64]. The initial study is further extended in the current research work for genre-level multi-to-multi domain knowledge transfer using unsupervised learning. Our approach, to the best of our knowledge, is the first to address the source domain selection process, which overcomes the major challenge observed in the multidomain literature. The proposed genre-based multidomain learning approach is presented in Section III.

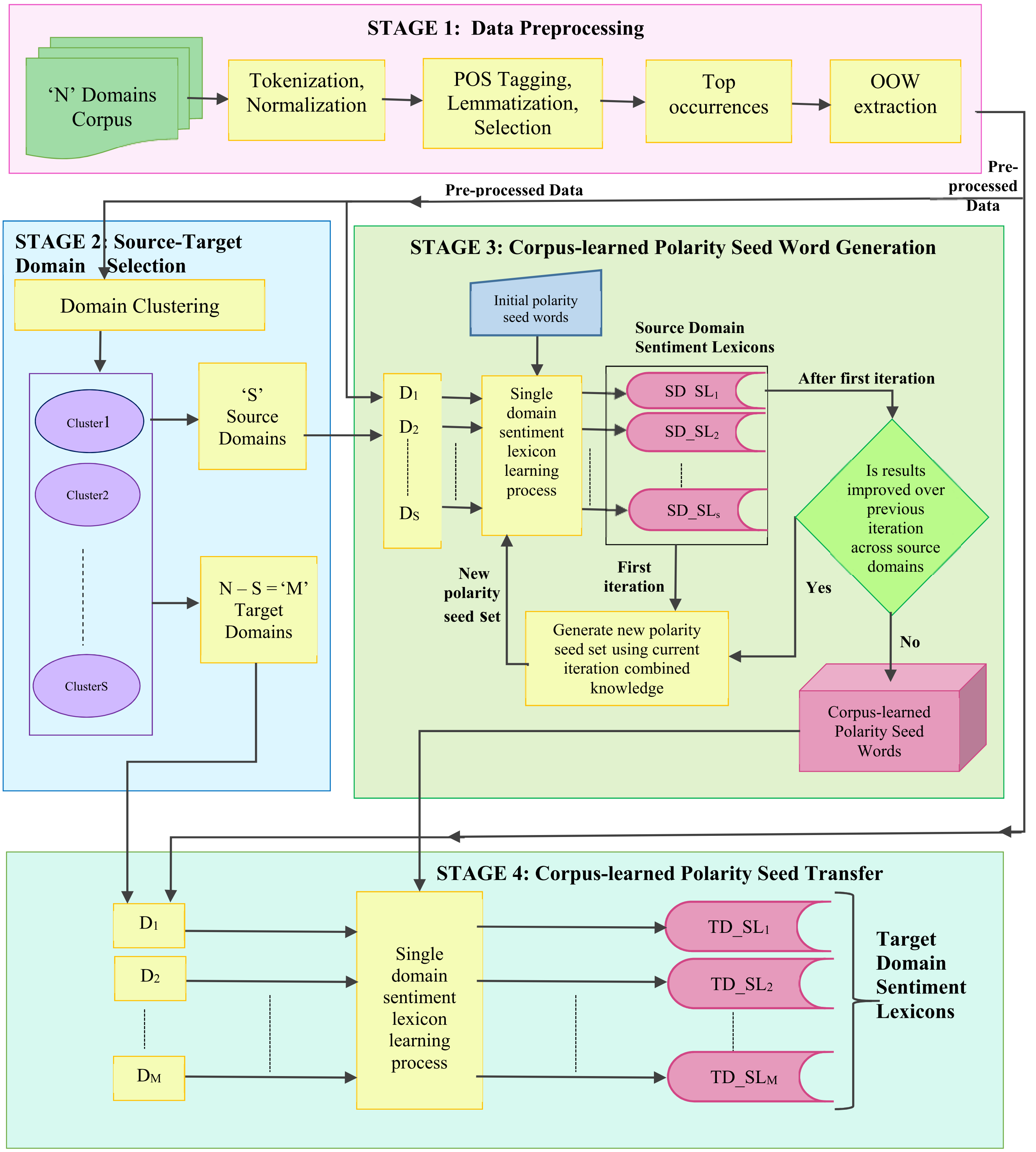


FIGURE 1. Schematic diagram of genre-based multidomain sentiment lexicon generation model.

III. GENRE-BASED MULTIDOMAIN SENTIMENT LEXICON GENERATION MODEL

Figure 1 illustrates the functionality of our approach. The process is divided into four stages. In Stage 1, basic data preprocessing and OOW extraction are performed for all domains. Stage 2 focuses on the process of selecting representative domains. The selected domains are used as source domains, and the remaining domains are considered target domains. Stage 3 is the core of the proposed model. Multiple source domains participate in the learning process and generate knowledge in the form of corpus-learned polarity seed words. These outcomes are passed to Stage 4, which is the corpus-learned polarity seed word transfer stage. Our approach addresses the problem of homogenous transfer learning [65], in which seed-based knowledge is transferred to different but related target domains. The proposed approach considers multiple domains that belong to the same genre, i.e., consumer product reviews. The domains are related because they belong to the same language domain and are written by consumers who like, dislike, and leave opinions about consumer products. The four stages are described in detail in the following subsections.

A. STAGE 1: DATA PREPROCESSING

In Stage 1, unstructured reviews for N consumer product domains are preprocessed, and a set of meaningful features is selected. At the end of this stage, each review is represented in the form of OOWs. Noise removal includes the removal of unwanted characters, such as URL or non-ASCII characters. Repeated characters are replaced by single characters; for example, "loooong" is replaced by "long" using a basic heuristic model that includes spell check. Data are tokenized at the sentence level and processed for POS tagging and lemmatization using the Stanford POS tagger and lemmatizer [66]. The meaningful features are extracted based on their semantic orientation. Early studies on sentiment analysis considered only adjectives and adverbs as candidate OOWs [23], [45]. However, it was later observed that verbs and nouns are also important candidate OOWs [2], [25]. Based on these studies, our approach considers adjectives, adverbs, verbs, and nouns (except proper nouns and auxiliary verbs) OOWs. In the final step, the extracted candidate OOWs are converted into lowercase letters. The reviews are then passed to the next stage of the model.

B. STAGE 2: SOURCE-TARGET DOMAIN SELECTION

This stage segregates domains of the same genre into source and target domains. The preprocessed reviews for the N domains from the previous stage are processed separately for each domain. The OOWs that correspond to the reviews are grouped together to form a separate "review pool" for each domain. This process generates N review pools corresponding to the N domains. Stage 2 aims to identify the source and target domains using a clustering approach based on the observation that domains with a similar word distribution share common characteristic. Manning *et al.* [67] used cosine similarity as a document similarity measure. Documents were clustered into groups [68], such that the degree of similarity tended to be high within the group and low across groups. Our clustering process works in a similar way but at the domain-level. The proposed model uses the affinity propagation algorithm to cluster the domains [69]. The input into the clustering algorithm is a domain similarity matrix that is derived from the N review pools. A term frequency inverse document frequency (TF-IDF) feature matrix D_{LxN} is constructed from the review pools, where each row OOW_i represents an OOW from the collection of L unique OOWs across all N review pools and each column D_j corresponds to N domains. Each value d_{ij} in the matrix represents the TF-IDF score of the i^{th} OOW in the j^{th} domain. d_{ij} is calculated using Equation (1), where the first term, tf, is $tf(OOW_i, D_j)$, which denotes the number of occurrences of an OOW_i in the j^{th} domain. The second term is idf, the logarithm of N divided by C, where N is the total number of domains, and C is the number of domains in which OOW_i appears.

$$d_{ij} = tf\left(OOW_{i,}D_j\right) * log\left(\frac{N}{C}\right) \tag{1}$$

A similarity matrix CS_{NxN} is created from matrix D_{LxN} by calculating the cosine similarity between each pair of column vectors of D_{LxN} using Equation (2), where CS_{ij} represents the similarity score between the pair of domains D_i and D_j.

$$CS_{ij} = \begin{cases} \dfrac{<\vec{D}_i.\vec{D}_j>}{\|D_i\| \ \|D_j\|} & i \neq j \\ & \qquad i, j = 1, 2, .., N \\ 1 & i = j \end{cases} \tag{2}$$

The similarity matrix forms an input of the affinity propagation algorithm [69]. Affinity propagation is a clustering algorithm widely used in text clustering. This graph-based method uses a message-passing technique. In particular, it initially considers all points as centers and gradually builds clusters by exchanging messages. It automatically identifies new cluster centers without prior information on the number of clusters or the cluster centers. This process creates S clusters of N same genre domains. Each created cluster is characterized by its center domain and cluster member domains. The center domain is a semantic representation of a cluster. The center of each cluster is considered the source domain, while the remaining cluster members are considered the target domains. This process forms S source domains and $M = (N - S)$ target domains. The source and target domain selection process constitute the initial step of the corpus-learned polarity seed word generation stage.

C. STAGE 3: CORPUS-LEARNED POLARITY SEED WORD GENERATION

In Stage 3, the selected source domains are used to generate corpus-learned polarity seed words. The seed words are generated by the "combined learning" process. The single-domain sentiment lexicon learning process is a subprocess of this stage, which is discussed in Section III-C-1. Section III-C-2 describes the combined learning process, which combines knowledge from multiple source domains to generate corpus-learned polarity seed words.

1) SINGLE-DOMAIN SENTIMENT LEXICON LEARNING

The single-domain sentiment lexicon learning process is based on the distributional semantic hypothesis, which states that "words that are used and occur in the same contexts tend to have similar meanings" [70]. This process also applies the latent relation hypothesis, which states that "pairs of words that co-occur in similar patterns tend to have similar semantic relations" [68]. The single-domain sentiment lexicon learning process learns a sentiment lexicon for a domain based on LSA [71], which in turn relies on a truncated SVD [72], [73] of an OOW review matrix of TF-IDF values to discover a higher order association between OOWs. LSA creates a semantic space that uses the second level of associations between OOWs in a multidimensional review data space. At the first level, associations between OOWs are observed using a co-occurrence approach. At the second level of association, OOWs may not co-occur, but they can co-occur in similar contexts. For example, the terms "RAM" and "main memory" are used interchangeably in domains such as Electronics or Cell Phone.

The process has four main steps: matrix construction, matrix decomposition, matrix reconstruction, and sentiment lexicon learning. Matrix A_{pxq} contains unique OOWs as rows and single-domain reviews as columns. Each cell a_{ij} stores a TF-IDF score of the i^{th} OOW in the j^{th} review as in Equation (1). SVD is applied to matrix A_{pxq}. The matrix is decomposed into three matrices: a matrix U of OOWs, a matrix V_q of reviews, and a matrix Σ_{pxq} of singular values. In particular, U_p contains a row vector for each OOW, V_q contains a column vector for each review, and Σ_{pxq} stores singular values that represent the importance of dimensions in descending order. A truncated SVD is used to reconstruct the matrices, where only the first k singular values are considered. Let T denote the pxk matrix that is derived from $U\Sigma$ by truncating all rows to length k, as illustrated in Equation (3).

$$T_{pxk} = U_{pxk} \sum_{k} \tag{3}$$

The review matrix V_q is not used in matrix reconstruction, as the learning process focuses on OOWs to build the sentiment lexicon. The matrix construction and decomposition processes are performed only once for the largest k value. Matrix reconstruction, which uses Equation (3), is performed for different k values. This modification significantly reduces the number of computations and improves time complexity and scalability.

The final step of this submodule involves the calculation of sentiment scores for each OOW in the domain that corresponds to every row vector in T. The process uses a small, predefined, and strong semantic opinion orientation seed set that contains positive and negative seed words. *Pseeds* is a set of seed words with a strong positive orientation, while *Nseeds* is a set of negative seed words. The *Pseed_Score* of an OOW is the summation of the cosine similarity between the word vector of the OOW and the positive seed word vectors, as given by Equation (4). Similarly, *Nseed_Score* is calculated using Equation (4).

$$Pseed\_Score(oow_i) = \frac{1}{|Pseeds|} \sum\nolimits_{\forall s, p_s \in Pseeds} \frac{\vec{oow_i}\vec{p_s}}{\|\vec{oow_i}\| \, \|\vec{p_s}\|} \tag{4}$$

$$Nseed\_Score\,(oow_i) = \frac{1}{|Nseeds|} \sum\nolimits_{\forall s, n_s \in Nseeds} \frac{\vec{oow_i}\vec{n_s}}{\|\vec{oow_i}\| \, \|\vec{n_s}\|} \tag{5}$$

$$Score\,(oow_i) = Pseed_{Score(oow_i)} - Nseed\_Score(oow_i), \ \forall \text{ i}, \ \text{i} = 1 \ldots .p \tag{6}$$

The final sentiment score of an OOW is calculated using Equation (6). Each OOW sentiment score ranges between −1.0 and +1.0. The sentiment lexicon learning process is completed by calculating *Score* for all OOWs. Subsequently, the sentiment lexicon is evaluated using the in-domain test on a validation set, which is an unseen labeled review dataset. The score of a review is calculated by summing the sentiment scores of all OOWs from the review, where the generated sentiment lexicon is used for OOW scores. OOWs not found in the sentiment lexicon are assigned a score of 0. A review is classified as positive or negative based on the polarity of the aggregate score. The sentiment lexicon is evaluated in terms of accuracy, as illustrated in Equation (7).

$$Accuracy = \frac{Total\ number\ of\ correctly\ classified\ reviews}{Total\ number\ of\ \ reviews} \tag{7}$$

The matrix reconstruction, OOW score calculation, and evaluation processes are repeated for different k values. The value of k that corresponds to the highest accuracy on the validation set is selected, and the corresponding sentiment lexicon is selected as the final sentiment lexicon for the domain under consideration.

2) POLARITY SEED WORD LEARNING PROCESS

The focus of this stage is to generate corpus-learned polarity seed words by combining knowledge from multiple source domains. The sentiment lexicon learning process is applied to S source domains. The sentiment lexicons that are created for each source domain are denoted SD_SL_i, $i = 1\ldots..,S$. OOWs that have the same polarity orientation across all S source domains are listed in the positive and negative sets, and each listed OOW is assigned the calculated average of the scores across all S source domains. The absolute scores in each set are arranged in descending order. The top P seed words from each set are considered a new seed word set. Equal numbers of positive and negative top words are used in each experiment to avoid a polarity class bias. The top P seed word set is used to build the sentiment lexicon for each source domain for different values of P. The optimal value of P that corresponds to the highest accuracy across all source domains is referred to as the best seed word set. If the best seed words in the current iteration produce an improvement in accuracy of the sentiment lexicon across all source domains (compared with the previous iteration for the corresponding domains), then the process continues with the next iteration for a new seed generation. This process stops when the best seed words in the current iteration do not improve the accuracy of the sentiment lexicon across all source domains (compared with the previous iterations for the corresponding domains). In this case, the best seed words in the previous iteration are declared the final corpus-learned seed word set. Table 1 presents the

TABLE 1. Algorithm: Corpus-learned polarity seed word learning process.

Inputs: The set D_1, D_2......, Ds of S source domains.
SD_SL_1, SD_SL_2......, SD_SL_s are the source domain sentiment lexicons from Iteration 1*.
Output: Corpus-learned polarity seed sets and improved source domain sentiment lexicons.

1. Set Iteration = 2
2. Generate corpus-learned polarity seed set
   - 2.1. Collect the same polarity OOW set (SPOOW) across SD_SL_1, SD_SL_2…..., SD_SL_s of the previous iteration
   - 2.2. N_SPOOW denotes a negative SPOOW set and P_SPOOW denotes a positive SPOOW set, with scores averaged across S source domains
   - 2.3. Sort P_SPOOW and N_SPOOW in ascending order of the absolute score
   - 2.4. For the varying value of "*P*"
     - Select the top *"P"* seed words from P_SPOOW and N_SPOOW to form a new seed set
     - Apply the new polarity seed set to the single-domain sentiment lexicon learning process* across S source domains
   - 2.5. Select the optimal *"P"* value and the corresponding seed set that produces the highest accuracy across S source domains.
   - 2.6. Compare the accuracy of the current iteration with that of the previous iteration across S source domains
     - If the accuracy is improved across S source domains
       - Store the selected top *"P"* seed set as the best seed set and use it as the seed set for the next iteration
       - Set Iteration = Iteration + 1
       - Repeat process: go to Step 2
     - else
       - The top seed set from (Iteration - 1) is considered to be a corpus-learned polarity seed word set
       - Stop iterations and exit the learning process: go to Step 3
3. Exit

** The application of the single-domain sentiment lexicon learning process is explained in Section III-C-1.*

corresponding algorithm. These corpus-learned polarity seed words are used in Stage 4, which is described in Section III-D.

D. STAGE 4: CORPUS-LEARNED POLARITY SEED TRANSFER

In the previous stage, knowledge from the source domains was combined to generate corpus-learned polarity seed words. In this stage, the corpus-learned polarity seed word knowledge learned from source domains of the same genre is transferred to the target domain. The target domain uses the set of corpus-learned seed words as labeled input and learns the domain-based sentiment lexicon using the single-domain sentiment-lexicon learning process presented in Section III-C-1. The generated sentiment lexicons for the target domain are denoted TD_SL_i, $i = 1, 2\ldots, M$. Thus, genre-level seed knowledge-transfer learning is accomplished through the corpus-learned polarity seed word set. The review data that were used to create and evaluate the sentiment lexicons are described in Section IV.

IV. DATA SOURCE

The proposed model was experimentally validated using the Amazon consumer review dataset [74], [75], a large multidomain dataset. Our experiments used data from 24 domains (e.g., Beauty & Personal Care, Cell Phones, Clothing, Shoes, Jewelry, etc.). The dataset was obtained from the Stanford Network Analysis Project database [74] from the period 2007–2013 and from the University of California San Diego database [75] from the period 2013–2014. An average review contained approximately 60–100 words. The model's performance was evaluated using the hold-out sampling technique. The source domain data were split into a training set and a validation set, while the target domain data were split into a training set and a test set. There was no overlap between the training, test, and validation sets. The training set for each domain contained 80,000 unlabeled reviews. The validation and test sets contained 3,000 and 20,000 labeled reviews, respectively, with a balanced polarity class distribution. Reviews with one and two stars were labeled as negative, while reviews with four and five stars were labeled as positive. The proposed model only considered OOWs that appeared a minimum of five times in the domain-level data. The OOW count at the domain level was in the ranges of 2–3 and 0.4–0.7 million for the training and test datasets, respectively.

The proposed model was compared with several baselines using standard evaluation measures. All baselines and evaluation measures are discussed in Section V.

V. BASELINE AND EVALUATION MEASURES

The experimental results of the proposed model were compared with existing lexicons and models using standard evaluation measures. The baselines and evaluation measures are described in subsections V-A and V-B, respectively.

A. BASELINES

The experimental results of the proposed model were compared with those of six baselines. The first four baselines were standard sentiment lexicons that are widely used as baselines in sentiment analysis research [28], [40]. The fifth and sixth baselines were widely used models, that have been adapted from the method originally proposed by Turney and

Littman [25]. These baselines are mainly used in in-domain sentiment lexicon learning.

1) HU & LIU LEXICON

The first baseline was the Hu & Liu lexicon [23], a popular polarity lexicon created from consumer product review data. It contains a list of 2,006 positive and 4,783 negative words that were manually extracted from consumer reviews. The Hu & Liu lexicon was evaluated by assigning a score of $+1$ if the word was in the positive list, -1 if the word was in the negative list, and 0 if the word was absent from the lexicon.

2) MULTIPERSPECTIVE QUESTION ANSWERING (MPQA) LEXICON

The second baseline was the MPQA subjectivity lexicon [12], which contains 2,301 positive words and 4,149 negative words. It also contains neutral words that were not considered in this study. This manually created lexicon stores subjectivity clues for words in terms of their discrete strength of intensity (strong or weak). Each word entry has a predefined POS tag: noun, verb, adverb, adjective, and other tags. For evaluation, POS tags were not considered, and conflicting words were removed. The polarity orientation and polarity strength were used to assign a score of -1, -2, $+1$, and $+2$ to weak negative, strong negative, weak positive, and strong positive words, respectively; otherwise, a score of 0 was assigned.

3) NRC EMOTION LEXICON

The NRC emotion lexicon, which is described in [76], is a polarity lexicon that was manually created using Amazon Mechanical Turk. It contains frequently occurring nouns, verbs, adjectives, and adverbs, eight emotions, and positive/negative polarity. For evaluation, 2,231 words with positive polarity were assigned a score of $+1$, while 3,324 words with negative polarity were assigned a score of -1.

4) SentiWordNet (SWN)

The fourth baseline was SentiWordNet 3.0 (SWN) [10], which is one of the most commonly used baselines in lexicon learning research [40]. SWN synsets for a word contain positive, negative, and objective scores. In this study, a word's score was calculated as the difference between the average positive score and average negative score across all synsets without considering the POS tags. The resulting score's polarity orientation was used to label a word as positive or negative. Accordingly, 2,527 positive polarity sentiment words and 2,380 negative polarity sentiment words were extracted from SWN.

5) SEMANTIC ORIENTATION POINT-WISE MUTUAL INFORMATION (SOPMI) MODEL

The fifth baseline was a sentiment lexicon learning model that is based on SOPMI [25], [77]. For a given sentiment, *word*, and a standard polarity word, *seed,* a PMI score was calculated based on the probability of co-occurrence and individual occurrence at the domain level, as given by Equation (8).

$$PMI\left(word, seed\right) = log(\frac{p\left(word, seed\right)}{p\left(word\right).p\left(seed\right)}) \tag{8}$$

Using the PMI score, the SOPMI score for the sentiment *word* was calculated as given by Equation (9), which uses the seven standard pairs of seed words that were proposed in reference [25]. Initial Pseeds (positive seed words) and initial Nseeds (negative seed words) are defined as follows:

- ***Initial Pseeds***: "excellent," "good," "fortunate," "correct," "nice," "superior," "positive."
- ***Initial Nseeds***: "poor," "bad," "unfortunate," "wrong," "nasty," "inferior," "negative."

Each word expressing a sentiment was assigned a score by subtracting the average positive PMI score from the average negative PMI score. A positive score for a word that expresses a sentiment is the average score of the mutual association of the word with positive seed words. Similarly, a negative score for a word that expresses a sentiment is the average score of the mutual association of the word with negative seed words. The positive average score was subtracted from the negative average score and assigned to the sentiment *word.*

$$\begin{aligned} SOPMI\left(word\right) &= \left(\frac{1}{|Pseeds|}\sum\nolimits_{Pi\in Pseeds\forall i=1to7} PMI\left(word, Pi\right)\right) \\ &\quad - \left(\frac{1}{|Nseeds|}\sum\nolimits_{Ni\in Nseeds\forall i=1to7} PMI\left(word, Ni\right)\right) \end{aligned} \tag{9}$$

Using this baseline model, each domain created its own lexicon. The number of positive sentiment words was in the range of 2,500–5,200, and the number of negative sentiment words was in the range of 3,600–7,700. Table 2 presents the statistics of the sentiment lexicons generated using the SOPMI model for the 24 domains.

6) SPS MODEL

This baseline is the single-domain sentiment lexicon learning process described in Section III-C-1. The SPS model is based on LSA and uses the abovementioned seven standard pairs of seed words [25]. Table 2 provides the SPS model sentiment lexicon statistics for 24 domains. The number of positive sentiment words was in the range of 3,255–6,748, and the number of negative sentiment words was in the range of 3,623–7,723.

B. EVALUATION MEASURES

The focus of our approach was not on classification task but on the evaluation of the target domain sentiment lexicon. A review score was calculated by summing the scores of all sentiment words from the processed target domain sentiment lexicon. The sentiment words not found in the sentiment lexicon were assigned a score of 0. A review was classified as positive or negative based on the polarity of the aggregate score.

The proposed model sentiment lexicon and the six baseline sentiment lexicons were evaluated using four standard

TABLE 2. Domain-wise sentiment lexicon statistics generated using the semantic orientation point-wise mutual information (SOPMI) and seven pair seed (SPS) models.

SL. No	Domains	SPS model #Sentiment words		SOPMI model #Sentiment words	
		Positive	Negative	Positive	Negative
1	Apps for Android	4,867	5,052	4,111	4,525
2	Baby Products	4,429	3,991	3,471	4,043
3	CD Vinyl	5,303	7,723	5,171	6,374
4	Patio Lawn Garden	4,320	4,856	3,370	4,778
5	Kindle Store	6,419	7,175	5,286	6,640
6	Movies & TV	5,596	6,488	4,841	5,854
7	Amazon Instant Video	5,705	5,542	4,619	5,219
8	Automotive	4,425	4,862	3,606	4,724
9	Beauty Products	3,311	4,971	2,916	4,356
10	Book	6,748	6,853	4,255	7,747
11	Cell Phone	3,255	3,623	2,520	3,635
12	Clothe Shoe Jewel	3,682	4,469	2,922	4,320
13	Digital Music	5,436	5,132	3,211	6,062
14	Electronics	4,427	4,093	3,289	4,317
15	Grocery & Gourmet Food	3,933	4,902	3,281	4,554
16	Home & Kitchen	3,539	5,086	2,904	4,822
17	Health & Personal Care	5,543	5,563	4,309	5,578
18	Musical Instrument	4,787	4,712	3,801	4,703
19	Office Products	4,429	3,910	3,026	4,416
20	Pet Supplies	3,791	5,300	3,300	4,724
21	Sports & Outdoor	4,554	4,925	3,464	5,001
22	Tools Home	4,073	4,612	3,150	4,625
23	Toys & Games	4,525	4,967	3,490	4,833
24	Video Game	4,499	4,808	3,681	4,567

performance measures: precision, recall, F1 score, and accuracy. These measures were calculated separately for positive and negative reviews using inputs from the confusion matrix representing the target and predicted classes, as given in Table 3.

TABLE 3. Confusion matrix representing the target and predicted classes.

Class		Predicted	
		'+' Class	'−' Class
Target	'+' Class	True Positive (*tp*)	False Negative (*fn*)
	'−' Class	False Positive (*fp*)	True Negative (*tn*)

Precision is the fraction of correctly tagged reviews among the total number of tagged reviews; it was calculated using Equation (10). *Recall* is the ratio of correctly tagged reviews to the total number of reviews with correct tags, as given by Equation (11). The F1 score represents the harmonic mean of the precision and recall values; it was calculated using Equation (12)

$$Precision_i = \frac{tp}{tp + fp} \tag{10}$$

$$Recall_i = \frac{tp}{tp + fn} \tag{11}$$

$$F1\ score_i = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{12}$$

where $i = p$ or n

In Equations (10)–(12), $i = p$ indicates a positive class and $i = n$ indicates a negative class. Accuracy represents the total number of reviews that are correctly classified; it was calculated using Equation (13), where *tp* and *tn* represent the number of correctly classified positive and negative reviews, respectively.

$$Accuracy = \frac{tp + tn}{tp + fp + tn + fn} \tag{13}$$

Section VI discusses the parameter settings of the proposed model.

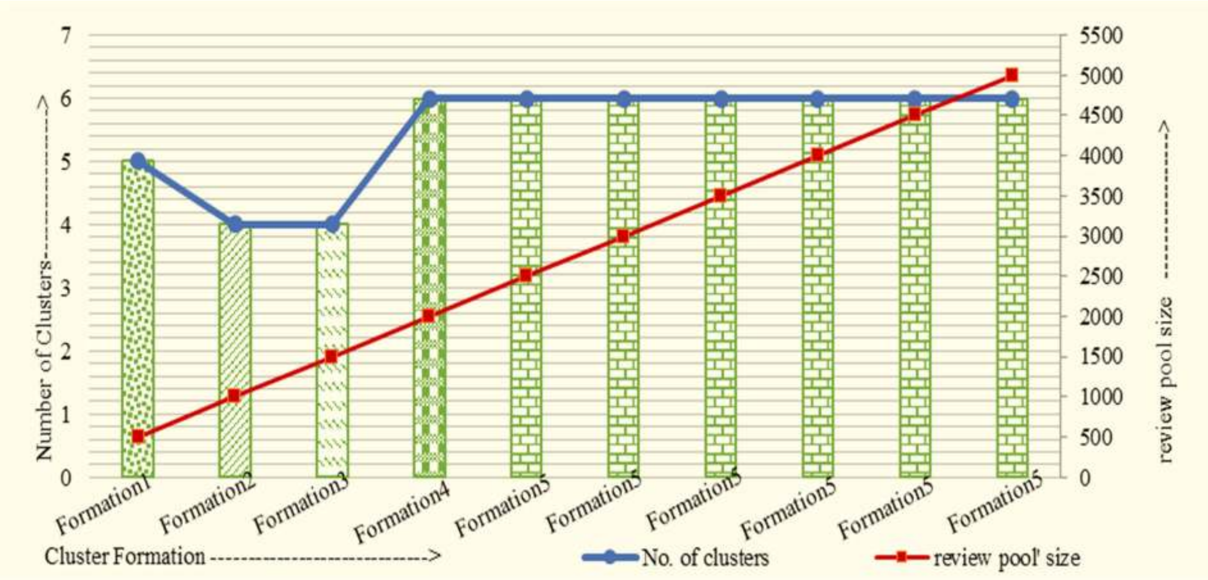


FIGURE 2. Domain cluster formation results.

VI. EXPERIMENTAL SETUP

The proposed model uses parameter settings in Stages 2 and 3, and the parameter settings of Stage 2 are used for identifying the source and target domains. Details are provided in Section VI-A. The Stage 3 parameters are used for corpus-learned seed word learning; they are discussed in Section VI-B. The proposed model was developed in Python using NLTK 3.09 [78] and Gensim 10 [79]. The computational complexity of the experiment is discussed in Section VI-C.

A. PARAMETER SETTING OF SOURCE AND TARGET DOMAIN SELECTION

The experiment used 24 domains that were grouped into clusters using the affinity propagation algorithm to identify the source and target domains. The process involved the creation of a review pool using positive reviews. The review pool was created for each domain by combining a fixed number of reviews. For example, a review pool for the Baby Products domain of size 1,000 was created by combining 1,000 reviews from the Baby Products domain. Each review pool represented a single domain. Thus, a review pool was created for all domains. Review pools of the same size were fed as input to the source and target domain selection process described in Section III-B. The selection process was repeated for different review pool sizes by starting at 500 reviews and increasing the pool size in increments of 500. The results of the selection process are presented in Figure 2. As illustrated in Figure 2, cluster formations can be visualized in terms of segregation into a number of clusters and different cluster formations, review pool sizes, and members of clusters.

Cluster formation using different cluster members is represented using a unique pattern. For example, for Formation 1 in Figure 2, the number of clusters (4) is denoted by a blue line mapped on the left vertical axis, the review pool size (500) is denoted by a red line mapped on the right vertical axis, and the cluster formation and cluster members are denoted by a green dotted pattern. The cluster formations are designated Formation 1 to Formation 5 for different review pool sizes, as shown in Figure 2. Formations 1–4 correspond to review pool sizes of 500, 1,000, 1,500, and 2,000, respectively, and differ in terms of the number of clusters and the cluster members. The review pool sizes of 2,500, 3,000, 3,500, 4,000, 4,500, and 5,000 were found to generate the same cluster formation, which is designated Formation 5 in Figure 2. The cluster formation continued to be Formation 5 when the experiment was repeated with increased pool sizes; therefore, the selection process was terminated. Formation 5, which contained six clusters, was selected as the output of the source and target domain selection stage. Table 4 lists the cluster representatives and members. The six cluster representatives were selected as the source domains, and the 18 cluster members were selected as the target domains.

B. PARAMETER SETTINGS OF CORPUS-LEARNED POLARITY SEED WORD GENERATION

The model's learning process requires the tuning of two main parameters. The first parameter is k in the single-domain sentiment lexicon learning process, while the second parameter is the best seed word set P in the corpus-learned seed word generation process. The single-domain sentiment lexicon learning process experimentally selects the value

TABLE 4. Cluster centers and cluster members in formation 5.

Cluster No.	Cluster representatives (Source Domains = S)	Cluster members (Target Domains = M)
1	Apps for Android	Video Game
2	Baby Products	Beauty Products, Cell Phone, Clothe Shoe Jewel, Grocery & Gourmet Food, Musical Instrument, Pet Supplies, Toys & Games
3	CD Vinyl	Digital Music
4	Patio Lawn Garden	Automotive, Health & Personal Care, Home & Kitchen, Office Products, Electronics, Sports & Outdoor, Tools Home
5	Kindle Store	Book
6	Movies & TV	Amazon Instant Video

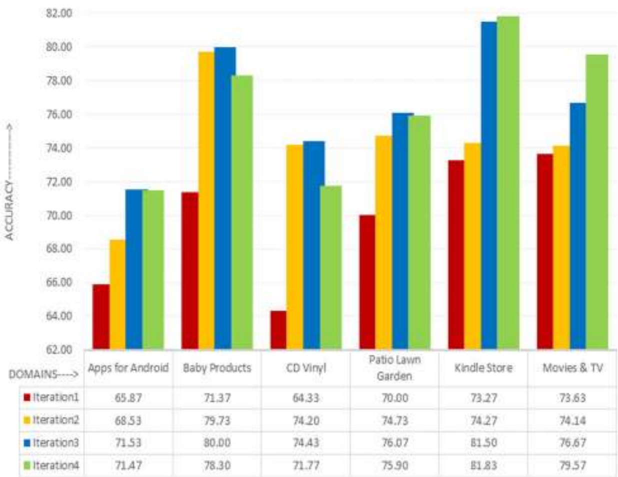


FIGURE 3. Source domain sentiment lexicon accuracy comparison on the validation set across iterations.

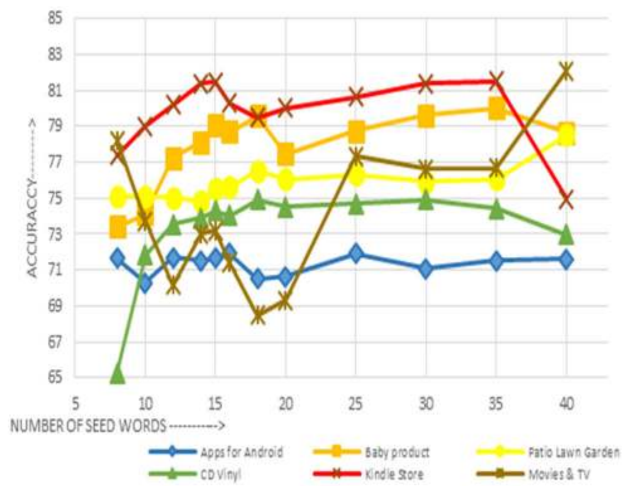


FIGURE 4. Source domain sentiment lexicon accuracy comparison of top P seed words used in the last iteration of corpus-learned polarity seed word learning process on the validation set.

of k as part of the matrix reconstruction step described in Section III-C-1. The process was optimized to identify the best value of k. The sentiment lexicon generation experiment was repeated for different values of k, which varied from 50 to 550 in increments of 25. We selected the sentiment lexicon that produced the highest accuracy on the validation set.

The corpus-based polarity seed word learning process involved the selection of new best seed words in each iteration. The initial iteration of the sentiment lexicon learning process used the initial Pseeds and initial Nseeds, as described for the SPS model in Section V. In subsequent iterations, seed words were selected experimentally from the top P seed words generated by the learning process of the previous iteration. The value of P was selected such that it corresponded to the P word polarity-seed-set that achieved the highest accuracy across all source domains. Figure 3 presents the accuracy of all source domains on the validation set in successive iterations. It can be observed that in iteration 4, there is accuracy drop for the source domains Apps for Android, Baby Products, and Patio Lawn Garden. Thus, the results of iteration 3 were selected as the final output of the proposed model. As illustrated in Figure 3, a significant 3–10 points accuracy improvement was achieved over the first iteration of the proposed model. Figure 4 displays the variation in the source domain's classification accuracy in the final iteration for different numbers of top P seed word sets. Stable results were observed for sets with over 20 seed words, while inconsistent accuracies were observed for sets with over 35 seed words. The set with the top 35 seed words was considered the corpus-learned polarity seed word set. The seed words are listed in Table 5 and are used in the proposed model to build target domain sentiment lexicons.

C. COMPUTATIONAL COMPLEXITY

The learning process uses LSA (in particular, truncated SVD) to generate a single-domain sentiment lexicon. The time complexity of the single-domain sentiment lexicon learning process is $O(qk^2)$, where q is the number of sentiment words. Although the complexity may appear quadratic, the value of k remains constant and does not change with the domain. In our approach, k is set to the largest singular value, which is 550. Matrix construction and decomposition are performed only once in the single-domain sentiment lexicon learning

TABLE 5. Proposed corpus-learned polarity seed words.

Pseeds Positive Seed Words	Nseeds Negative Seed Words
elegant, perfect, pleasantly, bonus, compliment, fabulous, terrific, detract, affordable, highly, beginner, suprb, companion, complement, easy, excellent, organize, pleased, fantastic, beautifully, outstanding, overjoyed, exciting, magical, nicely, thoughtful, favorite, instructor, hesitate, comfort, contemporary, relax, wonderful, drawback, ample	waste, junk, insult, crap, garbage, worthless, disgusted, piss, awful, horribly, substandard, donate, refund, ashamed, worst, horrible, trash, pathetic, gimmick, nonsense, nerve, stupid, aggravation, badly, recourse, defective, crappy, joke, damn, annoys, reimburse, terrible, useless, scam, torture

TABLE 6. Statistics of proposed model's target domain sentiment lexicons (TD_SL) generated using corpus-learned polarity seed words.

SL. No.	Target Domain	#Positive Sentiment words	#Negative Sentiment words
1	Amazon Instant Video	5,792	5,460
2	Automotive	4,983	4,304
3	Beauty Products	4,105	4,177
4	Book	6,968	6,639
5	Cell Phone	3,780	3,099
6	Clothe Shoe Jewel	3,612	4,539
7	Digital Music	5,661	4,901
8	Electronics	4,215	4,304
9	Grocery & Gourmet Food	4,521	4,314
10	Home & Kitchen	3,957	4,661
11	Health & Personal Care	5,972	5,137
12	Musical Instrument	5,081	4,429
13	Office Products	4,292	4,046
14	Pet Supplies	4,516	4,581
15	Sports & Outdoor	4,944	4,548
16	Tools Home	4,237	4,460
17	Toys & Games	4,970	4,527
18	Video Game	4,552	4,770

process. In the corpus-based seed word learning process described in Section III-C-2, the single-domain sentiment lexicon learning process is repeated for t sets of top seed words and S source domains, and the entire process is repeated i times. The aggregate time complexity is $O(qtiSk^2)$. Therefore, the time complexity can be considered linear, making the model scalable in practice. The results of the evaluation and baseline comparison are presented in Section VII.

VII. EVALUATION AND COMPARISON OF EXPERIMENTAL RESULTS

This section presents the experimental results and their statistical significance. Section VII-A presents the experimental results in two setups. In the first setup, the target domain results are evaluated and compared with existing lexicons and baseline models. In the second setup, the proposed model is compared with recent studies [39]. In Section VII-B, the statistical significance test is discussed.

A. EXPERIMENTAL RESULTS

In the first setup of the experimental results, the corpus-learned polarity seed word knowledge generated from the source domains was transferred to learn sentiment lexicons for 18 target domains of the same genre. This process transferred knowledge of the corpus-learned polarity seed words listed in Table 5. These seed words were used for the target domain sentiment lexicon generation using the process described in Section III-C-1. Table 6 presents the sentiment word polarity statistics of the target domain sentiment lexicons. The number of positive sentiment words varied between 3,600 and 7,000, and the number of negative sentiment words varied between 3,100 and 6,600. The target domain sentiment lexicons were evaluated using a classification test on the target domain test set, as described in Section V.

The results of the proposed model were compared with those of the baseline models. The accuracy results summarized in Table 7 indicate that independent use of general-purpose lexicons did not improve performance, whereas the use of domain-based sentiment lexicons led to improved results. The SPS and SOPMI baseline models learned a domain-based sentiment lexicon and performed best among the six baselines. Overall, our unsupervised genre-based model outperformed the six baselines in terms of accuracy (Table 7). On average, the accuracy of our model was up to 15.50 and 24.77 points higher than those of the SPS and SOPMI baselines, respectively. In addition, our model's accuracy was up to 27.70 points higher than that of the popular SentiWordNet lexicon. The highest accuracy of 86.89% was observed in the Toys & Games domain.

We conducted a deeper analysis in terms of precision, recall, and F1 score for positive and negative classes. The results are presented in Table 8. Positive and negative class F1 score results are comparable with the exception of some domains, such as Amazon Instant Video and Office Products. Negative class recall was observed to be higher than positive class recall in almost all domains except for Amazon Instant Video and Office Products. Generally, negative reviews are more descriptive, as reviewer express annoyance and dislike through overelaboration, whereas positive reviews are generally more precise. The average number of OOWs per review in the test data across all domains was 23.4 and 32.5 in the positive and negative labeled data, respectively. We compared our model with a model recently proposed by Xing *et al.* (2019) [39]. Their rule-based iterative model is a cognitive-inspired approach that uses incorrectly predicted outcomes as feedback. The authors conducted experiments using Hu & Liu, SentiWordNet, and SenticNet lexicon adaptation. They observed that the SenticNet lexicon adaptation demonstrated the best performance. The model results were tested on the popular dataset from Blitzer *et al.* (2007) [29]. A comparison of results was performed on five domains, i.e., the Apparel, Electronics,

TABLE 7. Comparison of accuracy of target domain sentiment lexicon of the proposed and baseline models.

SL. No	Target Domain	Accuracy (%)						
		Hu & Liu	MPQA	NRC	SWN	SOPMI	SPS	Our Model
1	Amazon Instant Video	65.55	64.95	56.27	62.65	71.63	72.34	**75.14**
2	Automotive	58.42	59.09	47.39	59.55	68.60	72.60	**77.72**
3	Beauty Products	61.23	60.59	51.64	60.41	72.13	69.36	**80.14**
4	Book	62.64	62.23	53.66	57.77	55.03	64.41	**77.84**
5	Cell Phone	61.13	60.02	50.56	59.29	71.94	73.50	**77.30**
6	Clothe Shoe Jewel	62.03	61.44	53.45	61.07	70.34	69.95	**79.22**
7	Digital Music	64.13	61.69	55.07	60.32	53.01	77.75	**77.78**
8	Electronics	60.84	60.45	50.53	59.01	70.40	74.78	**78.44**
9	Grocery & Gourmet Food	62.77	61.33	52.87	63.08	74.94	77.69	**79.28**
10	Home & Kitchen	64.03	63.94	53.76	60.19	75.69	78.52	**82.12**
11	Health & Personal Care	57.62	58.13	48.91	59.18	65.94	69.73	**76.78**
12	Musical Instrument	62.51	61.27	52.36	58.57	75.32	77.22	**80.42**
13	Office Products	63.99	64.05	53.75	60.80	75.32	**79.20**	77.56
14	Pet Supplies	57.84	59.14	53.00	60.06	61.85	61.14	**74.25**
15	Sports & Outdoor	61.89	60.84	52.88	60.04	69.08	71.09	**78.75**
16	Tools Home	62.61	61.53	52.10	59.73	71.75	72.09	**80.03**
17	Toys & Games	66.57	58.59	56.15	59.29	80.44	77.07	**86.89**
18	Video Game	64.18	61.03	55.62	61.77	70.13	66.22	**81.76**

TABLE 8. Experimental results of the proposed model for target domain sentiment lexicon (TD_SL) for positive and negative classes with precision, recall, and F1 score evaluation.

SL. No	Target Domain	Positive Class			Negative Class		
		Precision (%)	*Recall* (%)	*F1 score* (%)	*Precision* (%)	*Recall* (%)	*F1 score* (%)
1	Amazon Instant Video	81.50	65.04	72.35	70.92	85.24	77.42
2	Automotive	77.31	78.49	77.90	78.16	76.96	77.56
3	Beauty Products	78.81	82.43	80.58	81.58	77.84	79.67
4	Book	78.28	77.05	77.66	77.41	78.62	78.01
5	Cell Phone	77.34	77.22	77.28	77.26	77.38	77.32
6	Clothe Shoe Jewel	80.71	76.81	78.71	77.88	81.64	79.72
7	Digital Music	71.94	91.09	80.39	87.86	64.47	74.37
8	Electronics	83.41	71.00	76.71	74.76	85.88	79.94
9	Grocery & Gourmet Food	75.18	87.43	80.84	84.98	71.14	77.45
10	Home & Kitchen	90.92	71.37	79.97	76.44	92.87	83.86
11	Health & Personal Care	75.93	78.44	77.16	77.70	75.13	76.39
12	Musical Instrument	83.47	75.88	79.49	77.89	84.97	81.28
13	Office Products	89.07	62.82	73.68	71.28	92.29	80.44
14	Pet Supplies	78.40	66.94	72.22	71.16	81.56	76.01
15	Sports & Outdoor	81.06	75.03	77.93	76.76	82.47	79.51
16	Tools Home	85.00	72.93	78.50	76.30	87.13	81.36
17	Toys & Games	87.54	86.03	86.78	86.27	87.75	87.00
18	Video Game	84.15	78.26	81.10	79.68	85.26	82.38

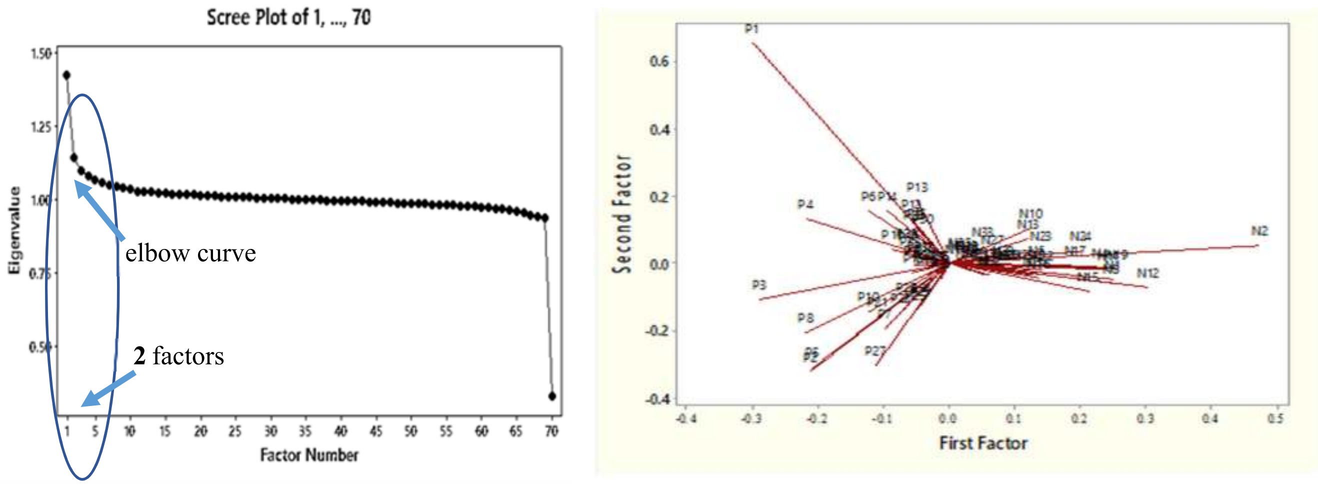
Kitchen, Healthcare, and Movie domains. We used the proposed model's respective target domain sentiment lexicons for evaluation. We excluded results from the finance domain, as they were unrelated to our genre-based approach.

TABLE 9. Comparison of the accuracies of the proposed model and model proposed by Xing *et al.* [39].

Model\Domains	Apparel	Electronics	Kitchen	Health Care	Movie
Xing et al. 2019 [39]	74.70	69.20	69.30	65.70	77.90
Our Model	**79.20**	**77.65**	**76.45**	**78.40**	**82.80**

TABLE 10. Comparison of the results of the proposed model and baselines (column 1) using Tukey's post-hoc analysis test.

Our Model Vs. Baseline	Difference in Mean	Std. Err. Of difference	t-value	Adjusted p-value	95% Conf. Interval
HU&LIU	16.747	1.261	13.280	0.000	(12.963 - 20.529)
MPQA	17.839	1.261	14.140	0.000	(14.056 - 21.622)
NRC	26.192	1.261	20.770	0.000	(22.408 - 29.974)
SWN	18.813	1.261	14.920	0.000	(15.030 - 22.596)
SOPMI	9.327	1.261	7.390	0.000	(5.543 - 13.109)
SPS	6.487	1.261	5.140	0.000	(2.703 - 10.269)



(a). Scree plot displaying elbow curve with two factors.

(b). Factor analysis results displaying seed-word-polarity class binding.

FIGURE 5. Analytical visualization of corpus-learned polarity seed words. (a). Scree plot displaying elbow curve with two factors. (b). Factor analysis results displaying seed-word-polarity class binding.

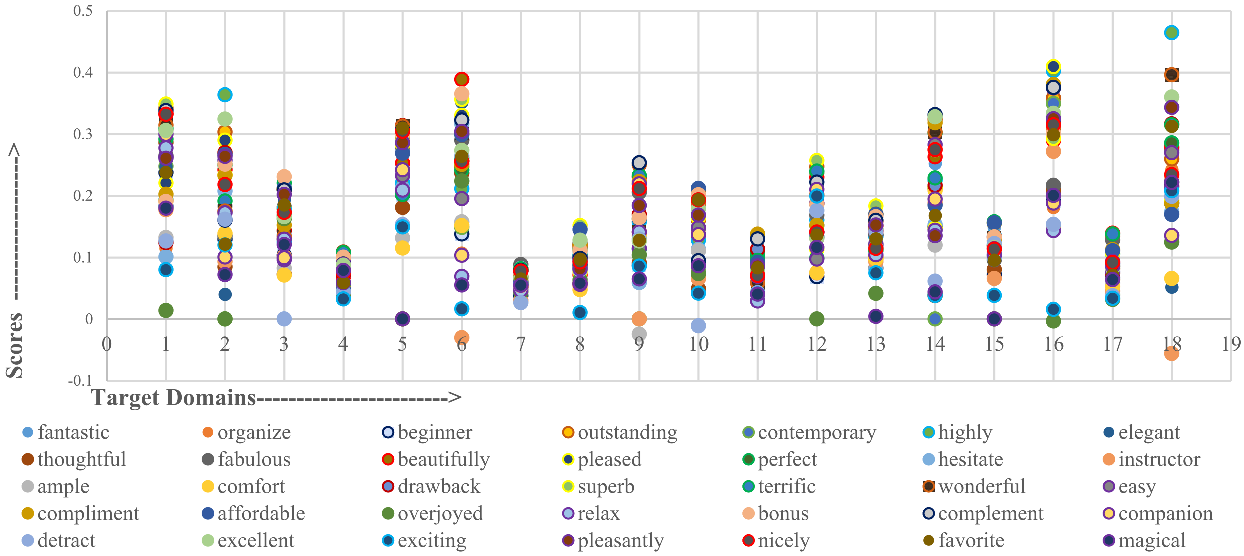
Our evaluation used 1,000 positive and negative reviews each from Blitzer *et al.* (2007) [29]. The comparative results displayed visible improvements in all domains, as presented in Table 9. Our model showed a 4.5–12.7 points accuracy improvement across domains. To further evaluate our proposed model, we performed a statistical significance test using the accuracy measure in comparison with the baselines.

B. STATISTICAL SIGNIFICANCE TEST

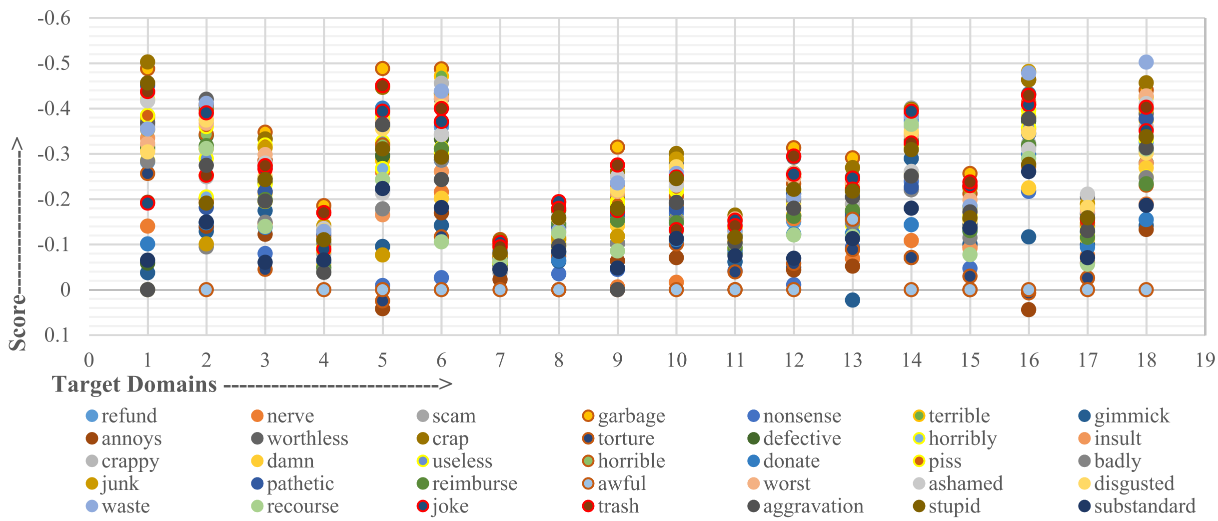
A single-factor analysis of variance (ANOVA) test was performed to determine whether there was a significant difference in accuracy between the proposed model and the baselines across various domains. Tukey's post-hoc test revealed models having honest significant differences. The results of statistical analysis with Tukey's post-hoc test at a 95% confidence level are presented in Table 10. The p-value obtained in the ANOVA test was less than 5% ($F(6, 119) = 97.95$, $p = 0.0000$), which implies that there was a significant difference between the models. The results reveal that the proposed model significantly differed from all baseline models, demonstrating a statistically significant increase in the mean of 6.487 points at a 95% confidence interval ($t(17) = 5.140$, $p < 0.05$) in comparison with the SPS model. Therefore, our proposed model's improvement in accuracy compared to baselines is statistically significant. The strength of our proposed model lies in two outcomes: the corpus-learned polarity seed words that are used as the source for transfer learning and the learned target domain lexicons. In Section VIII, we provide a deeper analysis of these two contributions.

VIII. ANALYSIS AND DISCUSSION OF EXPERIMENTAL RESULTS

A thorough analysis of the proposed model was performed from the following three perspectives:



(a). Positive polarity words.



(b). Negative polarity words.

FIGURE 6. **Corpus-learned seed word scores across 18 target domains. (a). Positive polarity words. (b). Negative polarity words.**

- Analysis of seed word polarity class binding and target domain representation
- Comparison of the target domain sentiment lexicon coverage and polarity strength analysis

Case study of sentiment word polarity, polarity strength, and domain-specific divergence across target domains.

A. SEED WORD POLARITY CLASS BINDING AND TARGET DOMAIN REPRESENTATION

The proposed model's target domain sentiment lexicon results demonstrate the representative nature of the corpus-learned polarity seed words and their relation with their respective polarity classes. A cross-investigation of these properties was performed using a factor-analysis technique [73]. A TF-DIF vector for a corpus-learned seed word was formed by concatenating the corpus-learned seed word vectors from all source domains. The concatenated TF-DIF vector acted as input to the factor-analysis technique. The process involved 35 positive and negative corpus-learned polarity seed words each.

Figure 5(a) presents the factor analysis by plotting corpus-learned seed words on the x-axis and eigenvalues on the y-axis in a scree plot. Through its elbow point, the scree plot reveals that the orientation of the corpus-learned seed words is along with two factors. In the loading plot illustrated in Figure 5(b), positive seed words are represented using the second factor, whereas negative seed words are represented using the first factor. The 35 positive corpus-learned polarity seed words are denoted by *P1–P35*, while the 35 negative corpus-learned polarity seed words are denoted by *N1–N35*. As shown in Figure 5(b), the positive and negative seed word sets are strongly bound to their respective

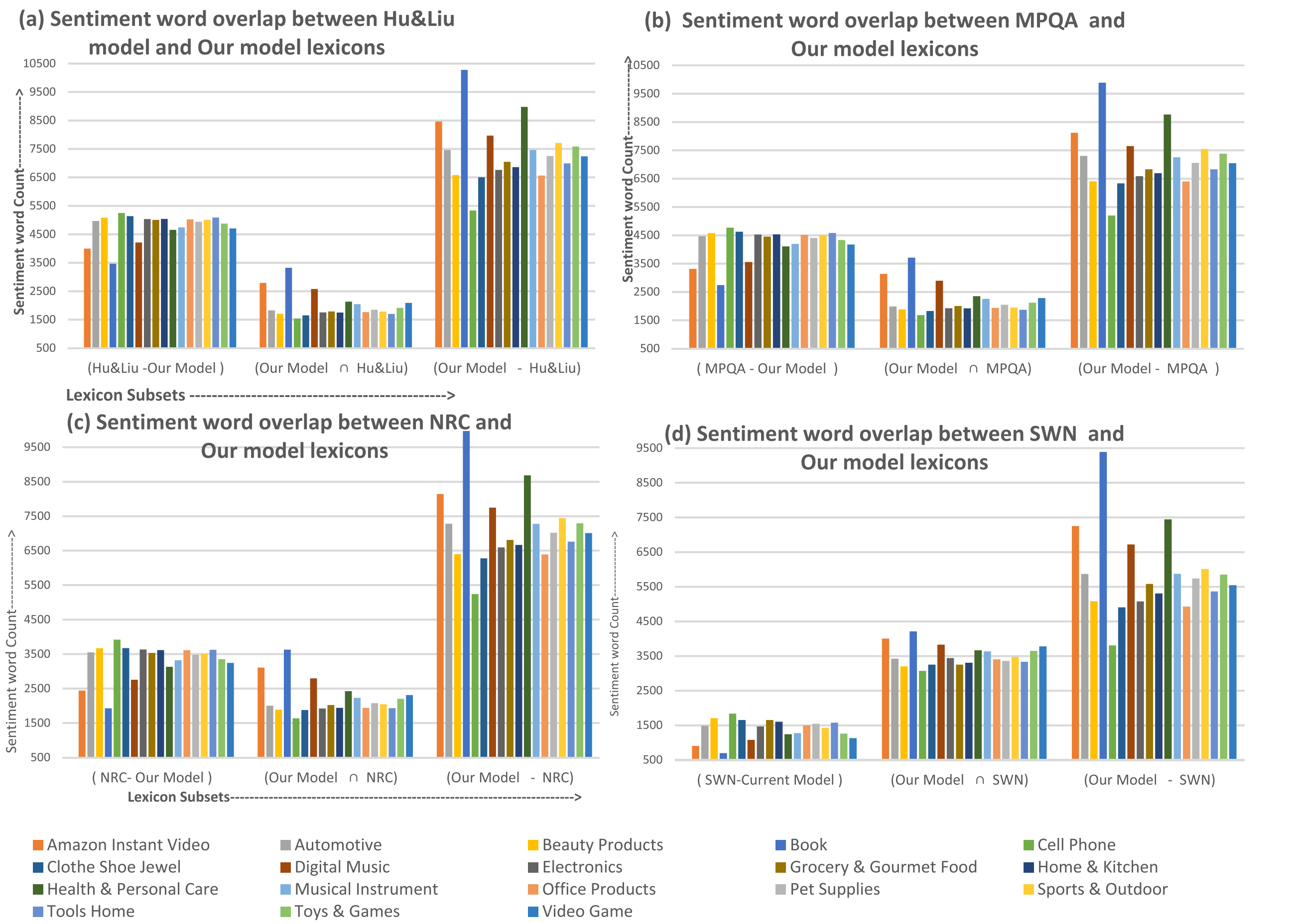


FIGURE 7. **Proposed model's domain-wise sentiment word overlaps and non-overlaps with baselines' sentiment lexicons.**

polarity classes. This reveals the polarity strength and class representativeness of the corpus-learned polarity seed words. Thus, the proposed model can generate lexicons that perform well across all target domains.

We further investigated the representativeness of corpus-learned polarity seed words by verifying their polarity classes in the target domains. Figures 6(a) and 6(b) display the polarity scores of positive and negative corpus-learned seed words across 18 target domains respectively. The same serial order of domains was followed as used in all tables. Each domain in Figures 6(a) and 6(b) (1–18) represent 35 corpus-learned seed word scores as a colored dot. The scores of seed words have the same polarity class and domain-based intensity variation.

Overall, the positive corpus-learned seed words were positive across the genre-based target domains, whereas the negative seed words were negative across these domains. However, there were some exceptions; for example, the seed word "awful" was absent from the Electronics and Digital Music domains, and the seed word "instructor" was absent from the Cell Phone domain. In addition, "instructor" was observed with slightly opposite polarity in Clothe Shoes Jewel and Video Games domains, as this word is rarely observed in these domains and has a negative connotation. The target domain seed word relevance was consistent with the seed binding results in the source domain. The analysis thus indicates that seed knowledge transfer was achieved by the proposed model.

B. TARGET DOMAIN SENTIMENT LEXICON COVERAGE AND POLARITY STRENGTH ANALYSIS

Sentiment lexicon coverage and sentiment word polarity strength have a major impact on sentiment evaluation. Therefore, it is necessary to evaluate the polarity lexicon coverage and sentiment scores used in the different baselines.

The first four baselines (Hu & Liu, MPQA, NRC, and SWN) are existing lexicons. The polarity lexicon coverage analysis of these four baselines in comparison with our model is presented in Figure 7, wherein the results are in three subgroups, comparing sentiment word overlap (i.e., words shared by our model and the baselines) and non-overlap (i.e., sentiment words present in our model but absent from the baseline models/lexicons and vice versa) statistics of the baseline model lexicon with our model's domain-based lexicon across target domains. The sentiment word overlap between the proposed model and the baseline models was in the range of 1,500–4,200. The number of domain-specific sentiment words present in the proposed

TABLE 11. Representative sentiment word score/polarity comparison between the proposed and baseline models for some target domains.

Domain	Sentiment word	SOPMI	SPS	Our Model	Domain	Sentiment word	SOPMI	SPS	Our Model
Book	Terrific	0.016	0.004	0.108	Pet Supplies	cuddle	0.000	0.012	0.266
	Incorrect	0.026	0.042	-0.061		negative	0.245	-0.096	-0.149
	Enjoy	-0.003	0.007	0.042		scuffle	0.000	-0.030	-0.157
	fold	-0.034	-0.031	-0.013		Stale	-0.061	-0.148	-0.158
	sensible	-0.055	0.018	0.010		degrade	-0.010	-0.071	-0.132
	Story	-0.020	-0.001	0.033		pathetic	-0.109	-0.187	-0.226
Cell Phone	Ease	0.045	0.070	0.334	Sports & Outdoor	stylish	0.131	0.036	0.133
	Wonderful	0.100	0.033	0.313		comfort	0.100	0.049	0.119
	Lightweight	0.050	0.067	0.288		rugged	0.024	0.049	0.112
	Affordable	-0.001	0.041	0.269		defective	-0.053	-0.072	-0.154
	Crap	-0.039	-0.067	-0.446		substandard	0.863	-0.066	-0.137
	Defective	-0.038	-0.097	-0.296		faulty	-0.087	-0.060	-0.145
Clothe Shoe Jewel	Dressy	0.082	0.018	0.292	Tools Home	convenient	0.026	0.043	0.325
	Elegant	0.043	0.008	0.279		useful	0.065	0.010	0.187
	Exquisite	0.004	0.053	0.270		adjustable	-0.022	0.040	0.198
	Terrific	-0.033	-0.033	0.236		Cost	-0.005	-0.006	-0.204
	Terrible	-0.083	-0.068	-0.400		innovation	-0.066	0.042	0.121
	Shred	-0.126	-0.069	-0.225		beautiful	-0.002	0.000	0.302
Health & Personal Care	Healthy	-0.014	-0.014	0.059	Toys & Games	education	-0.028	0.013	0.025
	Rest	-0.008	0.011	0.015		fabulous	0.045	0.026	0.128
	Soft	0.044	0.056	0.092		tedious	0.034	0.024	-0.015
	Fat	0.045	0.050	-0.018		brave	0.029	-0.005	0.039
	Careless	0.100	-0.078	-0.114		affordable	-0.039	0.058	0.111
	Calm	-0.036	-0.033	0.032		defect	-0.080	-0.044	-0.0813
Home & Kitchen	Casual	0.021	0.054	0.098	Video Game	loony	-0.041	-0.027	0.491
	Exquisite	0.018	0.026	0.091		winner	0.018	0.023	0.273
	Luxurious	-0.012	0.029	0.046		entertain	-0.006	-0.004	0.354
	Lousy	-0.011	-0.167	-0.102		Error	-0.017	0.012	-0.220
	Trash	-0.063	-0.172	-0.141		jeopardy	-0.020	-0.004	-0.216
	homemade	-0.027	0.030	0.033		Virus	-0.020	-0.004	-0.216

model and absent from the baseline models was in the range of 3,800–10,200. These statistics explain the poorer performance of existing lexicon baselines.

The SPS and SOPMI models learn sentiment lexicons from a target domain corpus; thus, sentiment lexicon coverage in these two baselines was similar to that of the proposed model. A sentiment word score represents the polarity orientation and strength (score) that is reflected in the results. The proposed model and SOPMI and SPS models assign a real value score to a sentiment word that captures the sentiment orientation strength. Table 11 presents the sentiment scores of representative sentiment words from the SOPMI and SPS baseline models and the proposed model sentiment lexicons for some target domains. In the Pet Supplies domain, the review orientation was inclined toward pets. Therefore, the sentiment words "cuddle" and "scuffle" were relevant and denoted positivity and negativity, respectively. With the proposed model, these two sentiment words had a higher strength (score) than with the SOPMI and SPS models. In the Sports & Outdoor domain, the sentiment words "stylish," "comfort," and "rugged" were observed to be important, having a positive connotation. Similarly, "defective," "substandard," and "faulty" were observed to be important and having a negative connotation. These sentiment words were highly relevant in our model compared with the baseline models. For example, a consumer prefers a cell phone that is "affordable" and "lightweight" and provides "ease" of handling. The proposed model reveals that these sentiment words have high domain relevance.

Similar behavior was observed in other sentiment words across all target domain sentiment lexicons of our model. Two important achievements of the proposed model include the ability to enhance the polarity orientation and retain domain-specific polarity.

C. CASE STUDY OF SENTIMENT WORD POLARITY, POLARITY STRENGTH, AND DOMAIN-SPECIFIC DIVERGENCE ACROSS TARGET DOMAINS

This case study analyzed the quality and domain-independence of the target domain sentiment lexicon learned using our genre-based seed transfer learning approach. This analysis addressed the transfer learning challenge of learning the domain-relevant polarity, domain-relevant polarity score, and domain-independent sentiment words. We present an analysis using 17 representative sentiment words.

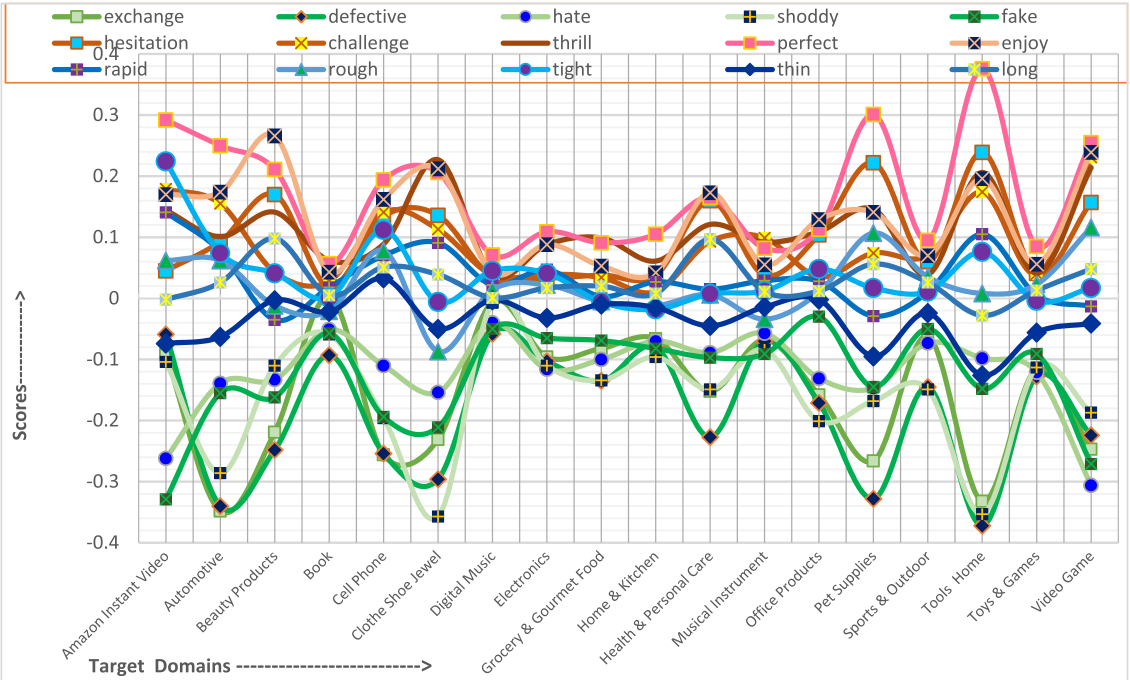


FIGURE 8. **Sentiment word scores displaying polarity and strength divergence across target domains.**

In our model, sentiment words have domain-based scores denoting their domain relevance. Thus, sentiment words were observed with varying polarity scores across target domains. The variation in scores of sentiment words is depicted in Figure 8. The positive sentiment words "hesitation," "challenge," "thrill," "perfect," and "enjoy" denoted positivity, and their scores varied across target domains based on their relevance to the domain. These words are indicated by shades of red in Figure 8. The following sentence from the Tools Home domain expressed positivity through the sentiment word "hesitation": "I bought this product without hesitation." Although "hesitation" generally has a negative connotation, it was always observed in a positive context. A sense of positivity was also expressed through the sentiment word "thrill" in the Video Games domain as follows: "My son loved this; I gave it to him as a gift, and he was thrilled."

The negative sentiment words "exchange," "defective," "hate," "shoddy," and "fake" exhibited negative score variations across domains. These words are indicated by shades of green in Figure 8. A consumer in the Beauty Products domain expressed anger about a low-quality product as follows: "This brand produces shoddy products and has lousy customer service." A sentiment word with positive polarity in one domain may be negative in another. The sentiment words "rapid," "rough," "tight," "thin," and "long" express varying polarity and are indicated by shades of blue in Figure 8. The poor quality of fabric in the Clothe Shoe Jewel domain was expressed using the sentiment word "rough," such as "This fabric is very rough." In contrast, the same sentiment word was used to express positivity in the Pet Supplies domain, such as " Our cat is rough and tough!" Similarly, the sentiment word "thin" expressed positivity in the Cell Phone domain, as thinner cell phones are more popular, whereas it expressed negativity in the Tools Home and Health & Personal Care domains.

Our proposed model also learned domain-specific sentiment words that were not present in all or source domains. The sentiment word "dressy" was present in the Clothe Shoes Jewel domain and several other domains but was absent from most domains. Similarly, the sentiment word "virus" was observed in the Video Games domain and other domains but was absent from the Automotive and Clothe Shoes Jewel domains.

The above analysis indicates that our proposed model learns target domains' specific polarity, polarity strength variation, and domain-specific sentiment words. The proposed model thus addresses the transfer learning challenge, the sparsity problem reported in a multidomain study [31].

IX. CONCLUSIONS AND FUTURE WORK

Our results suggest that the proposed genre-based model can significantly contribute to the field of sentiment lexicon learning. The strength of the model lies in the fact that it automatically learns from corpora and does not rely on existing seed words, lexicons, or labeled data. Using the proposed model, we demonstrated sentiment lexicon learning for multiple domains of the same genre. Thus, our model is scalable and can be applied to any set of domains of the same genre.

Source and target domains of the same genre were selected from multiple domains using the affinity propagation clustering algorithm. The selected source domains were used in the corpus-based polarity seed word learning process. The target domain sentiment lexicons were generated using the learned seed words. The seed word transfer learning process exhibited superior performance over the baselines with respect to almost all target domains. The most important output of the proposed approach was the genre-level set of corpus-learned polarity seed words. The representative nature and strong binding of the corpus-learned seed words were verified and reflected in the model's performance.

Parameter tuning in the model was performed using heuristic approaches. In the future, an optimization technique can be implemented to tune the parameters. This would further enhance the model's adaptability to various domains. Potential extensions of this research include investigation using different homogeneous knowledge-transfer tasks and the possible application of the model to different languages in circumstances wherein labeled data are unavailable.

REFERENCES

[1] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar. 2016.

[2] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool, 2012.

[3] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[4] A. Mabrouk, R. P. D. Redondo, and M. Kayed, "Deep learning-based sentiment classification: A comparative survey," *IEEE Access*, vol. 8, pp. 85616–85638, 2020.

[5] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proc. ACL*, vol. 2002, pp. 79–86.

[6] C. Priyanka and D. Gupta, "Identifying the best feature combination for sentiment analysis of customer reviews," in *Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI)*, Aug. 2013, pp. 102–108.

[7] B. Agarwal and N. Mittal, "Machine learning approach for sentiment analysis," in *Prominent Feature Extraction for Sentiment Analysis* (Socio-Affective Computing). Cham, Switzerland: Springer, 2016, pp. 21–45.

[8] Y. Fang, H. Tan, and J. Zhang, "Multi-strategy sentiment analysis of consumer reviews based on semantic fuzziness," *IEEE Access*, vol. 6, pp. 20625–20631, 2018.

[9] X. Fu, J. Yang, J. Li, M. Fang, and H. Wang, "Lexicon-enhanced LSTM with attention for general sentiment analysis," *IEEE Access*, vol. 6, pp. 71884–71891, 2018.

[10] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc. Int. Conf. Lang. Resour. Eval. (LREC)*, vol. 10, 2010, pp. 2200–2204.

[11] *General Inquirer*. [Online]. Available: http://www.wjh.harvard.edu/~inquirer/

[12] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proc. Conf. Hum. Lang. Technol. Empirical Methods Natural Lang. Process. HLT*, 2005, pp. 347–354.

[13] E. Cambria, S. Poria, D. Hazarika, and K. Kwok, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," in *Proc. 22nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.

[14] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for senti- ment analysis of social media text," in *Proc. 8th Int. AAAI Conf. Weblogs Social Media*, 2014, pp. 1–14.

[15] C. Hung, "Word of mouth quality classification based on contextual sentiment lexicons," *Inf. Process. Manage.*, vol. 53, no. 4, pp. 751–763, Jul. 2017.

[16] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 437–442.

[17] T. Loughran and B. McDonald, "When is a liability not a liability? textual analysis, dictionaries and 10-Ks," *J. Finance*, vol. 66, no. 1, pp. 35–66, 2011.

[18] N. Oliveira, P. Cortez, and N. Areal, "Stock market sentiment lexicon acquisition using microblogging data and statistical measures," *Decis. Support Syst.*, vol. 85, pp. 62–73, May 2016.

[19] A. Hussain and E. Cambria, "Semi-supervised learning for big social data analysis," *Neurocomputing*, vol. 275, pp. 1662–1673, Jan. 2018.

[20] Y. Han, L. Yuhong, and J. Zhigang, "Sentiment analysis via semi-supervised learning: A model based on dynamic threshold and multi-classifiers," *Neural Comput. Appl.*, vol. 32, pp. 5117–5129, Jan. 2019.

[21] V. Iosifidis and E. Ntoutsi, "Large scale sentiment learning with limited labels," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 1823–1832.

[22] F. Huang, S. Zhang, J. Zhang, and G. Yu, "Multimodal learning for topic sentiment analysis in microblogging," *Neurocomputing*, vol. 253, pp. 144–153, Aug. 2017.

[23] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining KDD*, 2004, pp. 168–177.

[24] Q. Guang, B. Liu, J. Bu, and C. Chen, "Expanding domain sentiment lexicon through double propagation," in *Proc. Int. Joint Conf. Artif. Intell., IJCAI*, 2009, pp. 1199–1204.

[25] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, no. 4, pp. 315–346, Oct. 2003.

[26] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[27] S. Sanagar and D. Gupta, "Roadmap for polarity lexicon learning and resources: A survey," in *Intelligent Systems Technologies and Applications ISTA* (Advances in Intelligent Systems and Computing), vol. 530. Cham, Switzerland: Springer, 2016, pp. 647–663.

[28] S. Deng, A. P. Sinha, and H. Zhao, "Adapting sentiment lexicons to domain-specific social media texts," *Decis. Support Syst.*, vol. 94, pp. 65–76, Feb. 2017.

[29] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, 2007, pp. 440–447.

[30] F. Wu and Y. Huang, "Sentiment domain adaptation with multiple sources," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics (Long Papers)*, vol. 1, 2016, pp. 301–310.

[31] N. A. Abdullah, A. Feizollah, A. Sulaiman, and N. B. Anuar, "Challenges and recommended solutions in multi-source and multi-domain sentiment analysis," *IEEE Access*, vol. 7, pp. 144957–144971, 2019.

[32] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguistics*, vol. 37, no. 2, pp. 267–307, Jun. 2011.

[33] F. H. Khan, U. Qamar, and S. Bashir, "SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection," *Appl. Soft Comput.*, vol. 39, pp. 140–153, Feb. 2016.

[34] H. Saif, M. Fernandez, L. Kastler, and H. Alani, "Sentiment lexicon adaptation with context and semantics for the social Web," *Semantic Web*, vol. 8, no. 5, pp. 643–665, Apr. 2017.

[35] C. S. Khoo and S. B. Johnkhan, "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons," *J. Inf. Sci.*, vol. 44, no. 4, pp. 491–511, Aug. 2018.

[36] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese text sentiment analysis based on extended sentiment dictionary," *IEEE Access*, vol. 7, pp. 43749–43762, 2019.

[37] F. Wu, Y. Huang, and J. Yan, "Active sentiment domain adaptation," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics (Long Papers)*, vol. 1, 2017, pp. 1701–1711.

[38] A. Dey, M. Jenamani, and J. J. Thakkar, "Senti-N-Gram: An n -Gram lexicon for sentiment analysis," *Expert Syst. Appl.*, vol. 103, pp. 92–105, Aug. 2018.

[39] F. Z. Xing, F. Pallucchini, and E. Cambria, "Cognitive-inspired domain adaptation of sentiment lexicons," *Inf. Process. Manage.*, vol. 56, no. 3, pp. 554–564, May 2019.

[40] M. Fares, A. Moufarrej, E. Jreij, J. Tekli, and W. Grosky, "Unsupervised word-level affect analysis and propagation in a lexical knowledge graph," *Knowl.-Based Syst.*, vol. 165, pp. 432–459, 2019.

[41] J. Bernabé-Moreno, A. Tejeda-Lorente, J. Herce-Zelaya, C. Porcel, and E. Herrera-Viedma, "A context-aware embeddings supported method to extract a fuzzy sentiment polarity dictionary," *Knowl.-Based Syst.*, vol. 190, Feb. 2020, Art. no. 105236.

[42] B. Shin, T. Lee, and J. D. Choi, "Lexicon integrated cnn models with attention for sentiment analysis," 2016, *arXiv:1610.06272*. [Online]. Available: https://arxiv.org/abs/1610.06272

[43] Y. Wang, Y. Zhang, and B. Liu, "Sentiment lexicon expansion based on neural PU learning, double dictionary lookup, and polarity association," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 553–563.

[44] W. Li, D. Li, H. Yin, L. Zhang, Z. Zhu, and P. Liu, "Lexicon-enhanced attention network based on text representation for sentiment classification," *Appl. Sci.*, vol. 9, no. 18, p. 3717, Sep. 2019.

[45] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proc. 8th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 1997, pp. 174–181.

[46] Y. Xia, E. Cambria, A. Hussain, and H. Zhao, "Word polarity disambiguation using Bayesian model and opinion-level features," *Cognit. Comput.*, vol. 7, no. 3, pp. 369–380, Jun. 2015.

[47] O. Vechtomova, "Disambiguating context-dependent polarity of words: An information retrieval approach," *Inf. Process. Manage.*, vol. 53, no. 5, pp. 1062–1079, Sep. 2017.

[48] D. Deng, L. Jing, J. Yu, S. Sun, and M. K. Ng, "Sentiment lexicon construction with hierarchical supervision topic model," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 4, pp. 704–718, Apr. 2019.

[49] W. L. Hamilton, K. Clark, J. Leskovec, and D. Jurafsky, "Inducing domain-specific sentiment lexicons from unlabeled corpora," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, p. 595.

[50] M. Fernández-Gavilanes, T. Álvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, and F. Javier González-Castaño, "Unsupervised method for sentiment analysis in online texts," *Expert Syst. Appl.*, vol. 58, pp. 57–75, Oct. 2016.

[51] O. Araque, M. Guerini, C. Strapparava, and C. A. Iglesias, "Neural domain adaptation of sentiment lexicons," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos (ACIIW)*, Oct. 2017, pp. 105–110.

[52] R. Xia, C. Zong, X. Hu, and E. Cambria, "Feature ensemble plus sample selection: Domain adaptation for sentiment classification," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 10–18, May 2013.

[53] B. Gong, K. Grauman, and F. Sha, "Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation," in *Proc. 30th ICML*, 2013, pp. 222–230.

[54] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2066–2073.

[55] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaption for large-scale sentiment classification: A deep learning approach," in *Proc. Int. Conf. Mach. Learn., ICML*, 2011, pp. 1–8.

[56] Z. Li, Y. Zhang, Y. Wei, Y. Wu, and Q. Yang, "End-to-end adversarial memory network for cross-domain sentiment classification," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2237–2243.

[57] S. M. Salaken, A. Khosravi, T. Nguyen, and S. Nahavandi, "Seeded transfer learning for regression problems with deep learning," *Expert Syst. Appl.*, vol. 115, pp. 565–577, Jan. 2019.

[58] Y. Yoshida, T. Hirao, T. Iwata, M. Nagata, and Y. Matsumoto, "Transfer learning for multiple-domain sentiment analysis identifying domain dependent/independent word polarity," in *Proc. 21th AAAI Conf. Artif. Intell.*, 2011, pp. 1286–1291.

[59] V. Jha, S. R, P. D. Shenoy, V. K R, and A. K. Sangaiah, "A novel sentiment aware dictionary for multi-domain sentiment classification," *Comput. Electr. Eng.*, vol. 69, pp. 585–597, Jul. 2018.

[60] C. Zhao, S. Wang, and D. Li, "Multi-source domain adaptation with joint learning for cross-domain sentiment classification," *Knowl.-Based Syst.*, vol. 191, Mar. 2020, Art. no. 105254.

[61] X. Ding, Q. Shi, B. Cai, T. Liu, Y. Zhao, and Q. Ye, "Learning multi-domain adversarial neural networks for text classification," *IEEE Access*, vol. 7, pp. 40323–40332, 2019.

[62] H. Keshavarz and M. S. Abadeh, "ALGA: Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs," *Knowl.-Based Syst.*, vol. 122, pp. 1–16, Apr. 2017.

[63] Z. Yuan, S. Wu, F. Wu, J. Liu, and Y. Huang, "Domain attention model for multi-domain sentiment classification," *Knowl.-Based Syst.*, vol. 155, pp. 1–10, Sep. 2018.

[64] S. Sanagar and D. Gupta, "Adaptation of multi-domain corpus learned seeds and polarity lexicon for sentiment analysis," in *Proc. Int. Conf. Comput. Netw. Commun.*, Dec. 2015, pp. 60–68.

[65] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, p. 9, Dec. 2016.

[66] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol. NAACL*, 2003, pp. 173–180.

[67] C. D. Manning, P. Raghavan, H. Schutze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[68] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics," *J. Artif. Intell. Res.*, vol. 37, pp. 141–188, Feb. 2010.

[69] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

[70] W. Weaver, "Translation," *Mach. Transl. Lang.*, vol. 14, nos. 15–23, pp. 1–10, 1955.

[71] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Amer. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, 1990.

[72] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge.," *Psychol. Rev.*, vol. 104, no. 2, pp. 211–240, 1997.

[73] P. Liang and S. Wongthanavasu, "Hybrid linear matrix factorization for topic-coherent terms clustering," *Expert Syst. Appl.*, vol. 62, pp. 358–372, Nov. 2016.

[74] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: Understanding rating dimensions with review text," in *Proc. 7th ACM Conf. Recommender Syst. RecSys*, 2013, pp. 165–172.

[75] R. He and J. McAuley, "Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering," in *Proc. 25th Int. Conf. World Wide Web WWW*, 2016, pp. 507–517.

[76] S. M. Mohammad and P. D. Turney, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon," in *Proc. NAACL-HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 26–34.

[77] K. W. Church and P. Hanks, "Word association norms, mutual information, and lexicography," *Comput. Linguistics*, vol. 16, no. 1, pp. 22–29, Mar. 1990.

[78] *NLTK3.09 Library*. Accessed: Dec. 2017. [Online]. Available: https://www.nltk.org/

[79] R. Rehurek and P. Sojka, "Software frame-work for topic modelling with large corpora," in *Proc. LREC Workshop New Challenges NLP Frameworks*, 2010, pp. 45–50.

SWATI SANAGAR received the bachelor's and master's degrees in computer science from Pune University, India. She is currently pursuing the Ph.D. degree with the Doctoral Program on Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham Bengaluru Campus, India. She also works as a Data Scientist. Her research interests include sentiment analysis, natural language processing, text analytics, machine learning, and neural networks.

DEEPA GUPTA was born in 1977. She received the Ph.D. degree in natural language processing (example-based machine translation) from the Department of Mathematics and Computer Application, IIT Delhi, in 2005. She worked as a Postdoctoral Researcher with FBKIRST (the Center for Scientific and Technological Research), Trento, Italy. She is currently an Associate Professor with the Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, India. Her research work has published in journals, such as *Information Processing and Management*, *Expert Systems with Applications*, the IEEE Access, and the *International Journal of Speech Technology*. She has been guiding the Ph.D. and graduate students, since 2009. She has given invited talks on machine learning and natural language processing in Government funded workshops. She has completed two government funded projects related to text plagiarism detection and speech recognition system for Kannada language, in last five years. She is also involved in consultancy projects with industry. Her research interests are in sentiment analysis, clinical data mining, speech processing, and other areas in natural language processing.