

SKEDS – An External Knowledge Supported Logistic Regression Approach for Document-Level Sentiment Classification

Nesar Ahmad Wasi^a, Muhammad Abulaish^{a,*}

^aDepartment of Computer Science, South Asian University, New Delhi, India

Abstract

Due to the enormous amount of user-generated content being generated on the web, labeling such data is a time-consuming and expensive endeavor. As a result, we have limited annotated data and the vast majority of data are unlabeled. Analysis reveals that extracting (external) knowledge from unlabeled data and integrating it with knowledge extracted from labeled data is a beneficial task for text information processing, in particular text classification. In this paper, we present a hybrid approach for classifying sentiments that employs external knowledge, which is categorized as either general-purpose sentiment knowledge or domain-related knowledge. General-purpose sentiment knowledge is extracted from sentiment lexicons, whereas domain-related knowledge is extracted from unlabeled data from the same or related domains. Similar domains for a given domain are identified based on their similarity score in terms of overlapping features. The proposed approach utilizes both forms of external knowledge and combines them with logistic regression to train an improved classification model. The classification model uses the conventional gradient descent algorithm for optimization, and its convergence analysis indicates that it is convex and converges to the global optimum. The proposed approach is empirically evaluated and compared to three baselines and one state-of-the-art method using standard performance evaluation metrics on a multi-domain sentiment dataset. The experiment results are encouraging, demonstrating that the proposed approach considerably outperforms the baseline approaches and outperforms the state-of-the-art approach by up to 2% in terms of both *f-score* and *accuracy*.

Keywords: Text Mining, Machine Learning, Logistic Regression, Sentiment Analysis, Knowledge-based System

1. Introduction

The evolution of the Web from Web1.0 to Web3.0 resulted in the generation of massive amounts of heterogeneous data. The extraction of information and knowledge from user-generated data is critical for the concerned enterprises. In the age of online social media, large corporations are more concerned with meeting customer expectations. Furthermore, businesses are more interested in learning what their customers think of their products. The user-generated data, on the other hand, is massive and unlabeled. Although working with labeled training data is generally straightforward, labeling data is a time-consuming and resource-intensive task, and it is difficult to extract valuable knowledge from unlabeled data.

Sentiment analysis is used to determine how customers or users feel about an organization's or service provider's products and services. Document-level sentiment classification refers to the task of classifying the polarity of a document. In general, sentiment classification approaches are classified into three types: lexicon-based approaches, machine learning-based approaches, and hybrid approaches

(Sazzed and Jayarathna, 2021). Lexicon-based approaches (Sazzed, 2020; Abulaish et al., 2020, 2009) are based on general-purpose sentiment lexicons. They use aggregated sentiment scores extracted from lexicons to classify the polarity of a document. In machine learning-based approaches (Tripathy et al., 2016; Li et al., 2021, 2018), a machine learning model is trained on a training dataset to predict the polarity of a document. Finally, hybrid approaches (Wasi and Abulaish, 2020; Malandrakis et al., 2013) incorporate knowledge extracted from lexicons into machine learning algorithms to classify a document's polarity.

In this study, we introduce a hybrid method that takes advantage of both traditional machine learning and external knowledge. We incorporate two types of external knowledge in the proposed method: *general-purpose sentiment knowledge* and *domain-related knowledge*. Domain-related knowledge is retrieved from unlabeled data of the present domain and other domains with similar characteristics. We compare the feature space of the current domain and other domains under consideration, and related domains are those that have a greater number of features with the current domain. To extract domain-related information, we employ the methodology outlined by Hatzivassiloglou and McKeown (1997). In Hatzivassiloglou and

*Corresponding author

Email addresses: nesarahmadwasi17@gmail.com (Nesar Ahmad Wasi), abulaish@sau.ac.in (Muhammad Abulaish)

McKeown (1997), two words connected by coordinating conjunctions are described as having the same semantic relationship. However, if two words are united by an adversarial conjunction or two words are joined by a coordinating conjunction and a negation word appears before either of the two terms, they are said to have an opposite semantic relationship. In this paper, we use external knowledge with logistic regression for developing an accurate classifier. The proposed method utilizes gradient descent algorithm for optimization. A preliminary version of this work has been published in a conference proceedings Wasi and Abulaish (2020). However, the work proposed in this paper is a considerably enhanced version, and it can be summarized as follows: (i) introducing a hybrid method that takes advantage of both traditional machine learning and external knowledge, (ii) an amalgamation of both general-purpose sentiment knowledge and domain-related knowledge for sentiment classification, (iii) a detailed convergence analysis of the proposed optimization model, and (v) an extensive empirical evaluation of the proposed approach and its comparative analysis with three baselines and an advanced sentiment analysis method on a multi-domain sentiment dataset.

The remainder of the paper is structured as follows. The section 2 provides a thorough examination of the related works. The section 3 describes the external knowledge extraction process as well as the overall functioning of the proposed approach. It also describes the optimization algorithm and its analysis of convergence. The experimental setup and results are presented in section 4. Finally, section 5 concludes the paper with future research directions.

2. Related Works

There are three approaches to sentiment analysis: lexicon-based approaches, machine learning-based approaches, and hybrid approaches. A general-purpose sentiment lexicon is used in lexicon-based approaches to assign scores to sentiment terms found in a document. Furthermore, the scores are aggregated to classify the document as positive or negative. A machine learning algorithm is trained over a training dataset in machine learning-based approaches, and the trained machine learning model is used to classify the polarity of a document. A machine learning model is trained with the help of a lexicon to classify the polarity of a document in hybrid approaches. There are two ways to use lexicon with machine learning models in hybrid approaches: first, train two models and combine them. The second method, however, is to directly use lexicon knowledge in a machine learning model.

Ahmed et al. (2020) proposed a neural network-based hybrid approach for generating a domain-dependent sentiment dictionary, and they used it to address aspect-level sentiment analysis in attention-based Long Short-Term Memory (LSTM). Appel et al. (2016) proposed a sentence-level hybrid approach in which Bing Liu’s sentiment lexicon is the starting point to generate their sentiment lex-

icon. Appel et al. (2016) also used the concept of fuzzy sets to calculate the semantic polarity of sentences. Further, the results are compared with Naïve Bayes and Maximum Entropy. Andreevskaia and Bergler (2008) proposed a two-step approach. In first step, two separate parametric models are trained. In second step, both the trained models are combined to form a single classification model. In the approach presented by Wilson et al. (2005) first, those terms that carry sentiment polarity are determined. Then BoosTexter Adaboost classifier is used to classify the contextual polarity of the identified sentiment terms. Dang et al. (2010) presented a hybrid sentiment classification approach in which two types of features – content-free and content-dependent features, are integrated with features generated using lexicon and Parts-Of-speech (POS) tagger. In approach proposed by Li et al. (2009), a sentiment lexicon is used to extract domain-independent sentiment terms. However, unlabeled data is used to extract domain-dependent features.

Fang and Chen (2011) proposed a hybrid sentiment classification approach in which domain-specific lexicon is extracted by employing aspect-based classifier. Further, another sentiment classifier is used to classify polarity of sentiment words. Finally, the classification model is produced by aggregating the results of both classifiers. In approach proposed by Han et al. (2018), a domain-specific sentiment lexicon is generated using concept of mutual information and SentiWordNet-based sentiment classifier. Moreover, a lexicon-based sentiment analysis framework is proposed in which the generated domain-specific sentiment lexicon is employed. In (Wasi and Abulaish, 2020), a hybrid document-level sentiment classification approach is proposed in which two types of prior knowledge are incorporated with logistic regression – general-sentiment knowledge and domain-specific sentiment knowledge. General-sentiment knowledge is extracted from Bing Liu’s lexicon (Hu and Liu, 2004). Domain-specific sentiment knowledge is extracted from unlabeled data of the same domain. Further, gradient descent approach is used to optimize the proposed optimization model.

In (Ennajari et al., 2022), the authors proposed an external knowledge enhanced approach that incorporate semantic knowledge extracted from knowledge graph and word embedding to enhance the learning process. Further, the extracted knowledge is leveraged in low-dimensional representation. In (Li et al., 2021), adaptive gate network is proposed to use corpus-level statistical features in deep neural architecture. In particular the adaptive gate network is used to mitigate the problem of overfitting caused by incorporating statistical features. In (Liu et al., 2017), external knowledge extracted from language resources such as pre-trained word embedding, WordNet, and knowledge bases are leveraged in deep neural classification models to enhance the classification task. In (Li et al., 2018), a text-concept-vector framework is proposed, in which concepts for original text is generated using a knowledge base. Further, a neural network is utilized to transform produced

Notations	Description
L	General-purpose sentiment lexicon
\mathcal{F}_c	Feature space of current domain
$ \mathcal{F}_c $	Size of the feature space where each feature $f_i \in \mathbb{R}$
K	External knowledge where $K \in \mathbb{R}^{ \mathcal{F}_c \times 1}$
G	General-purpose sentiment knowledge where $G \in \mathbb{R}^{ \mathcal{F}_c \times 1}$
\mathcal{D}	Domain-related knowledge where $\mathcal{D} \in \mathbb{R}^{ \mathcal{F}_c \times \mathcal{F}_c }$
dv	Domain-related knowledge vector where $dv \in \mathbb{R}^{ \mathcal{F}_c \times 1}$
$F_{i,j}^s$	Frequency of features f_i and f_j shares the same-semantic orientation
$F_{i,j}^o$	Frequency of features f_i and f_j shares the opposite-semantic orientation
\mathcal{F}	Feature space of all domains
\mathcal{F}_r	Feature space of remaining domains
D	List of all domains
D_c	Current Domain
D_r	List of remaining domains
SiD	Similar domains
ψ	Classification model
n	Number of training samples
Ψ	Sigmoid function

Table 1: Notations and their descriptions

concepts into a vector representation. The produced vectors preserve the semantic and concept information embedded in the original text. Further, it is incorporated into various neural network-based approaches to enhance the prediction task.

Khan et al. (2023) proposed a hybrid deep neural network-based approach for sentiment classification that is able to leverage linguistic sentiment knowledge and pre-trained BERT-based word embedding in the attention enhanced Bi-LSTM. In (Li et al., 2023), the authors proposed sentiment-specific word representations by using an external hybrid sentiment knowledge. The external sentiment knowledge is composed of knowledge expectation and context weight. Zhong et al. (2023) proposed a knowledge graph augmented network for aspect-based sentiment analysis that utilizes syntactic and contextual information in the form of external knowledge. Du et al. (2023) proposed a hybrid aspect-based sentiment classification approach for the financial sector, in which multiple lexical knowledge sources are merged as external knowledge. Yin et al. (2023) proposed a multi-modal approach for sentiment analysis, in which multi-head visual attention is used to capture sentimental features with the help of visual features. In addition, a hybrid fusion network is proposed to incorporate knowledge learned from various modal representations by employing base classifiers to determine the final prediction.

3. Proposed Approach

In this section, we discuss the functional details of our proposed SKEDS method for external knowledge-supported document-level sentiment classification. The process of the suggested methodology is depicted in the figure 1. Beginning with fundamental preprocessing, SKEDS focuses primarily on the extraction of external knowledge and its incorporation into logistic regression for sentiment classification. The following subsections contain additional procedure-related information of SKEDS. The Table 1 provides a brief explanation of the notations utilized in the following sections.

3.1. External Knowledge Extraction

External knowledge is extracted from the current domain’s and related domains’ data. There are two sorts of external knowledge: general-purpose sentiment knowledge and domain-specific knowledge. The source of general-purpose sentiment knowledge is general-purpose sentiment lexicons, while the source of domain-specific sentiment knowledge is unlabeled data from the present domain and related domains. We combine general-purpose sentiment knowledge and domain-specific knowledge to create external knowledge, which is then included into logistic regression. The following subsections describe additional information regarding various knowledge extraction processes.

3.1.1. General-Purpose Sentiment Knowledge Extraction

To extract general-purpose sentiment knowledge, a general-purpose sentiment lexicon is used. SentiWord-

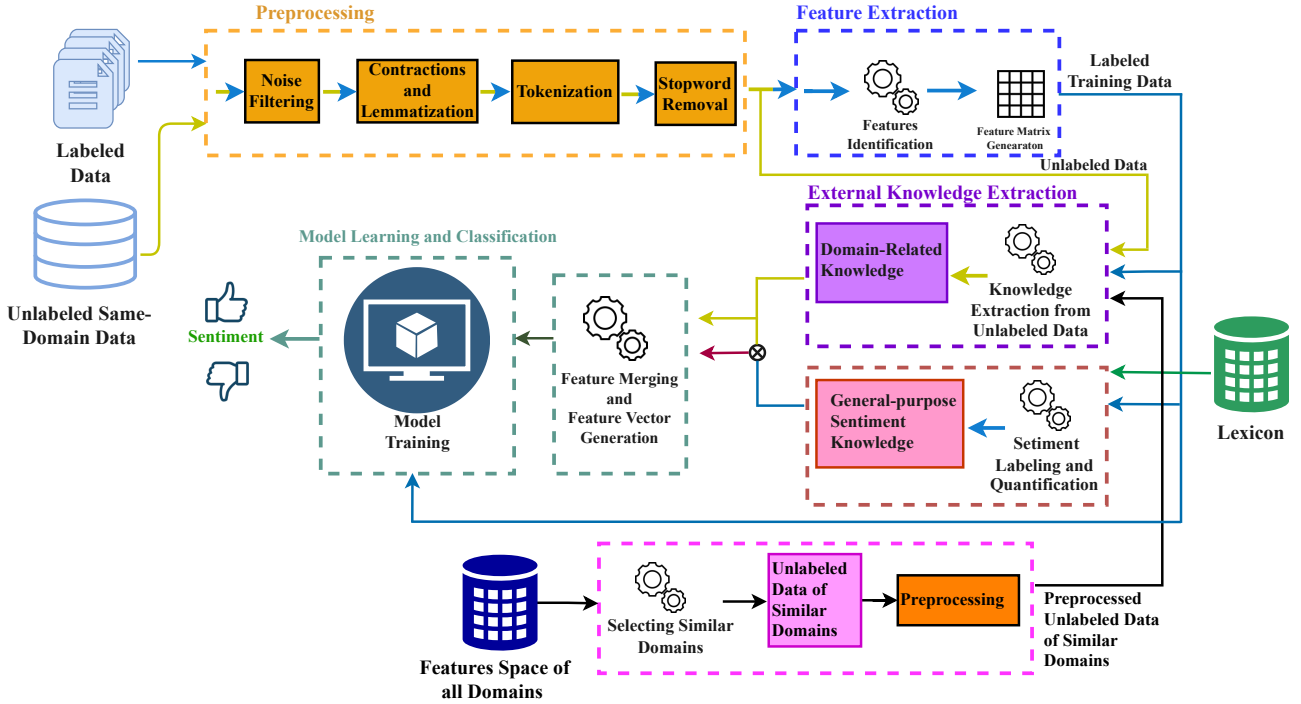


Figure 1: Functional details of the proposed SKEDS method.

Net (Sebastiani and Esuli, 2006), MPQA (Wilson et al., 2005), Bing Liu’s sentiment lexicon (Hu and Liu, 2004), and BiSAL (Rowaily et al., 2015) are only a few examples of general-purpose sentiment lexicons. Each feature f_i is compared to the general-purpose sentiment lexicon L in order to obtain general-purpose sentiment knowledge. The general-purpose sentiment knowledge g_i is assigned as 1 (or -1) if feature f_i is present and it is given a positive (or negative label) in the general-purpose sentiment lexicon; otherwise, it is assigned as 0. This process is presented formally in equation (1). It should be noted that as a result of the manual labeling of the general-purpose sentiment lexicon, the polarity of knowledge acquired from the general-purpose sentiment lexicon is more certain.

$$g_i = \begin{cases} 1, & \text{if } f_i \in L \text{ and labeled as positive in } L, \\ -1, & \text{if } f_i \in L \text{ and labeled as negative in } L, \\ 0, & \text{if } f_i \notin L \end{cases} \quad (1)$$

3.1.2. Domain-Related Knowledge Extraction

Unlabeled data is used to extract domain-related knowledge. Unlabeled data includes information from the current domain and related domains. To generate a list of similar domains SiD , we first extract the feature space \mathcal{F} of all domains D . Thereafter, the Jaccard similarity, as defined in equation (2), is used to find similarities between the current domain’s feature space \mathcal{F}_c and the feature space of the remaining domains \mathcal{F}_r .

$$JS_i(\mathcal{F}_c, \mathcal{F}_{r_i}) = \frac{|\mathcal{F}_c \cap \mathcal{F}_{r_i}|}{|\mathcal{F}_c \cup \mathcal{F}_{r_i}|} \quad (2)$$

In equation (2), JS is the list of similarity scores of the remaining domains \mathcal{F}_r with the current domain \mathcal{F}_c . The list generated by equation (2) is arranged and top 5 domains based on the similarity scores are selected similar domains.

Domain-related knowledge encompasses terms with $g_i = 0$ in general-purpose sentiment knowledge. The approach proposed by Hatzivassiloglou and McKeown (1997) is used to extract domain-related knowledge. In order to extract domain-related knowledge, two types of conjunctions are used: *adversarial* and *coordinating* conjunctions. Coordinating conjunctions are used to connect two words that have the same semantic meaning. An adversarial conjunction, on the other hand, is used to indicate the opposite or contrast between two words. Equation (3) formally defines the process of extracting domain-related knowledge.

$$\mathfrak{D}_{i,j} = \frac{F_{i,j}^{s,c} - F_{i,j}^{o,c}}{F_{i,j}^{s,c} + F_{i,j}^{o,c} + \alpha_0} + \frac{\sum_{k=1}^{|SiD|} F_{i,j}^{s,k} - \sum_{k=1}^{|SiD|} F_{i,j}^{o,k}}{\sum_{k=1}^{|SiD|} F_{i,j}^{s,k} + \sum_{k=1}^{|SiD|} F_{i,j}^{o,k} + \alpha_0} \quad (3)$$

In equation (3), domain-related knowledge is extracted using knowledge from the current domain and from similar domains SiD . In equation (3), $F_{i,j}^s$ represents the frequency of features f_i and f_j that share the same-semantic

orientation. $F_{i,j}^o$ represents the frequency of features f_i and f_j that share the opposite-semantic orientation. If features f_i and f_j have $F_{i,j}^s > F_{i,j}^o$, then it is more certain that both features share same orientation. In equation (3), $\alpha_0 > 0$, α_0 is a constant which is added to avoid division-by-zero problem.

3.2. External Knowledge Integration

Domain-related knowledge vector dv_i is computed by combining both general-purpose sentiment knowledge and domain-related knowledge using equation (4).

$$dv_i = \frac{\sum_{i \neq j} \mathcal{D}_{i,j} \cdot g_j}{\sum_{i \neq j} |\mathcal{D}_{i,j} \cdot g_j|} \quad (4)$$

Equation (4) makes sure that the range of dv vector remains in the range of $[-1, 1]$. In order to compute external knowledge, both general-purpose sentiment knowledge and domain-related knowledge vector are combined. If general-purpose sentiment knowledge $g_i \neq 0$ for feature f_i , then external knowledge is assigned by general-purpose sentiment knowledge as $K_i = g_i$; otherwise, external knowledge is assigned by domain-related knowledge vector as $K_i = dv_i$. Equation (5) defines the process of computing the final external knowledge formally.

$$K_i = \begin{cases} g_i & \text{if } g_i \neq 0, \\ dv_i & \text{otherwise,} \end{cases} \quad (5)$$

In equation (5), the priority is given to the general-purpose sentiment knowledge because the general-purpose sentiment lexicon is manually labeled; and accordingly, it is more certain that the features extracted from the lexicon carry the correct sentiment.

3.3. External Knowledge-Supported Logistic Regression

The goal is to incorporate external knowledge extracted from data from both current and similar domains into logistic regression to learn an accurate classifier. Equation (6) is the mathematical formulation of the proposed approach. In equation (6), n indicates the number of training samples. $\mathcal{L}(y^{(i)}, x^{(i)}; \psi)$ denotes the cost of inaccurate classification of sample $x^{(i)}$ to correct label $y^{(i)}$ using the classification model ψ . $\sum_{j=1}^{|\mathcal{F}_c|} \psi_j^2$ is L_2 -regularization (Zou and Hastie, 2005) term, which is added to increase the stability of the proposed approach. Using $K^T \psi$ term, external knowledge is incorporated into the proposed model. It ensures that the polarity of sentiment terms in the classification model ψ remains consistent with that in external knowledge. Domain-related knowledge extracted from unlabeled data of the current and similar domains is integrated using $\psi^T \mathcal{D} \psi$ term. In equation (6), α , β , and η are non-negative regularization constants for external knowledge, domain-related knowledge, and L_2 -regularization, respectively.

$$\arg \min_{\psi} \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y^{(i)}, x^{(i)}; \psi) + \frac{\lambda}{2n} \sum_{j=1}^{|\mathcal{F}_c|} \psi_j^2 - \alpha K^T \psi - \beta \psi^T \mathcal{D} \psi \quad (6)$$

The proposed approach's base model is logistic regression, which is a probabilistic classification approach that calculates the probability \hat{y} of correctly labeling an observation $x^{(i)}$. In the proposed approach, we use the binary classification variant of logistic regression. In a binary classification model, the prediction model's outcome is either *positive* or *negative*. The hypothesis function of logistic regression is equation (7), which is a non-linear sigmoid function. If the value of ψx_i is very large, the value of $\Psi(x; \psi)$ is close to 1; on the other hand, if the value of ψx_i is very small, the value of $\Psi(x; \psi)$ is close to 0. The sigmoid function produces a continuous value that is always between 0 and 1.

$$\Psi(x; \psi) = \frac{1}{1 + e^{-\psi x}} \quad (7)$$

If the probability obtained from the equation (7) is greater than or equal to 0.5, the prediction model ψ is assigned 1. Otherwise, ψ is assigned 0. Equation (8) is used to assign a computed probability to the associated predicted class.

$$\psi^{(i)} = \begin{cases} 1 & \text{if } p(y^{(i)} = 1 | x^{(i)}) \geq 0.5, \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

3.3.1. Optimization Algorithm

For the formulated model shown in equation (6), the optimal parameters ψ are determined using an optimization algorithm. Finding the parameters that allow the model to provide predicted labels that are the closest possible to the actual labels is known as finding the optimum parameters. While connecting sample $x^{(i)}$ and the actual label $y^{(i)}$, the loss function calculates how inaccurate the prediction model ψ is. The logistic loss function is the loss function associated to logistic regression. The logistic loss function, which is a monotonic function, is shown in equation (9). If a function f is totally non-increasing or non-decreasing, it is referred to as a monotonic function. Logistic loss is preferred over Mean Square Error (MSE) due to its convexity property (Shalev-Shwartz and Ben-David, 2014).

$$\mathcal{L}(\psi, X, Y) = \log \frac{1}{1 + e^{-\psi X}} \quad (9)$$

Equation (10) represents cost function for all data samples.

$$\begin{aligned} Cost(\psi) &= -\frac{1}{n} \sum_{i=1}^n (y^{(i)} \log(\Psi(x^{(i)}; \psi) \\ &+ (1 - y^{(i)}) \log(1 - \Psi(x^{(i)}; \psi))) \end{aligned} \quad (10)$$

$$\begin{aligned}
J(\psi) &= \arg \min_{\psi} \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y^{(i)}, x^{(i)}; \psi) \\
&+ \frac{\eta}{2n} \sum_{j=1}^{|\mathcal{F}_c|} \psi_j^2 - \alpha K^T \psi - \beta \psi^T \mathfrak{D} \psi
\end{aligned} \tag{11}$$

The cost function given in equation (11) is optimized to determine the ideal parameter vector ψ for the suggested model in equation (6). The equation (11) is optimized using the gradient descent method. A model’s optimum is discovered using gradient descent by differentiating the cost function. The suggested model uses a convex cost function. No matter where the search for the optimal begins, the convexity characteristic guarantees that it will eventually converge to the optimum. In the section that follows, convergence and convexity are covered in more detail. We differentiate cost, and the quantity of δ further discounts it in each iteration of the gradient descent search. The δ constant is the step-size by which the gradient descent approach moves in search space (Jurafsky and Martin, 2000). Equation (12) is used to update ψ .

$$\begin{aligned}
\psi_{j+1} &= \psi_j - \delta \frac{d}{d\psi} \Psi(x; \psi) + \frac{\eta}{2n} \frac{d}{d\psi} \sum_{j=1}^{|\mathcal{F}_c|} \psi_j^2 \\
&- \alpha \frac{d}{d\psi} K^T \psi - \beta \frac{d}{d\psi} \psi^T \mathfrak{D} \psi
\end{aligned} \tag{12}$$

If the log loss function is substituted in equation (12), then equation (13) is obtained.

$$\begin{aligned}
\frac{\partial}{\partial \psi} J(\psi) &= -\frac{1}{n} \sum_{i=1}^n \left(y^{(i)} \frac{\partial}{\partial \psi} \log(\Psi(x^{(i)}, \psi)) \right. \\
&+ \left. (1 - y^{(i)}) \frac{\partial}{\partial \psi} \log(1 - \Psi(x^{(i)}, \psi)) \right) \\
&+ \frac{\partial}{\partial \psi} \frac{\eta}{2n} \sum_{j=1}^{|\mathcal{F}_c|} \psi_j^2 - \frac{\partial}{\partial \psi} \alpha K^T \psi \\
&- \frac{\partial}{\partial \psi} \beta \psi^T \mathfrak{D} \psi
\end{aligned} \tag{13}$$

After performing several simplification steps, equation (14) is obtained. In each iteration of gradient descent, equation (14) is used to update ψ .

$$\begin{aligned}
\psi_{j+1} &= \psi_j - \delta \frac{1}{n} \sum_{i=1}^n (\Psi(x^{(i)}, \psi) - y^{(i)}) x_j^{(i)} \\
&+ \frac{\eta}{n} \sum_{j=1}^{|\mathcal{F}_c|} \psi_j - \alpha K^T - \beta \psi^T \mathfrak{D}
\end{aligned} \tag{14}$$

3.3.2. Convergence Analysis

The proposed approach can be non-convex due to the $-\beta \psi^T \mathfrak{D} \psi$ term. Based on Wu et al. (2016), the model presented in equation (6) can be guaranteed to be convex if appropriate values are selected for η and β . If η is strictly greater than $\beta \lambda_{\mathfrak{D}_{max}}$, i.e.,

$$\eta > \beta \lambda_{\mathfrak{D}_{max}} \tag{15}$$

In equation (15), $\lambda_{\mathfrak{D}_{max}}$ is the maximum eigenvalue of domain-related knowledge S , η is non-negative coefficient of $L_2 - regularization$, and β is non-negative coefficient of domain-related knowledge \mathfrak{D} .

The logistic loss function is differentiable and convex. A convex function is a combination of two or more convex functions defined over a convex set, so the model presented in equation (6) is convex (Chandra et al., 2013). Gradient descent converges to a global optimum due to the convexity property. According to Nesterov (2014) and Gordon and Tibshirani (2012), if a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and differentiable, then the differential of function f is Lipschitz continuous with a constant L greater than zero. Lipschitz gradient means that the gradient of function f does not change quickly. The equation (16) is satisfied by a gradient descent algorithm with a fixed step size $t \leq 1/L$. It means that gradient descent has a fixed progress bound.

$$f(\psi^{(k)}) - f^* \leq \frac{\|\psi^{(0)} - \psi^*\|_2^2}{2tk} \tag{16}$$

Gradient descent finds the ϵ -suboptimal, i.e., $f(\psi^{(k)}) - f^* \leq \epsilon$ point in $O(1/\epsilon)$ iterations. If the logistic loss function is accompanied with $L_2 - regularization$, the loss function becomes strongly convex, resulting in a unique and optimal solution (Shalev-Shwartz and Ben-David, 2014). If the function f is strongly convex, it means that the regularized function, i.e., $f(\psi) - \frac{m}{2} \|\psi\|_2^2$, is convex for $m > 0$, where m is the smallest eigenvalue. Under the assumption of Lipschitz gradient and strong convexity for function f , gradient descent with fixed step-size $t \leq 2/(m + L)$ satisfies equation (17), where $0 < \gamma < 1$.

$$f(\psi^{(k)}) - f^* \leq \gamma^k \frac{L}{2} \|\psi^{(0)} - \psi^*\|_2^2 \tag{17}$$

According to Gordon and Tibshirani (2012), the convergence rate of a strongly convex function is greater than $\frac{1}{\epsilon}$, and the rate of convergence under strong convexity is of order $O(\gamma)$, which is exponentially fast. In other words, it finds the ϵ -suboptimal point in $O(\log(\frac{1}{\epsilon}))$ iterations.

4. Experimental Setup and Results

The experimental setup for the proposed approach is presented in this section. We begin by discussing the multi-domain sentiment dataset and evaluation metrics. The performance of the proposed approach is then evaluated by comparing it to three baselines and one state-of-the-art method.

4.1. Dataset and Evaluation Metrics

We ran numerous experiments on a multi-domain sentiment dataset, which was compiled by Blitzer et al. (2007). It contains data from 25 different domains. There are two types of data in each domain of the dataset: labeled data and unlabeled data. Table 2 displays the dataset’s detailed statistics. Each sample in the dataset includes information such as *product name*, *product type*, *helpful*, *rating*, *title*, *date*, *reviewer name*, *reviewer location*, and *review text*.

Domain	# Positive instances	# Negative instances	# Unlabeled instances	# Total instances
Apparel	1000	1000	7252	9252
Automotive	584	152	0	736
Baby	1000	900	2356	4256
Beauty	1000	493	1391	2884
Book	1000	1000	973194	975194
Camera and photo	1000	999	5409	7408
Cell phones and service	639	384	0	1023
Computer and video games	1000	458	1313	2771
DVD	1000	1000	122438	124438
Electronics	1000	1000	21009	23009
Gourmet food	1000	208	367	1575
Grocery	1000	352	1280	2632
Health and personal care	1000	1000	5225	7225
Jewelry and watches	1000	292	689	1981
Kitchen and housewares	1000	1000	17856	19856
Magazines	1000	970	2221	4191
Music	1000	1000	172180	174180
Musical instruments	284	48	0	332
Office products	367	64	0	431
Outdoor living	1000	327	272	1599
Software	1000	915	475	2390
Sports and outdoors	1000	1000	3728	5728
Tools and hardware	98	14	0	112
Toys and games	1000	1000	11147	13147
Video	1000	1000	34180	36180

Table 2: Statistics of the dataset.

Because we perform sentiment classification for a given review, we only used two of the above features in our experiments: *rating* and *review text*. The *rating* feature is used as the class label and converted to the binary class label. A positive class label is assigned to those reviews with ratings greater than 3. A negative class label is assigned to those reviews with a rating of less than 3, and the rest of the reviews with a rating of 3 are labeled as neutral reviews. Due to the nature of the binary classification problem, the neutral reviews are discarded.

In order to evaluate the performance of the proposed approach and baselines, we have used standard data mining evaluation metrics that are formally defined in equations (18), (19), (20), and (21). In equations (18) and (19), TP indicates the number of positive instances that are correctly predicted as positive, FP indicates the number of negative instances that are incorrectly predicted as positive, and FN indicates the number of positive instances that are incorrectly predicted as negative.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (18)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (19)$$

$$\text{F-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (20)$$

$$\text{Accuracy} = \frac{\text{\#correctly predicted samples}}{\text{\#total samples}} \quad (21)$$

4.2. Experimental Setup

All data samples undergo preprocessing, which includes noise filtering, contraction and lemmatization, tokenization, and stop-word removal. Following sample preprocessing, the mutual information classifier-based feature selection step is applied, and the top 5000 features are chosen as the feature space. A one-gram term-frequency feature matrix is created for labeled data after the selection of the most prominent features.

Experiments are conducted on those domains that contain unlabeled samples. The results of the *automotive*, *cell phones and service*, *musical instruments*, *office products*, and *tools and hardware* domains are not given since these domains lack unlabeled samples. According to Table 2, there are domains with an unbalanced ratio of positive to negative samples. In order to address the issue of imbalanced datasets, samples are randomly selected from the dominant class. The ratio of training to testing data samples was maintained at 80:20, and hold-out strategy was utilized to evaluate the efficacy of the proposed method. The baselines' and suggested approach's hyper-parameters are manually tweaked.

The Bing Liu lexicon (Hu and Liu, 2004) is used to extract general-purpose sentiment knowledge. This lexicon has been collected over the years and contains over 6800 terms with positive and negative connotations. Since some terms are absent from the general-purpose sentiment lexicon, we extracted domain-specific knowledge from unlabeled data in the present domain and related domains to identify the polarity of the remaining words.

As previously explained, similar domains are selected

Domain	Lex	Lex+SDK	LR	SPDKE	SKEDS
Apparel	73.95	73.38	73.76	76.58	80.83
Baby	73.49	73.79	75.30	76.99	78.00
Beauty	80.52	81.31	77.52	82.39	83.39
Book	69.72	68.11	54.55	73.36	73.91
Camera and photo	73.47	73.47	74.58	77.81	84.53
Computer and video games	77.32	79.69	75.31	82.35	83.14
DVD	74.27	74.75	68.70	78.19	78.63
Electronics	73.36	73.14	63.95	79.00	80.66
Gourmet food	81.08	81.63	76.92	84.51	85.71
Grocery	78.18	78.34	78.48	83.44	84.88
Health and personal care	72.59	73.33	72.30	79.82	80.39
Jewelry and watches	82.24	83.41	83.24	86.34	87.83
Kitchen and housewares	72.80	72.69	72.99	77.05	78.34
Magazines	69.46	69.60	71.75	74.23	75.08
Music	69.76	68.41	74.32	77.75	78.74
Outdoor living	74.00	73.74	76.3	81.44	82.35
Software	76.46	77.08	71.69	80.11	84.68
Sports and outdoors	74.49	74.69	75.13	80.00	81.01
Toys and games	75.45	75.60	68.11	82.51	83.51
Video	70.71	71.46	75.00	80.00	80.85

Table 3: F-score of the proposed method and all baselines across all domains of the multi-domain sentiment dataset.

Domain	Lex	Lex+SDK	LR	SPDKE	SKEDS
Apparel	66.00	65.00	72.25	74.00	79.25
Baby	65.26	65.79	67.37	71.84	76.84
Beauty	72.60	73.97	68.49	77.17	77.63
Book	62.00	55.75	66.25	71.50	73.00
Camera and photo	64.25	64.25	77.50	79.75	83.25
Computer and video games	69.80	73.76	70.79	77.72	78.71
DVD	69.00	68.25	69.25	74.75	75.00
Electronics	65.50	64.75	71.25	79.00	79.50
Gourmet food	72.55	73.53	70.59	78.43	81.37
Grocery	70.19	70.81	78.88	83.23	83.85
Health and personal care	64.50	66.00	70.50	77.50	79.75
Jewelry and watches	72.66	74.82	79.14	82.01	83.45
Kitchen and housewares	65.25	65.25	72.25	72.75	76.50
Magazines	61.17	61.42	74.62	78.68	79.95
Music	59.25	54.75	71.50	77.25	78.00
Outdoor living	66.67	66.67	73.72	80.13	80.77
Software	69.45	71.28	75.46	81.46	84.60
Sports and outdoors	68.50	69.00	75.50	80.50	81.25
Toys and games	69.25	69.50	74.25	84.00	84.50
Video	63.75	64.25	76.5	81.50	82.00

Table 4: Accuracy of the proposed approach and all baselines across all domains of the multi-domain sentiment dataset.

based on the degree of similar features between the current domain and the remaining domains. To identify related domains, the unigram feature space of all domains is extracted, and the Jaccard similarity metric is used to compute the similarity score between the current domain and the remaining domains. Top similar domains are those that are most similar to the current domain, i.e., those

that share the most features with the current domain. For the aim of retrieving domain-related knowledge, we regarded top 5 domains to be similar domains. Table 5 displays the five most similar domains for each domain in the multi-domain sentiment dataset. It also presents the proportion of shared features between domains. For example, the *health* domain shares 40.96%, 38.54%, 38.10%,

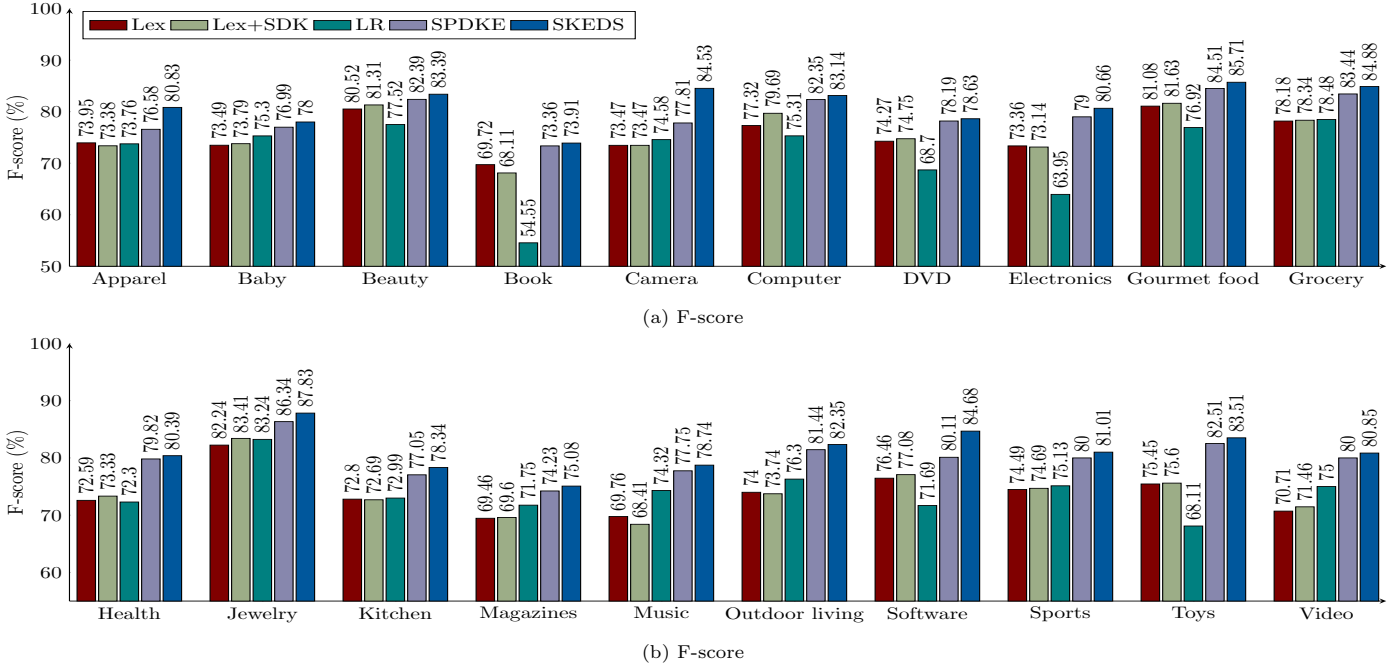


Figure 2: Visualization of f-score of the proposed approach and baseline methods for each domain of the multi-domain sentiment dataset.

Domain	Top 5 similar domains with percentage of shared features				
Apparel	(Baby, 44.48%)	(Beauty, 42.92%)	(Sports, 41.9%)	(Outdoor, 40.18%)	(Camera, 39.00%)
Baby	(Apparel, 44.48%)	(Sports, 41.24%)	(Beauty, 40.3%)	(Kitchen, 39.76%)	(Toys, 39.24%)
Beauty	(Apparel, 44.81%)	(Health, 42.76%)	(Baby, 42.07%)	(Grocery, 40.21%)	(Kitchen, 39.59%)
Book	(DVD, 44.54%)	(Video, 43.52%)	(Magazines, 40.12%)	(Computer, 37.08%)	(Music, 36.88%)
Camera	(Apparel, 39.00%)	(Electronics, 38.6%)	(Sports, 37.64%)	(Baby, 37.28%)	(Software, 36.74%)
Computer	(Book, 37.08%)	(DVD, 36.68%)	(Software, 36.22%)	(Apparel, 36.16%)	(Toys, 35.90%)
DVD	(Video, 47.78%)	(Book, 44.54%)	(Music, 39.18%)	(Magazines, 37.9%)	(Computer, 36.68%)
Electronics	(Camera, 38.6%)	(Software, 38.54%)	(Apparel, 37.44%)	(Sports, 35.88%)	(Baby, 35.4%)
Gourmet food	(Grocery, 56.72%)	(Beauty, 55.56%)	(Apparel, 50.68%)	(Outdoor, 46.57%)	(Kitchen, 46.44%)
Grocery	(Beauty, 52.87%)	(Gourmet food, 48.17%)	(Apparel, 48.04%)	(Baby, 45.48%)	(Kitchen, 45.37%)
Health	(Beauty, 40.96%)	(Baby, 38.54%)	(Apparel, 38.10%)	(Kitchen, 36.52%)	(Sports, 36.40%)
Jewelry	(Apparel, 62.11%)	(Beauty, 56.22%)	(Outdoor, 56.15%)	(Baby, 55.08%)	(Camera, 52.39%)
Kitchen	(Baby, 39.76%)	(Apparel, 38.46%)	(Beauty, 37.92%)	(Outdoor, 37.34%)	(Sports, 37.18%)
Magazines	(Book, 40.12%)	(DVD, 37.90%)	(Apparel, 36.80%)	(Video, 36.54%)	(Beauty, 34.50%)
Music	(DVD, 39.18%)	(Video, 38.26%)	(Book, 36.88%)	(Computer, 33.06%)	(Apparel, 32.96%)
Outdoor	(Apparel, 53.88%)	(Baby, 51.89%)	(Beauty, 50.23%)	(Kitchen, 50.07%)	(Sports, 49.56%)
Software	(Electronics, 38.54%)	(Apparel, 36.80%)	(Camera, 36.74%)	(Computer, 36.22%)	(Book, 35.98%)
Sports	(Apparel, 41.90%)	(Baby, 41.24%)	(Camera, 37.64%)	(Toys, 37.60%)	(Kitchen, 37.18%)
Toys	(Baby, 39.24%)	(Apparel, 38.80%)	(Sports, 37.60%)	(Computer, 35.90%)	(Beauty, 35.80%)
Video	(DVD, 47.78%)	(Book, 43.52%)	(Music, 38.26%)	(Magazines, 36.54%)	(Apparel, 35.54%)

Table 5: Top 5 similar domains with the percentage of shared features.

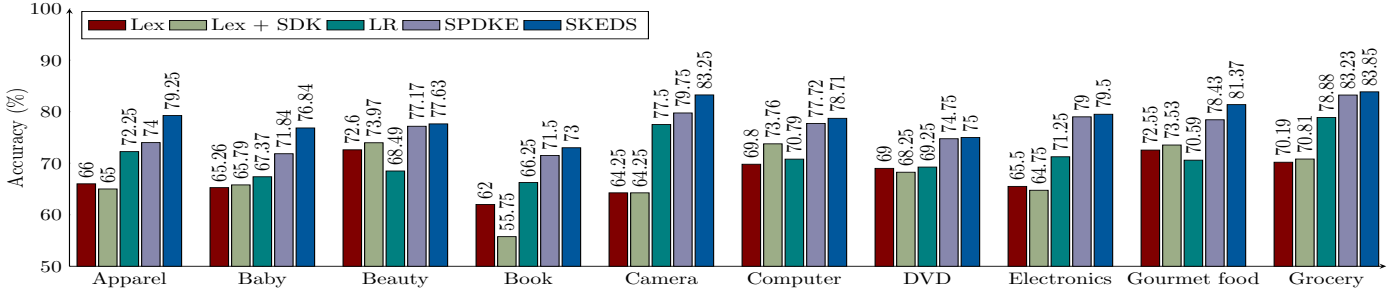
36.52%, and 36.40% of its features with the *beauty*, *baby*, *apparel*, *kitchen*, and *sports* domains, respectively.

During the training phase of the proposed method, sentiment knowledge and domain-specific knowledge are combined with training data. Using gradient descent, the parameter vector ψ of the proposed model is optimized. The δ learning rate is set at 0.01. If the difference between the cost functions of two successive iterations is less than 0.00001, gradient descent is terminated and considered to have converged to the optimal value.

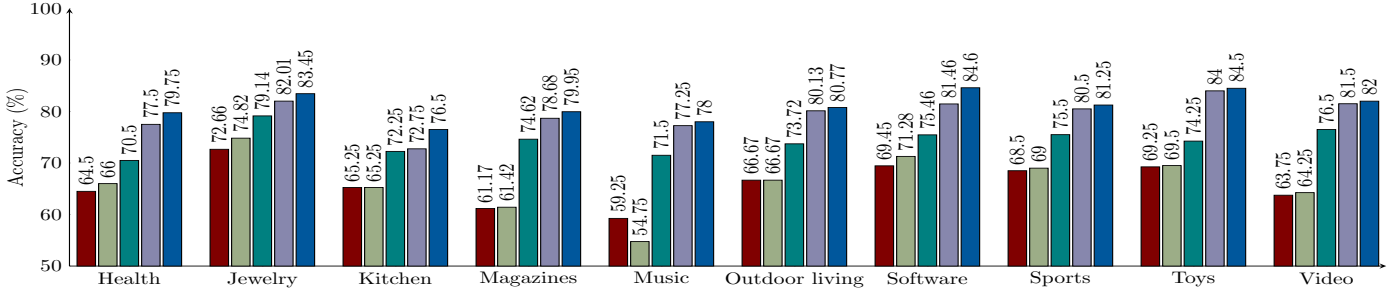
4.3. Comparative Analysis

In this section, we present a comparison of the proposed approach’s performance with baselines and a state-of-the-art method. We compare the proposed method against three baselines, namely a lexicon-based sentiment classifier, a combined classifier of the lexicon and single domain knowledge, and logistic regression, as well as a state-of-the-art method that employs a single domain prior knowledge-enhanced approach. The following paragraphs provide brief descriptions of all three baselines.

- *Lexicon-based sentiment classifier (Lex)*: In this base-

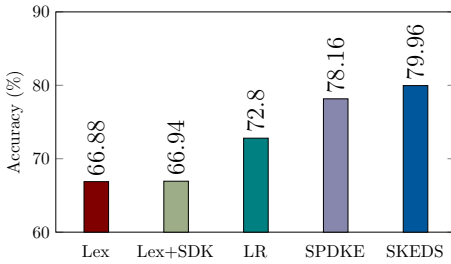


(a) Accuracy

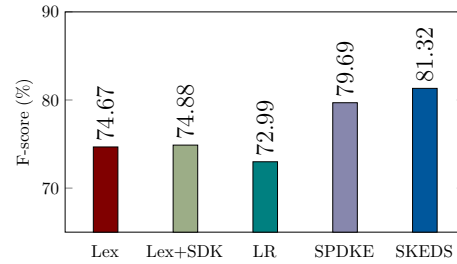


(b) Accuracy

Figure 3: Visualization of the accuracy of the proposed approach and baseline methods for each domain of the multi-domain sentiment dataset.



(a) Average accuracy over all domains



(b) Average f-score over all domains

Figure 4: Average accuracy and f-score of proposed approach and all baselines over all domains of multi-domain sentiment dataset.

line, we relied solely on the sentiment knowledge extracted from Bing Liu’s lexicon (Hu and Liu, 2004). This lexicon has been updated often for many years. It comprises approximately 6800 English sentiment terms, of which approximately 4800 are negative and approximately 2000 are positive.

- *A combined classifier of lexicon and single domain knowledge (Lex+SDK)*: In this baseline, we coupled the general-purpose sentiment knowledge with domain knowledge extracted from unlabeled data samples of a specific domain. In this technique, we do not utilize any optimization algorithm; instead, we combine the lexicon knowledge with the knowledge extracted from a particular domain.
- *Logistic regression (LR)*: This baseline is the foundational model for the proposed method. In this method, we optimize the parameter weights using gradient descent. However, prior knowledge is not incor-

porated into logistic regression.

- *Single-domain prior knowledge enhanced approach (SPDKE)* (Wasi and Abulaish, 2020): It is a modified logistic regression that can take into account prior sentiment knowledge. Prior knowledge is comprised of general-purpose sentiment knowledge and single-domain knowledge. Furthermore, gradient descent is employed to optimize the learning model.

The f-score and accuracy of the proposed and the baseline approaches are reported in Tables 3 and 4, respectively. In Table 3, it can be seen that the proposed approach has the lowest f-score, i.e., 73.91% in *Book* domain and the highest f-score, i.e., 87.83% in *Jewelry and watches* domain. From Table 4, it can be observed that the proposed approach has the lowest accuracy, i.e., 73% in *Book* domain and has the highest accuracy, i.e., 84.6% in *Software* domain. Figure 4 shows each approach’s average accuracy and f-score over all domains. It is apparent

from baseline *Lex+SDK* that simply blending knowledge extracted from unlabeled data of the current domain with the knowledge extracted from the lexicon is a worthwhile task. However, the improvement is minimal over *Lex* but encouraging. We may also note that knowledge extracted from unlabeled data needs to be incorporated appropriately. In Figures 2 and 3, we can observe that in domains such as *book* and *music*, knowledge extracted from unlabeled data in each of the two domains degrades the performance of the *Lex+SDK*. Because both domains have a significant amount of unlabeled data, the extracted domain knowledge must be incorporated appropriately. *SPDKE* combines *Lex+SDK* with *LR* and optimizes the extracted knowledge of a single domain using gradient descent. Figures 2 and 3 show that *SPDKE* appropriately incorporates the extracted knowledge in each of the two domains. From Figure 4, we can observe that *SKEDS*, which extracts external knowledge from unlabeled data of more than one domain, i.e., similar domains, outperforms *SPDKE* by around 2% in terms of both accuracy and f-score. From Figures 2 and 3, it is evident that *SKEDS* surpasses the baseline approaches in all domains in terms of both accuracy and f-score, which is in line with the hypothesis of utilizing unlabeled data from similar domains.

4.4. Ablation Study

In this section, we conduct the ablation study of the proposed approach to systematically validate the contributions and impacts of various core components within the proposed approach. In order to perform the ablation study, each core component is excluded from the model to check its associated effectiveness. In our approach, the core components are lexicon (*Lex*), single-domain knowledge (*Lex+SDK*), logistic regression (*LR*), single-domain prior knowledge enhanced approach (*SPDKE*), and finally *SKEDS*, which combines all previous core components with the knowledge from related domains. Each of the aforementioned core components is regarded as the baseline for the proposed method that is described in detail in section 4.3.

It can be observed from Figures 2 and 3 that merely using a lexicon for classification has the lowest performance score. However, incorporating single-domain knowledge into the lexicon can marginally boost its performance. In addition, when we combine single-domain knowledge and lexicon with logistic regression, performance is significantly enhanced. Finally, when we combine all core components together that constitutes the proposed approach, it outperforms all the baselines and state-of-the-art approach. The comparative analysis section and the results reported in Tables 3 and 4, as well as, Figures 2, 3, and 4 indicate that each core component contributes positively to the final model, *SKEDS*, enabling it to significantly outperform baseline approaches.

5. Conclusion and Future Work

In this paper, we have presented a hybrid approach for document-level sentiment classification that integrates external knowledge acquired from a sentiment lexicon and unlabeled domain-related datasets. This approach incorporates two forms of external knowledge: general-purpose sentiment knowledge and domain-specific knowledge. The general-purpose sentiment lexicon is used to extract general-purpose sentiment knowledge; whereas domain-related knowledge is retrieved from unlabeled data of the present domain and similar domains. The proposed classification model employs gradient descent for optimization. In addition, we have provided the convergence study of the proposed model, which indicates that our model converges to the optimal solution. The empirical results on a multi-domain sentiment dataset support the applicability of the proposed approach for document-level sentiment classification, as it outperforms multiple baselines in terms of both f-score and accuracy. On the basis of the empirical analysis, it can be concluded that including external sentiment knowledge acquired from related domains into a document classification model can greatly enhance its sentiment classification efficacy.

As an immediate future work, we can extract external knowledge not only from textual data but embedded in different modalities of data such as videos and images. To extend the proposed approach, we also look forward to exploring approaches that can utilize external knowledge in deep learning-based approaches. We are keen to see if it further improves the performance.

References

- Abulaish, M., Fazil, M., Anwar, T., 2020. A contextual semantic-based approach for domain-centric lexicon expansion, in: Proceedings of the Australasian Database Conference, Springer. pp. 216–224. doi:https://doi.org/10.1007/978-3-030-39469-1_18.
- Abulaish, M., Jahiruddin, Doja, M.N., Ahmad, T., 2009. Feature and opinion mining for customer review summarization, in: Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence, PReMI, Springer Berlin Heidelberg. pp. 219–224. doi:https://doi.org/10.1007/978-3-642-11164-8_35.
- Ahmed, M., Chen, Q., Li, Z., 2020. Constructing domain-dependent sentiment dictionary for sentiment analysis. *Neural Computing and Applications* 32, 14719–14732. doi:<https://doi.org/10.1007/s00521-020-04824-8>.
- Andreevskaia, A., Bergler, S., 2008. When specialists and generalists work together: Overcoming domain dependence in sentiment tagging, in: Proceedings of the 46th Annual Meeting of the ACL, ACL. pp. 290–298.

- Appel, O., Chiclana, F., Carter, J., Fujita, H., 2016. A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems* 108, 110–124. doi:<https://doi.org/10.1016/j.knosys.2016.05.040>.
- Blitzer, J., Dredze, M., Pereira, F., 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, in: *Proceedings of the 45th Annual Meeting of the ACL*, ACL. pp. 440–447.
- Chandra, S., Jayadeva, Mehra, A., 2013. *Numerical Optimization with Applications*. 1st ed., Narosa Publishing House.
- Dang, Y., Zhang, Y., Chen, H., 2010. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems* 25, 46–53. doi:<https://doi.org/10.1109/MIS.2009.105>.
- Du, K., Xing, F., Cambria, E., 2023. Incorporating multiple knowledge sources for targeted aspect-based financial sentiment analysis. *ACM Transactions on Management Information Systems* 14. doi:<https://doi.org/10.1145/3580480>.
- Ennajari, H., Bouguila, N., Bentahar, J., 2022. Knowledge-enhanced spherical representation learning for text classification, in: *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, SIAM. pp. 639–647. doi:<https://doi.org/10.1137/1.9781611977172.72>.
- Fang, J., Chen, B., 2011. Incorporating lexicon knowledge into SVM learning to improve sentiment classification, in: *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, Asian Federation of Natural Language Processing. pp. 94–100.
- Gordon, G., Tibshirani, R., 2012. Gradient descent revisited. *Optimization* 10, 725.
- Han, H., Zhang, J., Yang, J., Shen, Y., Zhang, Y., 2018. Generate domain-specific sentiment lexicon for review sentiment analysis. *Multimedia Tools and Applications* 77, 21265–21280. doi:<https://doi.org/10.1007/s11042-017-5529-5>.
- Hatzivassiloglou, V., McKeown, K.R., 1997. Predicting the semantic orientation of adjectives, in: *Proceedings of the 35th Annual Meeting of the ACL and 8th Conference of the European Chapter of the ACL*, ACL. pp. 174–181. doi:<https://doi.org/10.3115/976909.979640>.
- Hu, M., Liu, B., 2004. Mining and summarizing customer reviews, in: *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM. p. 168–177. doi:<https://doi.org/10.1145/1014052.1014073>.
- Jurafsky, D., Martin, J.H., 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3rd ed., Prentice Hall PTR.
- Khan, J., Ahmad, N., Khalid, S., Ali, F., Lee, Y., 2023. Sentiment and context-aware hybrid dnn with attention for text sentiment classification. *IEEE Access* 11, 28162–28179. doi:<https://doi.org/10.1109/ACCESS.2023.3259107>.
- Li, T., Zhang, Y., Sindhwani, V., 2009. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge, in: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, ACL. pp. 244–252. doi:<https://doi.org/10.5555/1687878.1687914>.
- Li, X., Li, Z., Xie, H., Li, Q., 2021. Merging statistical feature via adaptive gate for improved text classification, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 13288–13296. doi:<https://doi.org/10.1609/aaai.v35i15.17569>.
- Li, Y., Lin, Z., Lin, Y., Yin, J., Chang, L., 2023. Learning sentiment-enhanced word representations by fusing external hybrid sentiment knowledge. *Cognitive Computation* x, 1–15. doi:<https://doi.org/10.1007/s12559-023-10164-1>.
- Li, Y., Wei, B., Liu, Y., Yao, L., Chen, H., Yu, J., Zhu, W., 2018. Incorporating knowledge into neural network for text representation. *Expert Systems with Applications* 96, 103–114. doi:<https://doi.org/10.1016/j.eswa.2017.11.037>.
- Liu, M., Haffari, G., Buntine, W., Ananda-Rajah, M., 2017. Leveraging linguistic resources for improving neural text classification, in: *Proceedings of the Australasian Language Technology Association Workshop 2017*, pp. 34–42.
- Malandrakis, N., Kazemzadeh, A., Potamianos, A., Narayanan, S., 2013. SAIL: A hybrid approach to sentiment analysis, in: *Proceedings of the Second Joint Conference on Lexical and Computational Semantics and Seventh International Workshop on Semantic Evaluation*, ACL. pp. 438–442.
- Nesterov, Y., 2014. *Introductory Lectures on Convex Optimization: A Basic Course*. 1 ed., Springer Publishing Company, Incorporated. doi:<https://doi.org/10.1007/978-1-4419-8853-9>.

- Rowaily, K.A., Abulaish, M., Haldar, N.A.H., Rubaian, M.A., 2015. BiSAL – A bilingual sentiment analysis lexicon to analyze dark web forums for cyber security. *Digital Investigation* 14, 53–62. doi:<https://doi.org/10.1016/j.diin.2015.07.006>.
- Sazzed, S., 2020. Development of sentiment lexicon in bengali utilizing corpus and cross-lingual resources, in: *Proceedings of the 21st IEEE International Conference on Information Reuse and Integration for Data Science IRI*, IEEE. pp. 237–244. doi:<https://doi.org/10.1109/IRI49571.2020.00041>.
- Sazzed, S., Jayarathna, S., 2021. Ssentia: A self-supervised sentiment analyzer for classification from unlabeled data. *Machine Learning with Applications* 4, 100026. doi:<https://doi.org/10.1016/j.mlwa.2021.100026>.
- Sebastiani, F., Esuli, A., 2006. Sentiwordnet: A publicly available lexical resource for opinion mining, in: *Proceedings of the 5th International Conference on Language Resources and Evaluation*, European Language Resources Association (ELRA). pp. 417–422.
- Shalev-Shwartz, S., Ben-David, S., 2014. *Understanding Machine Learning: From Theory to Algorithms*. 1st ed., Cambridge University Press. doi:<https://doi.org/10.1145/2983323.2983851>.
- Tripathy, A., Agrawal, A., Rath, S.K., 2016. Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications* 57, 117–126. doi:<https://doi.org/10.1016/j.eswa.2016.03.028>.
- Wasi, N.A., Abulaish, M., 2020. Document-level sentiment analysis through incorporating prior domain knowledge into logistic regression, in: *Proceedings of the 19th IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, IEEE. pp. 969–974. doi:<https://doi.org/10.1109/WIIAT50758.2020.00148>.
- Wilson, T., Wiebe, J., Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment analysis, in: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, ACL. pp. 347–354. doi:<https://doi.org/10.3115/1220575.1220619>.
- Wu, F., Wu, S., Huang, Y., Huang, S., Qin, Y., 2016. Sentiment domain adaptation with multi-level contextual sentiment knowledge, in: *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, ACM. pp. 949–958. doi:<https://doi.org/10.1145/2983323.2983851>.
- Yin, C., Zhang, S., Zeng, Q., 2023. Hybrid representation and decision fusion towards visual-textual sentiment. *ACM Transactions on Intelligent Systems and Technology* 14. doi:<https://doi.org/10.1145/3583076>.
- Zhong, Q., Ding, L., Liu, J., Du, B., Jin, H., Tao, D., 2023. Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering* , 1–14doi:<https://doi.org/10.1109/TKDE.2023.3250499>.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of The Royal Statistical Society: Series B (Statistical Methodology)* 67, 301–320. doi:<https://doi.org/10.1111/j.1467-9868.2005.00503.x>.