IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 2, 2017, accepted February 15, 2017, date of publication March 31, 2017, date of current version August 29, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2690342

Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review

TAREQ AL-MOSLMI[1], NAZLIA OMAR[1], SALWANI ABDULLAH[1], AND MOHAMMED ALBARED[2]
[1]Faculty of Information Science and Technology, Center for Artificial Intelligence Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
[2]Faculty of Computer and Information Technology, Sana'a University, Sana'a, Yemen

Corresponding author: Tareq Al-Moslmi (almoslmi2010@gmail.com)

This work was supported in part by the Ministry of Higher Education, Malaysia, under Grant FRGS/1/2016/ICT02/UKM/02/11 and Grant FRGS/1/2015/ICT02/UKM/01/2, and in part by the Universiti Kebangsaan Malaysia under Grant DIP-2016-024.

ABSTRACT A sentiment analysis has received a lot of attention from researchers working in the fields of natural language processing and text mining. However, there is a lack of annotated data sets that can be used to train a model for all domains, which is hampering the accuracy of sentiment analysis. Many research studies have attempted to tackle this issue and to improve cross-domain sentiment classification. In this paper, we present the results of a comprehensive systematic literature review of the methods and techniques employed in a cross-domain sentiment analysis. We focus on studies published during the period of 2010–2016. From our analysis of those works, it is clear that there is no perfect solution. Hence, one of the aims of this review is to create a resource in the form of an overview of the techniques, methods, and approaches that have been used to attempt to solve the problem of cross-domain sentiment analysis in order to assist researchers in developing new and more accurate techniques in the future.

INDEX TERMS Cross-domain sentiment analysis, domain adaptation for sentiment analysis, multi-domain sentiment analysis, sentiment analysis, systematic literature review, transfer learning.

I. INTRODUCTION

Sentiment analysis is the computational study of people's attitudes, appraisals, and opinions about individuals, issues, entities, topics, events, and products as well as their attributes [1]–[9]. This type of analysis is very useful in a whole range of practical applications, but it is technically challenging. Due to the expansion in the types and uses of social media (e.g., blogs, forum discussions, reviews, and social networks), individuals can freely share their opinions in many domains [10]–[15]. However, capturing all of these opinions involves a huge computational cost because the training data for such an endeavor needs to be annotated for a large number of domains. This challenge prevents the exploitation of a plethora of information that is shared across domains.

Indeed, one of the main issues in cross-domain sentiment analysis is the lack of annotated data, which is crucial for accurate sentiment classification. Moreover, customer reviews, for example, can cover several types of services or products for which the usage of terminology varies, which further complicates the process of classification. Therefore, in recent years, research efforts have shifted toward developing cross-domain techniques to solve this issue. Given the multitude of approaches and techniques that now exist, this study aims to survey only the most recent and therefore focuses on works published during the period 2010–2016.

The remainder of this paper is organized as follows: The next section provides the basic terminology used in research on sentiment analysis. Then the following section presents the aims and criteria of this systematic review of the literature on cross-domain sentiment analysis. This is followed by a section that explains the data extraction and analysis process and a further section that discusses the dominant cross-domain sentiment analysis techniques. This is followed by a section that concludes the paper.

II. BASIC TERMINOLOGY

First, it would be useful to define the key terminology related to our research focus:

A. SENTIMENT ANALYSIS

The web has become the most important place for expressing opinions about products and services, as well as for commenting on social issues and government policies. The explosive growth in user-generated content, or "what people think?," is now an extremely important source of information.

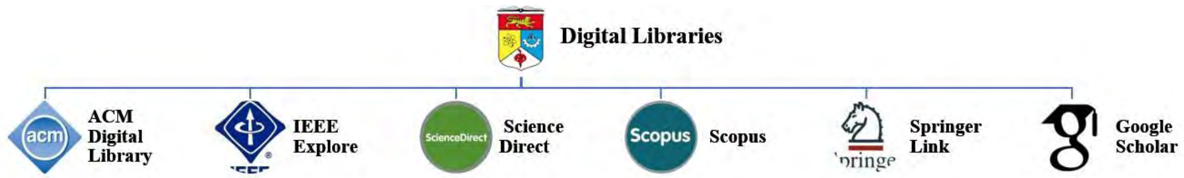

FIGURE 1. Digital libraries and databases searched.

For businesses, identifying and analyzing helpful reviews efficiently and accurately can help them to satisfy both current and potential customer needs, and the need to do so has become a critical challenge for market-driven product design.

B. DOMAIN

There are several definitions for domain, but in this context a domain is a class that contains different entities. For instance, products, such as electronics, software, and DVDs are considered different domains in the field of sentiment analysis.

C. DOMAIN ADAPTATION

Also known as cross-domain learning, domain adaptation uses annotated data from the source domain either alone or in conjunction with some from the target domain to learn a model in order to predict the unannotated occurrences in the target domain. The ease or complexity of domain adaptation depends on how closely related or otherwise the source and target domains are to each other. Therefore, cross-domain learning can influence the quality of the knowledge learned from the source domain and the understanding of the target domain.

III. SYSTEMATIC LITERATURE REVIEW

In order to survey the current state-of-the-art on and around the topic of cross-domain sentiment analysis, we undertook a systematic literature review following the procedures in [16]. Based on that work, we formulated the following research questions:

Question 1: For a given source–target domain, are the cross-domain approaches able to perform as well as they do in in-domain sentiment analysis?

Question 2: Are the cross-domain algorithms sensitive to variations in data representation?

Question 3: Does the cross-domain result depend on whether the source and target domains are heterogeneous or homogeneous?

A. SEARCH PROCESS

We started our survey by searching for relevant research studies on internet websites and in the online library of Universiti Kebangsaan Malaysia. The internet search was conducted by using search engines to trawl the digital libraries and databases illustrated in Figure 1.

The key terms or search strings that we used were "cross domain sentiment analysis," "cross domain opinion mining," "multi domain sentiment analysis," "cross domain sentiment classification," and "domain adaptation for sentiment analysis". In fact, these five key terms are mainly used to search for cross domain sentiment analysis articles. More articles also identified through scanning the reference list of each one of these articles

B. INCLUSION AND EXCLUSION CRITERIA

There is a lot of literature on sentiment analysis. Therefore, to ensure that the search would be manageable and focused we defined some inclusion and exclusion criteria to select the papers for review as follows:

1) INCLUSION CRITERIA

Studies published during the period 2010–2016 related to cross-domain sentiment analysis and studies on tools and techniques related to cross-domain sentiment analysis. If studies had been published in more than one journal or conference proceeding, we chose the most complete version for inclusion.

2) EXCLUSION CRITERIA

Informal studies (unknown conferences or journals); papers irrelevant to the above research questions.

TABLE 1. Number of research studies identified.

Digital library	Number of papers found		
	Based on key terms	Based on title	Based on abstract
ACM Digital Library	143	20	18
IEEE Xplore	92	15	11
Science Direct	28	7	5
Scopus	135	43	25
Springer Link	72	11	8
Google Scholar	712	68	46
Total	**1182**	**164**	**113**

C. BIBLIOGRAPHY MANAGEMENT AND DOCUMENT RETRIEVAL

We used Mendeley Desktop 1.12.1 to manage all the bibliographic details and citations. The studies that were identified by the above-described search process were scanned by title and abstract according to the inclusion and exclusion criteria. Then, all the papers that were identified as relevant to our research were then downloaded for data extraction and

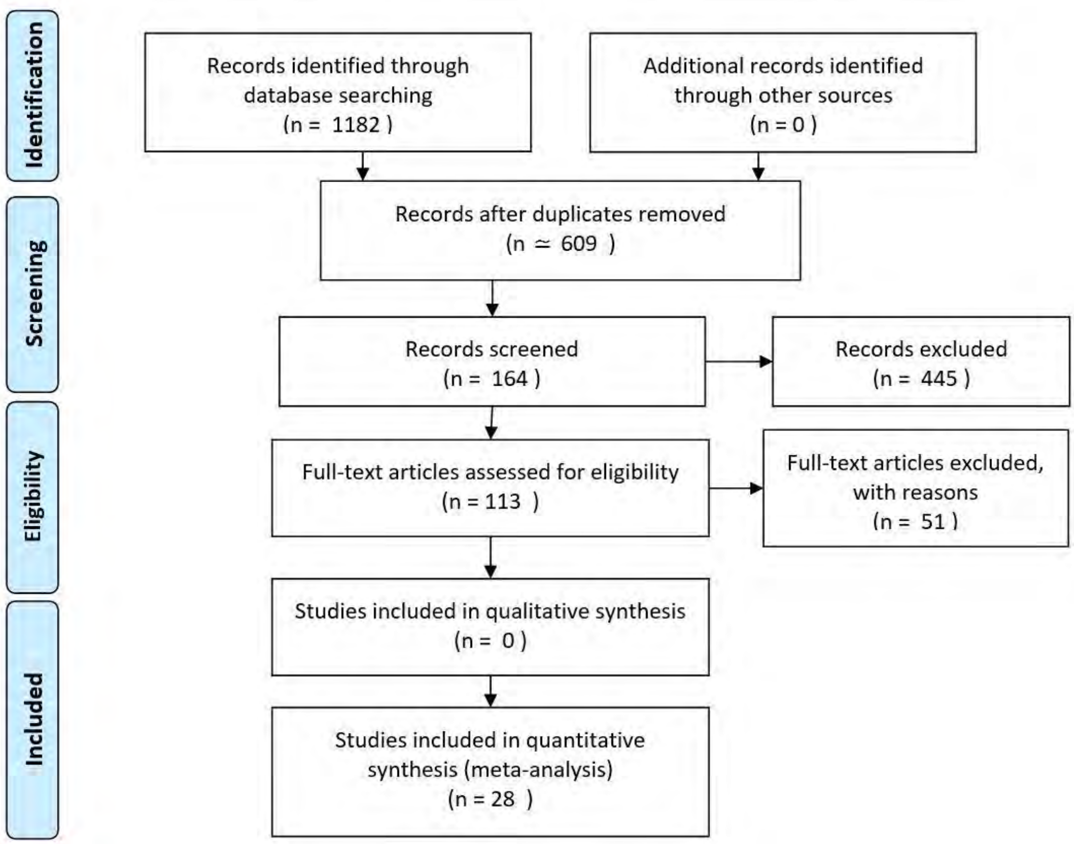

FIGURE 2. PRISMA flowchart adopted from [91].

further study. Table 1 provide details on the number of research studies that were discovered by the search of the digital libraries and databases in Figure 1 above and Figure 2 illustrates the search process in the form of a PRISMA flowchart.

IV. DATA EXTRACTION AND ANALYSIS

The 28 papers that were selected for more detailed study are summarized in Table 2. This table shows the main information extracted from the selected papers, which are presented in order of most recent year of publication.

A. CHALLENGES IN CROSS-DOMAIN SENTIMENT ANALYSIS

The key technical challenge in sentiment analysis is that it is highly domain-dependent. In other words, a method that performs well in one domain might underperform in another. This issue is especially challenging because machine learning for cross-domain classification performs well only with labeled documents and hence is highly domain-sensitive. The extent of the similarity between training data and test data determines the performance capability of machine learning techniques, implying that source–target domain pairs highly influence the results of cross-domain sentiment analysis. Complementing the sentiment scale with two or more sentiment classes results in a severe decrease in sentiment analysis performance. Although interclass differentiation is expected to be harder than binary classification and lead to a certain loss of performance, a sharp dip in performance is not acceptable. Moreover, a mismatch between review ratings and review text for ratings that are not fully binary also results in a drop in performance [17], [18]. Furthermore, a sentiment classifier may perform below par if it has been trained on labeled data from a source domain that has a rich corpora but is later used to classify sentiments from a target domain with a poor corpora. This inconsistent performance tends to occur because some words in the rich corpora (source) might not be used at all in the poor corpora (target), or they might be used in a different context and thus have a different meaning. Clearly, cross-domain sentiment analysis poses many challenges. The main ones in respect of natural language processing (NLP) are shown in Figure 3 below and can be summarized as follows:

1) SPARSITY

The sparsity problem arises when the target domain contains words or phrases that do not appear or rarely appear in the source domain.

TABLE 2. Summary of methodology and findings of selected research studies (n = 28).

Publication	Methodology / Finding	Key
Cross-Domain Sentiment Classification Using Sentiment Sensitive Embeddings [21]	An embedding technique was developed for the training phase of cross-domain sentiment classification, which considered the following objective functions together and in isolation: (a) distributional properties of pivots, (b) label constraints in source domain documents, and (c) geometric properties in unlabeled source and target domain documents.	K1
	The results showed that better performance can be achieved by optimizing all three of the above-mentioned objective functions than by optimizing each one individually. This confirms the importance of using task-specific embedding learning for cross-domain sentiment classification. As regards the best performance of an individual objective function, this is achieved by objective function (c)..	
Leveraging Latent Sentiment Constraint in Probabilistic Matrix Factorization for Cross-domain Sentiment Classification [22]	A latent sentiment factorization (LSF) algorithm based on probabilistic matrix factorization was developed to address the gap between the source and target domains by exploiting sentiment correlations between domain-shared and domain-specific words in a two-dimensional sentiment space. The proposed algorithm is designed for use where there are only labeled data in the source domain and unlabeled data in the target domain.	K2
	In experiments, LSF, SCL, SFA and TCT were applied to Amazon datasets and it was found that LSF performs better than SCL and SFA, and also achieves a level of accuracy that is comparable to that of TCT for cross-domain sentiment classification.	
Cross-Domain Sentiment Classification with Word Embeddings and Canonical Correlation Analysis [23]	A feature learning and feature subspace mapping method was proposed that uses. Word embeddings and canonical correlation analysis (CCA) to address vocabulary mismatches that can occur between source and target domains.	K3
	Experimental results revealed that while the proposed method is simple and generic, it is effective and competitive with other comparable approaches. Moreover, because it can access out-of-the-box of feature transfer learning methods, it is easy to adapt the proposed method for use on a range of other natural language processing tasks.	
Cross-domain sentiment classification-feature divergence, polarity divergence or both? [24]	The proposed Transferring the Polarity of Features (TPF) algorithm transfers the polarity of features from the source domain to the target domain with the independent features acting as the bridge.	K4
	The result showed that the TPF method enhances the performance of the linear classifier and that the TPF method outperforms state-of-the-art methods in cross-domain sentiment classification.	
Cross-domain polarity classification using a knowledge-enhanced meta-classifier [25]	The model uses the BabelNet multilingual semantic network for feature generation together with the proposed Knowledge-enhanced Meta-learning (KE-Meta) algorithm that classifies documents as per their polarity by summing up different types of classifiers.	K5
	The generic ability of KE-Meta is attained by not performing any domain adaptation. Analysis of the information gain of base classifiers showed that Word Sense Disambiguation (WSD) and vocabulary extension are better than bag-of-words-based or n-gram-based classifiers because the former provide extra information that is not given by the latter.	
Building domain-specific sentiment lexicons combining information from many sentiment lexicons and a domain specific corpus [26]	A stochastic formulation is used for the sentiment score assignment problem to cope with score adjustment across different domains.	K6
	The results showed that a dictionary combining information from all the source sentiment dictionaries and from the domain-explicit corpus performs better than one that derives information only from the source dictionaries.	
Cross-domain sentiment classification via topical correspondence transfer [27]	A Topical Correspondence Transfer (TCT) algorithm is developed to learn domain-specific information from different domains and combine them into unified topics, with the help of shared topics across all domains. In this way, the topical correspondence between the shared topics can be used as a bridge to minimize the gap between domains.	K7
	It was found that TCT performs better than the baseline method and at par with state-of-the-art methods for cross-domain sentiment classification.	
Supervised adaptive-transfer PLSA for Cross-Domain text classification [28]	A Supervised Adaptive-transfer Probabilistic Latent Semantic Analysis (SAtPLSA) method is developed for interdomain text classification by extending PLSA to make it a supervised learning algorithm, which is achieved by setting the latent variable as the observable one. Knowledge in the source domain is transferred to assist in classifying the text in the target domain.	K8
	Experimental evaluation of the SAtPLSA algorithm proves that the algorithm can perform efficiently on nine cross-domain text classification benchmark datasets.	
Exploring ensemble of models in taxonomy-based cross-domain sentiment classification [29]	An ensemble algorithm consisting of a Support Vector Machine (SVM) and a transfer learning algorithm named the Spectral Features Alignment (SFA) algorithm is tested on the Amazon dataset. This algorithm takes the tree information and the similarity between domains into account.	K9
	Experimental results showed that an ensemble algorithm consisting of a SVM and SFA algorithm which is able to comprehend the effect of different model application algorithm.	
Cross-domain opinion word identification with query-by-committee active learning [30]	The Opinion Word Identification (OWI) system is developed by utilizing a Query-by-Committee (QBC) active learning scheme and by selecting controlled amounts of data from the new domain for manual annotation to complement the annotated data of the pre-existing domain.	K10
	It was found that QBC gives amazing results even just by adding 1,000 annotated sentences from the new domain to the existing annotated data. The system attains approximate accuracy when it trained on 10,000 annotated sentences.	

TABLE 2. ***(Continued.)*** **Summary of methodology and findings of selected research studies (n = 28).**

An ensemble model for cross-domain polarity classification on Twitter [31]	Four different representations of tweets are created and a classifier is trained on each one of the four representations. An ensemble model combines the output of the different classifications. It was found that merging algorithms trained on different features in a standard training set shows beneficial results, with 81.81% precision on the ensuing training set.	K11
Data intensive review mining for sentiment classification across heterogeneous domains [32]	The proposed approach adopts empirical learning to implement sentiment classification technology, and uses a k-Nearest Neighbor (kNN) distance-based predictive model to combine computational efficiency and modularity. A suitably designed semantic-based metric is the cognitive core that measures the distance between two user reviews. Experimental results confirmed that the overall approach employing kNN attains satisfactory performance in terms of both cross-domain classification accuracy and computational efficiency.	K12
A link-bridged topic model for cross-domain document classification [33]	The link-bridged topic (LBT) model utilizes an auxiliary link network to discover the direct or indirect co-citation relationship among documents by embedding the background knowledge into a graph kernel. The mined co-citation relationship is leveraged to bridge the gap across different domains. Secondly, LBT simultaneously combines the content information and link structures into a unified latent topic model. The experimental results showed that LBT enhances the predicated precision of cross-domain document categorization considerably in comparison to some up-to-date standard algorithms cited in the study. The LBT model also achieves effective knowledge transformation between different domains.	K13
Cross-domain sentiment classification using a sentiment-sensitive thesaurus [34]	A method is proposed to automatically create a sentiment-sensitive thesaurus that is sensitive to sentiment words from different domains. The created thesaurus is used to expand the feature vector in training and testing a binary classifier. The sentiment-sensitive thesaurus exactly assembles words that express comparable sentiments after comparing them against the SentiWordNet.	K14
Feature ensemble plus sample selection: Domain adaptation for sentiment classification [35]	A labeling adaptation method named Feature Ensemble plus Sample Selection (SS-FE) is proposed based on using the Feature Ensemble (FE). The final model is a weighted ensemble of individual classifiers, in which the weights are tuned based on a small amount of labeled data in the target domain. A sample selection method based on Principal Component Analysis (PCA-SS) is also developed. The results showed that both FE and PCA-SS are effective for cross-domain sentiment classification and that SS-FE performs better than either approach because it comprehensively considers both labeling and instance adaptation. Furthermore, assigning weights to training samples or a soft approach seems to be a more appropriate way to extend the proposed method.	K15
Employing emotion keywords to improve cross-domain sentiment classification [36]	A few emotion keywords are used to build an automatically annotated dataset of samples (positive and negative reviews). After that, these automatically generated samples from the target domain are combined with labeled data from the source domain to form the new labeled dataset. Finally, the semi-supervised label propagation algorithm is used to accomplish cross-domain sentiment analysis. Experiments showed that the suggested method can significantly enhance the capacity of domain adaptation in sentiment categorization.	K16
Dynamic Joint Sentiment-Topic model [37]	A Dynamic Joint Sentiment-Topic (DJST) model is proposed as an extension of JST, which allows the detection and tracking of views about current and recurrent interests and shifts in topic and sentiment. Both topic and sentiment dynamics are captured by assuming that the current sentiment-topic-specific word distributions are generated according to the word distributions at previous epochs. The work derives efficient online inference procedures to sequentially update the model with newly arrived data and the effectiveness of the proposed DJST model is demonstrated on the Mozilla add-on reviews crawled between 2007 and 2011.	K17
Active learning for cross-domain sentiment classification [38]	A novel active learning approach is proposed for cross-domain sentiment classification by leveraging Query-by-Committee (QBC)-based sample selection and combination-based classifier classification. Two individual classifiers, the source classifier and the target classifier, are trained with the labeled data from the source and target, respectively. The effectiveness of the active learning approach employing QBC for cross-domain sentiment classification is proven over some strong baselines.	K18
Semi-supervised vs. cross-domain graphs for sentiment analysis [39]	One of the most popular graph-based algorithms, label propagation, is explored, together with its modifications. The impact of modified graph structures and parameter variations are studied and the performance of various graph-based algorithms is compared in cross-domain and semi-supervised settings. The results provide a strategy for selecting the most favorable algorithm and learning paradigm on the basis of the available labeled and unlabeled data.	K19
A case-based approach to cross domain sentiment classification [40]	A case-based method for cross-domain sentiment analysis is proposed that leverages sentiment lexicons and out-of-domain data to build a case-based system where solutions to past cases are reused to predict the sentiment of new documents from an unknown domain. The proposed approach is competitive when equated to a benchmark by means of the paramount outcome from a single-lexicon classifier; generating more solid results by removing the necessity to fix a lexicon preceding making forecasts.	K20
Domain adaptation using domain	A lightweight domain adaptation algorithm is developed that consists of two components: domain similarity	K21

TABLE 2. *(Continued.)* **Summary of methodology and findings of selected research studies (n = 28).**

similarity- and domain complexity-based instance selection for cross-domain sentiment analysis [41]	and domain complexity. Instances are selected from a source domain training set that are the most similar to the target domain.	
	The possible estimation of parameter settings by using domain complexity variance and domain similarity is demonstrated. Moreover, the proposed approach shows statistically significant improvements over natural baselines.	
Biographies or blenders: Which resource is best for cross-domain sentiment analysis? [42]	This work tries to find training data that are similar to those in the test domain, which is very important to obtain high performance in cross-domain sentiment analysis. Several measures of domain similarity are presented. Also, domain complexity is introduced as another independent vector influencing performance loss.	K22
	The proposed model was found to exhibit acceptable behavior, predicting performance loss with an error of 1.5%. For minor exactness drops somewhat the standard error and deviation of the forecast are lower.	
Do neighbours help? An exploration of graph-based algorithms for cross-domain sentiment classification [43]	This study analyzes the performance of two graph-based algorithms, OPTIM and RANK when applied to cross-domain sentiment classification. First, OPTIM is applied to the rating-inference problem in a semi-supervised setting. This study also utilizes domain characteristics to understand the ideal parameter values for the algorithms.	K23
	Analysis of the best parameters for the two graph-based algorithms reveals that there are no optimal values that are valid for all domain pairs and that these values are dependent on the characteristics of corresponding domains. However, there is no apparent regularity in the number of source- and target-domain neighbors.	
SocialTransfer: Cross-domain transfer learning from social streams for media applications [44]	A novel cross-domain real-time transfer learning framework named SocialTransfer is proposed in order to find a way to utilize real-time social streams. SocialTransfer is able to effectively learn from social streams, assuming an intermediate topic space can be built across domains. Experiments are conducted on a real-world large-scale dataset, including 10.2 million tweets and 5.7 million YouTube videos.	K24
	Results showed that SocialTransfer significantly outperforms traditional learners and plays a role in creating an interoperable connection across video and social domains, which could lead to a wide variety of cross-domain applications.	
A two-stage framework for cross-domain sentiment classification [45]	A two-stage framework is proposed for cross-domain sentiment transfer. The first stage is named "building a bridge," which builds a bridge between the source domain and the target domain. The second stage is named "following the structure" in which the structure of the target domain is followed by applying a manifold-ranking algorithm and using the manifold-ranking scores to label the target-domain data.	K25
	The experimental results showed that the suggested method can strongly enhance precision and that it can be used as a high-performance sentiment transfer scheme.	
Automatically extracting polarity-bearing topics for cross-domain sentiment classification [46]	Polarity-bearing topics are initiated by a Joint Sentiment-Topic (JST) model to augment the original feature space.	K26
	The results showed that augmenting the original feature space with polarity-bearing topics allows the in-domain supervised classifier to achieve high-quality results, outperforming Structured Correspondence Learning and providing similar results to Spectral Features Alignment. Moreover, the proposed method does not require demanding parameter tuning.	
Domain adaptation for large-scale sentiment classification: A deep learning approach [47]	Unsupervised feature extraction is performed by using a deep learning system founded on stacked denoising auto-encoders with sparse rectifier units.	K27
	The experiments showed that the proposed method scales well and allows the successful performance of domain adaptation on a larger industrial-strength dataset of 22 domains.	
Cross-domain sentiment classification via spectral feature alignment [19]	A bipartite graph is constructed between domain-independent and domain-specific structures. Subsequently, a Spectral Features Alignment algorithm is proposed in order to bring into line the domain-specific words originating from the source and target domains into expressive groups.	K28
	The viability of the suggested framework is proven by the experimental outcomes for both document-level and sentence-level sentiment classification activities.	

2) POLYSEMY
There is polysemy when the meaning of a word that appears in both the target and source domains changes due to the context of the respective domain. This makes it harder to test the accuracy of the feature representations.

3) FEATURE DIVERGENCE
Also known as feature mismatch or domain mismatch [19], [20], feature divergence refers to the mismatch between the domain-specific features being classified and the knowledge of the classifier if it has been trained on data containing many source-specific features. If this occurs, the classifier trained on domain X may not perform well when applied to domain Y.

4) POLARITY DIVERGENCE
All features have polarity (i.e. they are either positive or negative) and this may differ in the source and the target domain. For instance, ''easy'' could be positive in domain A but negative in domain B. The problem of polarity divergence is present among independent features.

V. PROFILE OF SELECTED STUDIES

A. RESEARCH FOCUS
The research studies on cross-domain sentiment analysis have tried to solve the above problems and issues in relation to the abovementioned three research questions. Some of them focus on evaluating new techniques for cross-domain

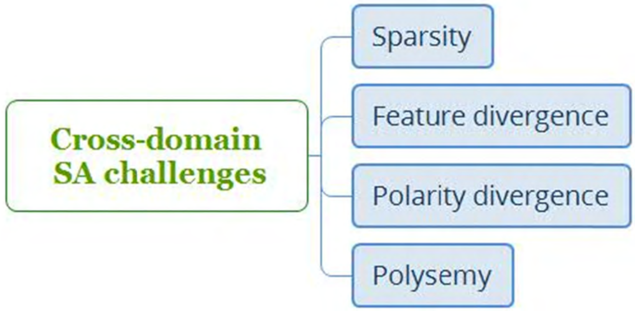

FIGURE 3. Main challenges in cross-domain sentiment analysis.

TABLE 3. Research questions and related key works.

Question	Selected studies
Q1	All
Q2	K1, K2, K3, K4, K10, K12, K13, KK14, K15, K16, K20, K21, K22, K26, K28
Q3	All except K14

sentiment analysis. Others try to study the effect of different data representations and also try to map the features of the source and target domains to discover the domain-independent features that can be used for cross-domain classification. Table 3 shows which of the 28 selected research studies attempt to answer the abovementioned research questions.

B. LANGUAGES AND DATASETS

Among the selected research studies on cross-domain sentiment analysis the vast majority (22 out of 28) are in the English language (K1, K2, K3, K4, K5, K7, K8, K9, K11, K12, K13, K14, K15, K16, K17, K18, K19, K20, K21, K22, K23, K24, K26, K27, K28). The other studies are in the Chinese language (K10, K25) and Norwegian language (K6).

Also, most of these works (17 out of 28) use an Amazon multi-domain dataset as a benchmark dataset. The Amazon dataset was created by [20] and is described in Table 4 below. The Amazon multi-domain dataset includes reviews from four different domains: DVDs (D), Books (B), Kitchen (K) appliances and Electronics (E), and each one contains 1,000 negative and 1,000 positive labeled reviews.

The other research studies (11 out of 28) use a range of publicly available datasets from different domains, as shown in Table 5. Some researchers have manually annotated their own datasets (K10, K11).

C. BASELINE METHODS AND EARLY RESEARCH

In this section, we give an overview of the early works on cross-domain sentiment analysis. In the early days of domain adaptation, groups of classifiers were trained on different source domains [48]. For instance, Transfer Probabilistic Latent Semantic Analysis (TPLSA) was developed by [49] to consider both unlabeled and labeled data. In TPLSA the learning is done via a joint probabilistic model and the documents in the training and test domains are bridged by using hidden variables. The idea behind TPLSA is that the concurrent decomposition of contingency tables related with information on the occurrence of terms in documents in both the training and test domains leads to the identification of prime topics in both the training and the test domains. This work was followed by the introduction of a heterogeneous transfer learning approach that was proposed by [50]. In this method, learning performance is improved if data can be denoted in feature spaces where there is no communication between the data in these spaces. Next, Collaborative Dual-PLSA was developed by [51] to capture domain distinction and commonality among multiple domains at the same time. The two latent aspects of this model are the word concept and the document class. Table 6 summarizes these and some of the other early research studies in the field of cross-domain sentiment analysis.

The selected studies use many baseline methods to evaluate their proposed methods, techniques, algorithms, and approaches. The most common baseline methods are Structured Correspondence Learning (SCL) [61], SFA [19], and Structured Correspondence Learning-Mutual Information (SCL-MI) [20]. Table 7 shows the baseline methods used by the selected studies.

VI. TECHNIQUES FOR CROSS-DOMAIN SENTIMENT ANALYSIS

In the last few years several techniques have been used for cross-domain sentiment analysis. In this section, we attempt to group these techniques according to the following: i) the type of learning model used, for example, topic learning and deep learning; ii) the feature representation method used, such as feature-based reasoning and case-based reasoning (CBR); iii) the resources used by the techniques, which mainly depends on whether they are automatically or manually developed (e.g. the lexicon, such as the sensitive-sentiment thesaurus and the meta-combination of lexical-enhanced classifiers). However, it should be noted that there is some overlapping of these groups and some of the techniques therefore could be classified under more than one group. Nevertheless, generally, these techniques can be categorized into the main groups shown in Figure 4. The key techniques in this figure are explained below.

A. SPECTRAL FEATURE ALIGNMENT (SFA) ALGORITHM

The SFA algorithm was proposed by [19] as a way to align the words from a range of domains into a unified cluster for a specific domain by utilizing knowledge on domain-independent words. The SFA algorithm does several things: it identifies the source and target domains, recognizes the domain-independent and domain-specific features, constructs a bipartite graph, and performs co-clustering and alignment to achieve successful sentiment classification. Pan et al. (2010) first constructed a bipartite graph to model the co-occurrence between domain-specific and domain-independent words into a set of clusters, with the aim of

TABLE 4. Description of Amazon multi-domain dataset.

Dataset	Domain	No. of features	Average length per review	Max/min length per review	Selected studies
Amazon product reviews[1]	Books	188,050	239	3827/13	K1, K2, K3, K4, K5, K7, K9, K12, K14, K15, K18, K19, K21, K22, K23, K26, K27, K28
	DVDS	179,879	234	1723/13	
	Electronics	104,027	153	1303/14	
	Kitchen	89,478	131	1214/15	

reducing any mismatch between the domain-specific words of the source and target domains. This cluster was then utilized to train a classifier for sentiment classification. The proposed SFA algorithm collates the domain-specific words originating from the source and target domains into expressive groups, and the domain-independent words are utilized as a channel to assist in this process. In this way the distance between the domain-specific words of the two domains is reduced. The algorithm is also used to train the sentiment classifier in the target domain. Later, [29] tested an ensemble algorithm consisting of a Support Vector Machine (SVM) and the SFA algorithm on an Amazon dataset. The authors enhanced SFA by the addition of words in shorthand notation and in n-gram form. Next, a SVM-based binary classifier was trained on positive and negative examples of customer reviews. The model takes the tree information and the similarity between domains into account during sentiment classification. In addition, the closest related models in terms of target node, model weight, and model application are selected by using two strategies, the cosine function and a taxonomy-based regression model.

B. STRUCTURED CORRESPONDENCE LEARNING (SCL) ALGORITHM

The SCL algorithm was proposed by [61] as a method for learning the features of a variety of domains. Essentially, the algorithm adapts the source domain to the target domain. The authors state in their work, which focuses on developing an effective binary classifier, that: "A domain is a pair consisting of a distribution D on X and a labeling function f : X [0,1]" [61]. To develop the model, they measured the distance between two distributions, one in the source and one in the target domain, by using hypothesized distance measures based on divergence. Also, features from different domains were discovered by modeling their correlations with pivot features. It should be noted that the pivot feature is particularly useful in semi-supervised machine learning. In the SCL algorithm, the non-pivot features are correlated with similar pivot features. Then a discriminative learner is used in training the classifier. In the SCL methodology, labeled training data is not used in testing, so it is important to be able to model and utilize correlations between features in different domains. The work in [61] was extended the following year by [20], who proposed a new model named the Structured Correspondence Learning-Mutual Information or SCI-MI model. The extension was necessary because SCL depends on the choice of pivot features, and if they are not well-chosen this can adversely affect performance. To address this issue, in SCL-MI [20], the top pivot features are selected by using the mutual information between a feature (unigram or bigram) and a domain label. After the necessary features have been selected, the binary classifier is trained by the SCL algorithm. More recently, [76] drew on the concept of SCL to develop a method that utilizes two auxiliary tasks to help induce sentence embedding because the authors expected the embedding technique to perform well across the sentiment classification domains. They also suggested that the sentence embedding step take place while the sentiment classifier is itself in the learning phase.

C. JOINT SENTIMENT-TOPIC (JST) MODEL

The JST model proposed by [46] is a probabilistic modeling framework based on Latent Dirichlet Allocation (LDA). Several machine learning approaches for sentiment classification require labeled corpora in order to train the classifier. In contrast, the JST model is totally unsupervised. The JST model is based on the authors' research on polarity-bearing topics, which they used to enhance the original feature space. Learning in the JST model is based on prior information about the domain-independent polarity words. The JST model is an extension of the LDA model proposed by [77] that was developed to detect a sentiment and a topic simultaneously from text. In the JST model, discriminative classifiers are used to search for a decision boundary that maximizes a certain measure of separation between classes. The posterior distribution is calculated by applying sequential sampling for each variable (known as Gibbs sampling). The JST method can cluster different terms that exhibit a similar sentiment. Information gain criteria are used to augment and select the features for cross-domain classification. The JST model was later improved upon by [37], who developed the Dynamic Joint Sentiment-Topic Model (DJST). As the name implies, the DJST model can be used to identify and track interests and shifts in topic and sentiment over time. Both the topic and sentiment are captured dynamically by assuming that the current sentiment-topic-specific word distributions are dependent on earlier word distributions. The authors looked at three different ways to acquire information on these dependencies: the sliding window, a skip model and a multiscale model. In the sliding window the current sentiment-topic word distributions are dependent on the previous sentiment-topic-specific word distributions in the last S epochs.

TABLE 5. Summary of datasets used for evaluation of cross-domain sentiment analysis methods.

Dataset	Author/Year	Domain	Description	Language	Key
Norwegian product reviews	Hammer et al. (2015)	Products	15,118 product reviews from the Norwegian online shopping sites www.komplett.no and mpx.no	Norwegian	K6
Newsgroups[2]	Lang (1995)	News	Comprises approximately 18,000 newsgroup posts on 20 subcategories	English	K8
Reuters-21578[3]	Lewis (1997)	News	Has a hierarchical category of structures suitable for cross-domain learning settings		
Manually annotated datasets	Tsai et al. (2014)	Restaurants	Consists of 10,000 review sentences from iPeen.com.tw	Chinese	K10
		Movies	Consists of 10,000 review sentences from atmovies.com.tw		
		Hotels	Consists of 10,000 review sentences from agoda.com		
Emoticons dataset “ED”	Tsakalidis (2014)	Emoticons	Contains 250,000 tweets containing sad /happy (“”:(” , “:)”) emoticons	English	K11
Stanford Twitter dataset test set (STS)	Go et al. (2009)	Tweets	First version consists of 184 negative and 177 positive tweets; second one consists of 75 negative and 108 positive tweets		
Obama Healthcare reform (HCR)	Speriosu et al. (2011)	Healthcare	Contains tweets related to Barrack Obama’s healthcare reform presented in 2010		
Obama-McCain debate (OMD)	Speriosu et al. (2011)	Debate	Related to the 2008 debate between Obama and McCain and contains 3,269 labeled tweets (1,190 negative and 707 positive)		
TripAdvisor dataset	Wang et al. (2010)	Hotel	Consists of 235,793 hotel reviews from TripAdvisor	English	K12
Cora datasets	McCallum et al. (2000)	Scientific	Contains roughly 37,000 papers and more than one million links among approximately 200,000 separate documents; the dataset’s documents are hierarchically categorized	English	K13
Industry sector datasets[4]		Companies	Collection of around 10,000 web pages belonging to firms from numerous economic segments; the corporate web pages are hierarchically categorized		
Multi-domain emotional comments corpus[5]	Blitzer et al. (2007)	Emotional	Contains four domains: DVD, Book, Electronic, and Kitchen	English English	K16 K22
IMDB dataset	Pang, Lee, & Vaithyanathan (2002)	Movies	An IMDB dataset of film reviews	English	K20
Hotel reviews dataset	Pang et al. (2002)	Hotels	A hotel reviews dataset		
Product reviews	Baccianella et al. (2009)	Products	Contains reviews for books, electronics, apparel, and music from Amazon.com		
Real-world large-scale dataset	Jindal & Liu (2008)	Tweets and videos	Contains 5.7 million YouTube videos and 10.2 million tweets	English	K24
Book reviews[6]		Books	Includes 4,000 annotated book reviews (2000 negative and 2000 positive) from www.dangdang.com/	Chinese	K25
Hotel reviews[7]		Hotels	Includes 4,000 annotated hotel reviews (2000 negative and 2000 positive) from www. ctrip.com/		
Notebook reviews[8]		Notebooks	Includes 4,000 annotated notebook reviews (2000 negative and 2000 positive) from www.360buy. com/		

TABLE 5. *(Continued.)* **Summary of datasets used for evaluation of cross-domain sentiment analysis methods.**

movie review (MR)[9]	Pang et al. (2002)	Movie	Includes 2,000 movie reviews (1,000 negative and 1,000 positive) from IMDB	English	K26

[1] http://www.cs.jhu.edu/˜mdredze/datasets/sentiment/index2.html
[2] http://people.csail.mit.edu/jrennie/20Newsgroups
[3] http://www.daviddlewis.com/resources/testcollections
[4] http://people.cs.umass.edu/~mccallum/data.html.
[5] http://www.seas.upenn.edu/~mdredze/datasets/sentiment/
[6] www.searchforum.org.cn/tansongbo/corpus/Dangdang_Book_4000.rar.
[7] www.searchforum.org.cn/tansongbo/corpus/Ctrip_htl_4000.rar.
[8] www.searchforum.org.cn/tansongbo/corpus/Jingdong_NB_4000.rar.
[9] 3http://www.cs.cornell.edu/people/pabo/movie-review-data

TABLE 6. **Overview of early studies on cross-domain classification.**

Author and year	Machine learning/ Natural language processing	Algorithm	Dataset	Classifier
Blitzer et al. (2006)	POS	SCL	Amazon	SVM
Blitzer et al. (2007)	POS	SCL-MI	Amazon	SVM
Xue et al. (2008)	TF and IDF	PLSA	Net news articles, SRAA & Newswire articles	SVM and NBC
Li & Zong (2008b)	N-gram method	Feature-level fusion	Amazon	LIBSVM
Guo et al. (2009)	LaSA and LDA	MI	Wikipedia, Chinese newspapers	PRM classifier
Wang et al. (2008)	TF and IDF	MI	Newsgroup & SRAA	NBC
Paltoglou & Thelwall (2010)	TF and IDF	SMART and BM25 tf.idf variants	Amazon, movie dataset	SVM and NBC
Jakob & Gurevych (2010)	POS	CRF	Amazon	SVM
Huang & Yates (2010)	CRF	HHM-based Model	Wall Street Journal	ME

However, in the skip model the sentiment-topic-word distributions are considered by skipping some of the earlier epochs that fall between the current and a selected previous epoch. Lastly, in the multiscale model a range of previous long- and short-timescale distributions are taken into consideration.

D. ACTIVE LEARNING AND DEEP LEARNING

To be effective, the active and deep learning methods need to choose the data from which they will learn in order to perform well with less training [78]. In other words: "Active learning is a special case of semi-supervised machine learning in which a learning algorithm is able to interactively query the user (or some other information source) to obtain the desired outputs at new data points" [79], [80]. In essence, this method uses the source domain information to get additional labeled target data. The challenge to overcome when designing an active learning model is the size of the labeled data both in the source domain and the target domain. The three types of scenario and the related query strategies that are used in active learning are illustrated in Figures 5 and 6 while Table 8 gives the pros and cons of the three active learning scenarios.

[17] http://www.cs.jhu.edu/~mdredze/datasets/sentiment/index2.html
[2] http://people.csail.mit.edu/jrennie/20Newsgroups
[3] http://www.daviddlewis.com/resources/testcollections
[4] http://people.cs.umass.edu/~mccallum/data.html.
[5] http://www.seas.upenn.edu/~mdredze/datasets/sentiment/
[6] www.searchforum.org.cn/tansongbo/corpus/Dangdang_Book_4000.rar.
[7] www.searchforum.org.cn/tansongbo/corpus/Ctrip_htl_4000.rar.
[8] www.searchforum.org.cn/tansongbo/corpus/Jingdong_NB_4000.rar.
[9] 3http://www.cs.cornell.edu/people/pabo/movie-review-data

Very few research studies have used an active learning approach to address the problems encountered in cross-domain sentiment classification. One such study that attempts to do so is that of [38], who proposed an active learning approach that incorporates a Query-by-Committee (QBC) method for sample selection and a two-classifier combination for classification. In their proposed method, a source classifier is trained on labeled data from the source domain and likewise a target classifier is trained on labeled data from the target domain. The two classifiers are trained by fully exploiting the unlabeled data in the target domain with the label propagation (LP) algorithm. Then, the two classifiers each select informative samples by using QBC. Lastly, the two classifiers are combined to make the final classification decision. Another work, that of [30], also uses QBC for active learning in order to identify opinion words. Controlled amounts of data are selected from the new domain for manual annotation to complement and supplement the annotated data of the pre-existing domain. The authors found that the utilization of QBC was able to generate good results after the addition of 1,000 annotated sentences from the new domain to the existing annotated data, and the system attained approximate accuracy when it was trained on 10,000 annotated sentences.

On the other hand, in deep learning, the aim is to obtain a meaningful representation of a sentiment. The deep learning approach is unsupervised and it aims to extract high-level features from unlabeled data. To be successful, a deep learning algorithm has to discover the intermediate concepts that are common to both source and target domains. For example,

TABLE 7. Baseline methods used by selected studies.

Method/name	Abbreviation	Author and year	Key
No Transfer	NoTransf	-	K1, K7, K28, K4
Structured Correspondence Learning	SCL	Blitzer et al. (2006)	K1, K2, K3, K7, K26, K27, K28, K4
SCL Mutual Information	SCL-MI	Blitzer et al. (2007)	K14, K21, K4
Partially Supervised Cross-Collection Latent Dirichlet Allocation	PSCCLDA	Bao et al. (2013)	K8
Spectral Feature Alignment	SFA	Pan et al. (2010)	K1, K2, K7, K8, K14, K21, K26, K27, K9, K4
Latent Semantic Analysis	LSA	D. C. T. Hofmann (2001)	K28
Feature Latent Semantic Analysis	FALSA	Serafin & Di Eugenio (2004)	K28
Transfer Probabilistic Latent Semantic Analysis	TPLSA	Xue et al. (2008)	K8
Collaborative Dual-Probability Latent Semantic Analysis Manifold	CDPLSA	Zhuang et al. (2010)	K8
Maximum Likelihood with Expectation Maximization	EM Algorithm	Dempster et al. (1977)	K25
Non-negative Matrix Tri-Factorization	NMTF	Li et al. (2009)	K7
Transfer Component Analysis	TCA	Li et al. (2012)	K7
Majority Class classifier	MC	Majority class classifier	K11
Multilabel Consensus Training SentiRank	MCT	Li & Zong (2008a)	K27
Transductive Support Vector Machines	TSVM	Joachims (1999)	K25
SentiRank	SentiRank	Wu et al. (2009)	K25
Traditional supervised classifier	Porto	Traditional supervised classifier	K25

product quality, product price, and customer service would be intermediate concepts in the case of product reviews. The features that are identified by the algorithm are then used to train the classifiers. In [47], unsupervised feature extraction is performed by using stacked denoising auto-encoders with sparse rectifier units for deep learning in a two-step process to classify sentiments across different domains. First, the high-level features are extracted by using a Stacked Denoising Auto-encoder (SDA) with rectifier units. The SDA is taught by using stochastic gradient descent in a greedy layer-wise manner. Second, the classifier is taught by using the transformed labeled data from the source domain.

Lately, [81] proposed a potential solution to the issue of domain adaptation in sentiment classification that utilizes both deep learning and ensemble methods, where the former is used for cross-domain high-level feature representation and the latter for reducing the amount of generalization errors across domains. When tested, the proposed method showed a significant improvement in terms of generalization compared to the state of the art, implying that the combined usage of deep learning techniques and ensemble methods could be a fruitful research direction for the task of domain adaptation.

Also in relation to deep learning, [82] demonstrated that, while deep neural networks can learn to undertake the task of abstract feature representation and thereby reduce the incidence of cross-domain discrepancy, such networks cannot entirely overcome the problem. In an attempt to address this, the authors decided to try to improve the invariance of deep representation and make it more transferable across domains by developing a unified framework for deep adaptation that consists of a method to learn transferable representation and a classifier to enable scalable domain adaptation. The framework draws on the advantages that are inherent in deep learning and in optimal two-sample matching. Specifically, the framework consists of (1) a method for unsupervised pre-training for the effective training of deep models that uses deep denoising autoencoders and (2) a method for supervised fine-tuning for effective exploitation of discriminative information using deep neural networks. These two methods are interdependent and both learn by embedding deep representations of reproducing kernel Hilbert spaces (RKHSs) and optimally matching different domain distributions. Also, to enable scalable learning, the authors developed a linear-time algorithm that uses the unbiased estimate to linearly scale large samples.

E. TOPIC MODELING

Topic modeling approaches are based on the concept of latent semantic indexing, the aim of which is to reduce high dimensionality in a term-document matrix into low dimensions that are denoted as latent topics. Topic modeling uses clustering techniques that do not require label information. The four main approaches that fall under this category are as follows:

1) TOPICAL CORRESPONDENCE TRANSFER (TCT)

Developed by [27], the TCT algorithm learns domain-specific information from several domains and creates unified topics with the help of knowledge about shared topics in all those domains. In TCT, the documents in each domain are represented in a term-document matrix. The term-document matrix is the approximated product of two specific matrices. One contains the domain-specific topics and the shared topics. The other contains a representation of the document(s) based on those topics. The TCT algorithm is essentially an optimization technique that uses joint non-negative matrix factorization. A document's sentiment labels are incorporated

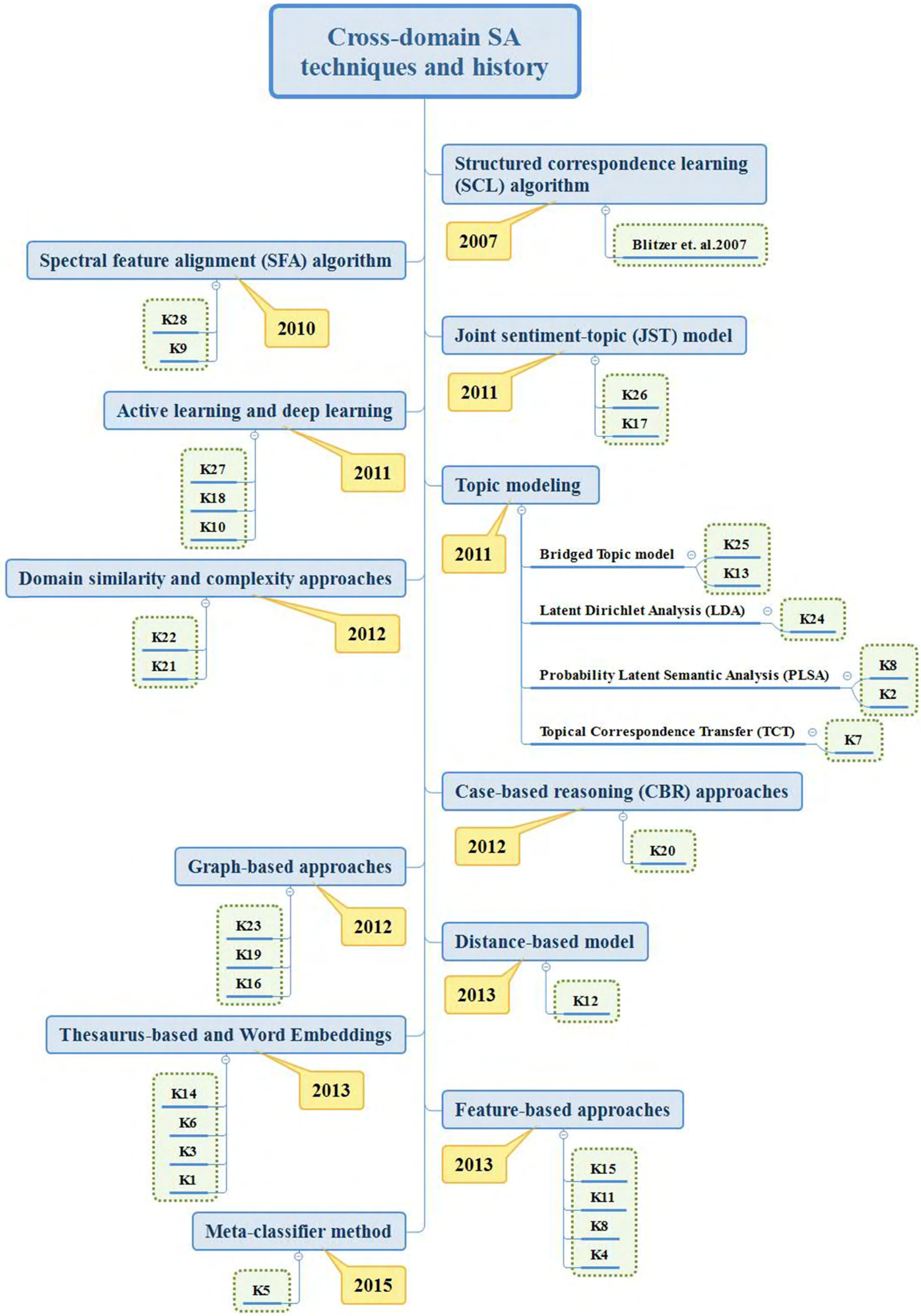

FIGURE 4. Techniques for cross-domain sentiment analysis.

by applying a least squares penalty based on a specific linear model. An objective function learns to represent the topics at the same time because it learns to predict the document's sentiment. The optimization of the objective function enables topics to be identified automatically and the sentiment to be classified according to the

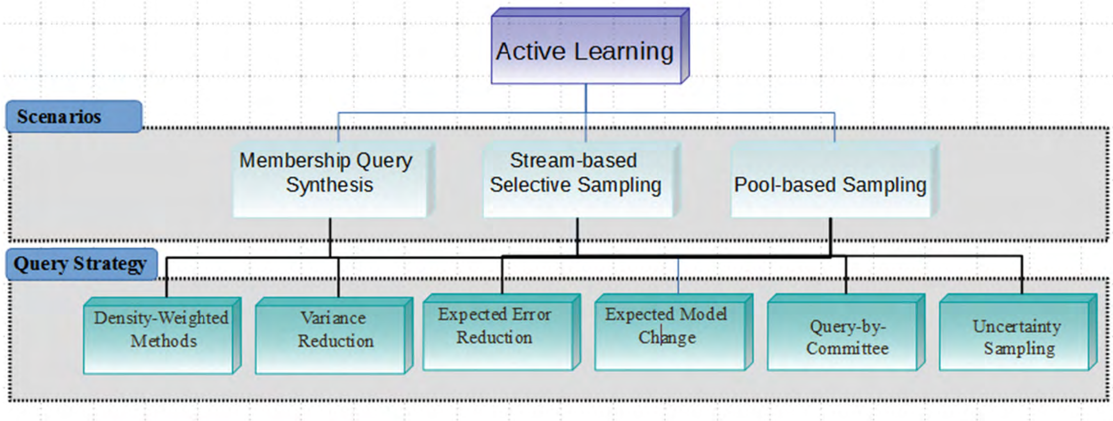

FIGURE 5. Active learning scenarios and query strategy.

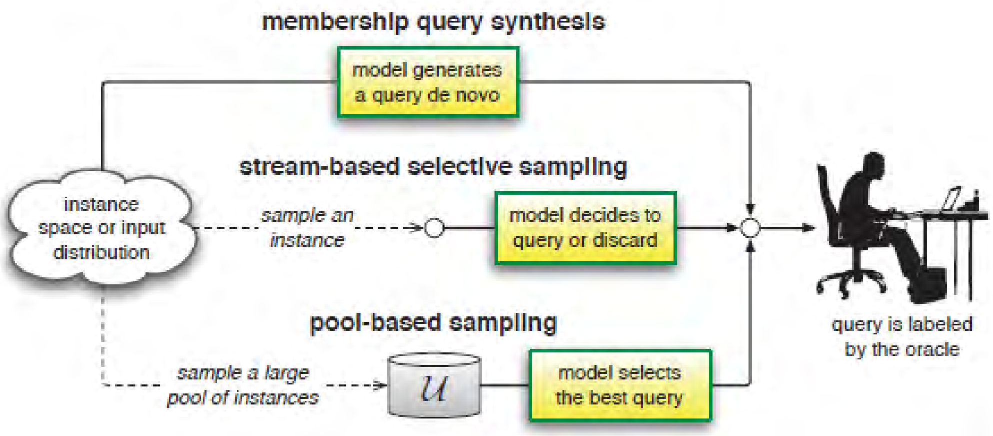

FIGURE 6. Three main active learning scenarios [79].

topic representation. Thus the hidden correspondence between the shared topics can be used to reduce the gap between the source and target domains, and can thus be described as a bridging technique.

2) BRIDGED TOPIC MODEL

The goal of transfer learning is to identify and exploit shared common structures and properties that reside in different domains in order to achieve knowledge transformation. However, if a cross-domain document classification method only focuses on the data within the source and target domains, it might fail to capture commonalities in the domains that are, in fact, indirectly connected. To address this issue, a Link-Bridged Topic (LBT) model was proposed in [33]. The LBT model uses an auxiliary link network to find direct and indirect co-citation relationships among documents. It does this by embedding background knowledge in a graph kernel. The identified co-citation relationships are then used to bridge the gap between domains. In addition, the LBT approach combines content information and link structures into a unified latent topic model. Other work on bridged topic models includes that of [38], who proposed a framework for cross-domain sentiment transfer that starts by "building a bridge" between source and target domains by using the SentiRank algorithm [75] to get the sentiment scores for the target-domain documents. The model then utilizes these sentiment

TABLE 8. Pros and cons of active learning scenarios.

Scenario	Pros	Cons
Membership query synthesis	- Computationally controllable (for finite domains) - Can be used for regression actitivities - Could be used in experiments that do not require human involvement in the annotation process	- A person may have difficulty interpreting and labeling arbitrary instances manually
Stream-based selective sampling	- Better option when memory or processing power is limited, such as in the case of mobile and embedded devices	- Access to free unlabeled data might not always be available
Pool-based sampling	- Seems to be most common sampling method as it is used for text, information, image, and video classification, as well as speech recognition, among others - Applied to many real-world tasks	- Has a high computational cost - The entire set has to be evaluated for each iteration

scores to identify a small number of the best-labeled documents that are employed to represent the intrinsic structure of the target domain. Next, by ''following the structure'' of this target domain, a manifold-ranking algorithm is applied and the resultant manifold-ranking scores are used to label the target-domain data.

3) LATENT DIRICHLET ANALYSIS (LDA)

This type of approach has already been mentioned above in the relation to the JST model. In relation to topic modeling more generally, [44] proposed a cross-domain real-time transfer learning framework based on LDA named SocialTransfer. It has a topic space that is learned in real time from social streams via Online Streaming LDA, also known as OSLDA. It also uses real-time cross-domain graph spectra analysis based on a transfer learning method that incorporates the topic models learned from social streams into its transfer learning framework.

4) PROBABILITY LATENT SEMANTIC ANALYSIS (PLSA)

This is a widely used probabilistic topic modeling approach for text mining [83]. This idea was extended by [28], who proposed the Supervised Adaptive-transfer PLSA (SAtPLSA) algorithm for cross-domain text classification. In that work, by making the latent variable the observable variable, the authors were able to modify the PLSA to make it a supervised learning algorithm. By doing this the latent topic becomes the category to which the document belongs. To train the algorithm on documents in the source domain, the class-conditional probability of a specific word conditioned on a class is estimated directly during initialization and is then fixed in the model fitting step. To test the algorithm on documents from the target domain, the word-category probabilities are assigned randomly and learned during the whole process. The classification tasks in the source and target domains are the same, so the class label sets of the two domains are also the same. Consequently, the word-category probabilities serve as a bridge to connect the two domains. Later on, [22] developed a latent sentiment factorization (LSF) algorithm that is based on probabilistic matrix factorization. The proposed method is suitable for use in sentiment classification when only labeled data exist in the source domain and only unlabeled data are present in the target domain. The LSF algorithm is able overcome the gap between the two domains by exploiting sentiment correlations between domain-shared and domain-specific words in a two-dimensional sentiment space.

F. THESAURUS-BASED APPROACHES

Another approach to cross-domain sentiment classification involves the use of a thesaurus. For instance, [34] created an automatic classifier based on a sentiment-sensitive thesaurus in order to avoid feature mismatches. The authors utilized labeled data from several source domains and unlabeled data from the source and target domains to calculate the relatedness of characteristics and to conceptualize the thesaurus. Then it was used to extend the feature vectors for a binary classifier that was applied to training and test data. In order to find the word co-occurrences, they used the pointwise mutual information approach, where each lexical element is either a unigram or bigram connected to a list of related lexical elements that appear most frequently in a sentence. To expand the feature vectors, they used their sentiment-sensitive thesaurus of related elements in the training step to overcome the feature mismatch problem. The relatedness measure was used to measure the neighbors for a particular lexical element from the list of different lexical elements. The suggested approach was found to be considerably better than numerous baselines. Moreover, its performance was analogous with other comparable cross-domain sentiment classification procedures when applied to a standard dataset. Furthermore, the sentiment-sensitive thesaurus exactly assembles words that express comparable sentiments after comparing them against SentiWordNet. Later, [84] developed an enhanced sentiment-sensitive thesaurus using Wiktionary. The thesaurus in that work aligns different words that express the same sentiment from reviews in different domains and from Wiktionary to improve classification performance in the target domain.

Recently, [85] proposed two methods for the domain adaptation of a polarity lexicon that are both corpus based and language independent. Moreover, both of the methods are suitable for any domain. One is based on term frequency (TF) and needs a corpus that has already been tagged with the polarity of the documents. The other is based on a bootstrapping (BS) algorithm and does not need an annotated corpus; instead, it just requires a set of patterns and a seed sentiment lexicon. The proposed TF approach was found to achieve very promising results, but to the authors' surprise, the BS approach did not result in a lot of improvement. They therefore combined the two methods to try to gain benefit from the best aspects of each and found that the combined approach was able to get results that were much better than those that could be obtained by the separate application of the individual methods.

On the other hand, [21] chose to model the problem of cross-domain sentiment classification as one of embedded learning. They constructed three objective functions that they applied together and insolation: (a) distributional properties of pivots (common features in both the source and target domains), (b) label constraints in source domain documents, and (c) geometric properties in source and target domain unlabeled documents. In contrast to earlier related works in which a lower-dimensional embedding is learned first, then the source domain sentiment labels are learned, and next a sentiment classifier is used in the embedding, the authors developed a joint optimization method that learns embeddings that are sensitive to sentiment classification. The results of some experiments revealed that by jointly optimizing the three objective functions better performance can be achieved than by optimizing each one individually. Their results clearly demonstrate the benefit of using task-specific embedding learning in cross-domain sentiment classification. As regards the best performance of an individual objective function, this is achieved by objective function (c). In a similar vein, [23] proposed a method that uses feature learning and feature subspace mapping, i.e. word embeddings and canonical correlation analysis (CCA), to address vocabulary mismatches that can occur between source and target domains. The authors showed that by using what are essentially quite simple, generic methods, it is still possible to obtain very competitive results compared to those produced by more complicated methods that have been designed to solve the same problem. A particular advantage of using word embeddings and CCA is that they can be accessed out-of-the-box, which is a key benefit with respect to the applicability of the proposed method.

G. CASE-BASED REASONING (CBR) APPROACHES

The idea behind CBR [86] is to draw on knowledge about similar past examples to predict the outcome of a new unseen one. This approach was applied to sentiment classification by [40], who developed a case base from a training set of labeled out-of-domain opinion documents. The case base consists of two key parts. One is a case description, which is a feature vector based on a document's statistics and acts as document signature that is used for retrieval purposes. The case description attempts to broadly capture a document's characteristics as a set of features; it does not focus on potential domain-specific aspects, such as specific terminology. The other part of the case base is a case solution, which stores information about successful predictions made during training. The case solution is comprised of all the lexicons that made accurate forecasts during training. The authors assessed their proposed CBR method by testing it on user-created reviews in six distinct domains. They found that the performance of their method was competitive when equated to a benchmark by means of the paramount outcome from a single-lexicon classifier; generating more solid results by removing the necessity to fix a lexicon preceding making forecasts. From the successful result reported in [40], it is clear that CBR approaches can play an important role in improving unsupervised sentiment classification. Indeed, other researchers have utilized sentiment dictionaries/ lexicons for sentiment analysis. For instance, [26] developed a process to build a domain-explicit sentiment dictionary by taking an undefined amount of data from a particular domain and combining it with information from numerous pre-existing sentiment dictionaries. Prior to their work, the combining of sources of information had rarely been investigated. To deal with domain-explicit variations, the authors adopted a stochastically formulated sentiment score assignment rather than a deterministic formulation. As a result, the authors were able to reduce the expected loss of a loss function that penalizes variations from the scores of the source sentiment dictionaries and the inhomogeneity of the sentiment scores for the same review.

H. FEATURE-BASED APPROACHES

As stated in [28], feature representation-transfer methods aim to obtain a good feature representation in order to improve the performance of classification in the target domain. In [35], a comprehensive approach is proposed named Feature Ensemble plus Sample Selection (SS-FE). The authors take both types of adaptation into account: (i) a Feature Ensemble (FE) model is first adopted to learn a new labeling function in a feature re-weighting manner and (ii) a Principal Component Analysis (PCA)-based sample selection method named PCA-SS is then used to help the FE. However, a deficiency in their approach concerns the way in which training samples are selected because the method uses unweighted training samples or a hard sampling approach, which results in occasional randomness. In [31], to tackle the domain-dependence issue, a range of representations, namely text-based representation, feature-based representation, lexicon-based representation, and combined representation, are all used to create an ensemble algorithm that consists of several classifiers where each one is trained by using one of these distinct feature representation methods. On the other hand, in [24], a novel approach is proposed named Transferring the Polarity of Features (TPF), which addresses the issue of not only feature divergence, but also polarity divergence. The TPF approach deals with these two issues first by selecting the independent features with higher polarities in both domains and creating a set of these high-polarity independent features. Then the algorithm transfers the polarity of the features from the source domain to the target domain by using these independent features as a bridge.

I. GRAPH-BASED APPROACHES

Graph-based algorithms are ideal for data that can be represented in the form of a weighted graph. In such a graph, the vertices are data instances and the edge weights denote the similarity between those instances. However, the data have to be available in the form of a manifold structure because this means that the instances are often strongly connected and belong to the same class. If the data do not form a graph in this way, then it is necessary to find a similarity function that can approximate the similarity between graph vertices as accurately as possible and thus enable the construction of a graph from such data [87]. The construction of a good graph that provides an appropriate estimation of the similarity between data instances is essential to the success of a graph-based algorithm [87]. The LP algorithm was one of the first graph-based algorithms to be developed [88] and it remains one of the most popular. It involves an iterative process that propagates information from labeled to unlabeled nodes, which in the case of sentiment classification are documents. This process continues until convergence is reached. The LP approach is, in effect, a weighted averaging of labels in the neighborhood of a node where the influence of neighbors is defined by the edge weights. In the context of sentiment classification, the edge weights indicate the closeness of document ratings. In another work, [39] explores LP and some modifications with the aim of improving sentiment classification. Specifically, the authors investigate the effect of adopting modified graph structures and varying the parameters in order to compare the performance of a range of graph-based algorithms for cross-domain and semi-supervised classification.

In [43], the performance of two previously proposed graph-based algorithms, OPTIM [89] and RANK [77], is compared in terms of effectiveness for cross-domain sentiment classification. The OPTIM algorithm deals with sentiments as an optimization problem, whereas the RANK algorithm uses a ranking to assign sentiment scores. As document similarity is key to the success of graph-based algorithms, [43] also investigate a range of sentiment similarity measures to determine which perform better. As the term implies, sentiment similarity compares documents with respect to the sentiment they convey. Finally, the authors determine how domain properties influence the values of algorithm parameters, and conclude that there are no universal parameter values that can be used for all domain pairs. Later, [36] suggested an original method to enhance the capacity of domain adaptation for sentiment categorization by using some emotion word groups. Initially, they utilized a few emotion keywords to mine an automatically tagged model from the target domain with a high level of accuracy. Then, using LP they ran a semi-supervised learning process to find the sentiment arrangement.

J. DOMAIN SIMILARITY AND COMPLEXITY APPROACHES

Domain similarity is a method that can also be used in domain adaptation. It is based on selecting the instances from a training set of a source domain that are the most similar to those in a target domain. The factor by which the original source-domain training set size is reduced is determined automatically by measuring domain similarity between source and target domains as well as the variance in the complexity of the domains. In a related work, [41] tried to find training data that are similar to those in the test domain because it is essential that there is as much similarity as possible in order for cross-domain sentiment analysis to achieve a high level of accuracy. The authors suggested that the domain complexity of a dataset can be ascertained by (i) the percentage of rare words, i.e. words occurring less than three times in the domain, (ii) the type/token-ratio of the domain, and (iii) the relative entropy of the domain, where the entropy normalized by the entropy of the uniform distribution. Among the three methods, they found that the percentage of rare words correlates best with in-domain accuracy in document-level polarity classification. In addition, they presented several measures for domain similarity and showed that domain complexity is an independent vector that can influence performance loss. In [42], domain similarity is measured as divergence between the term and unigram distribution, where domain complexity is measured as homogeneity (domain self-similarity). The authors showed the performance loss that can occur in various data groups. Determining this loss can help in the selection of data that are the most comparable to existing test data. Initially, they presented measures for domain complexity and similarity and then ascertained their effect on the performance loss of a cross-domain classifier. Then, by means of these measures, they developed a linear regression model that they used to validate the measures and then performed experiments using various parameters.

K. KNOWLEDGE-ENHANCED META-CLASSIFIER METHOD

The penultimate method discussed here is that proposed by [25], namely the Knowledge-enhanced meta-learning (KE-Meta) algorithm, which is used to combine and enrich bag-of-words n-grams or lexical resource-based classifiers by adding other knowledge features. The authors used a semantic network for Word Sense Disambiguation (WSD). Then, using the disambiguated terms, the semantic network was able to obtain a vocabulary expansion-based classifier. In another work [90], Babelnet was used as a multilingual semantic network to generate features from WSD and vocabulary expansion. The authors combined a range of established classifiers, namely the bag-of-words classifier, word n-gram classifier, lexical resources-based classifier, WSD-based classifier, and vocabulary expansion-based classifier, by using a meta-learning stacking method.

L. DISTANCE-BASED MODEL

Lastly, [32] adapted a distance-based predictive model for sentiment classification of review documents. The model consists of three main steps. First the training corpus and distance metric are defined. Second, a new review is classified and the distance metric is used to identify it in the training corpus. Third, an unlabeled review is tagged according to a majority-rule strategy.

VII. DISCUSSION

The aim of the vast majority of studies in the area of sentiment classification has been to reduce the distribution difference between the source and target domains. However, this a difficult task because most of the techniques suffer from a domain dependency problem and their efficiency falls if there is a feature-space distribution discrepancy between the source and target domains. In addition, the performance of some techniques is affected by the availability or otherwise of labeled data. Moreover, cross-domain approaches perform poorly when there is a large disparity between the training and test data. Studies on sentiment classification still mostly deal with binary classification, which unfortunately does not provide practically feasible results. Although manual effort is not required in cross-domain learning, there is an over-reliance on similarity between the source and target domains. Therefore, improving the accuracy of sentiment classification techniques requires the application of novel feature representation methods, further testing, and the investigation of the potential of a variety of combined or ensemble methods. There is also a need to develop methods that take into consideration the problems of feature deviation and polarity divergence. Moreover, a way needs to be found to reduce

TABLE 9. Pros and cons of the most common baseline methods.

Method	Pros	Cons
SFA	• Can be used efficiently to adapt multiple domains	• Costly to annotate data for each new domain because picking optimal pivot features can only be done empirically • Lacks unified theoretical guidance
SCL SCL-MI	• Can be applied with different machine learning models • Does not require labeled data in the target domain	• Identification of pivot features may fail if the correlation among features is low • Training multiple classifiers for pivot features may be too expensive
JST DJST	• Highly portable to other domains • Can detect sentiments and topics simultaneously • Can readily meet the demands of large-scale sentiment analysis	• Hard to analyze the sentiments of subtopics
Active learning	• Useful when human effort, becomes a scarce resource • Useful when collecting labeled data is expensive and there is a need to build classifiers by using fewer training data to expand the labeled data pool	• Active learning in cross-domain classification normally contains abundant labeled data in the source domain • Large amount of labeled data results in difficulties in both the sample selection strategy and the classification algorithm for active learning in the cross-domain case
Graph-based	• Clear, simple, and elegant math • Good result when the graph fits the task. • Can handle multi-class classification	• Result is unsatisfactory when the graph is bad • Highly dependent on graph structure and weighting methods • Unpredictable behavior and output
Topic-based	• More compact and meaningful representation of documents • Shared topics and domain-specific topics that are discovered are more coherent and easier to understand	• Reliance on only out-of-domain data might be risky • Largely depends on both in- and out-of-domain datasets
Thesaurus-based	• Obtains competitive results in single-domain polarity classification (K5) • Achieves significant results as long as there are enough unlabeled reviews from the target domain (K5)	• Lacks unified theoretical guidance • Results are highly sensitive to source and target domains

if not totally remove the need for external resources or human intervention. There also needs to be greater stability in the results across a wide range of source and target domains. Table 9 presents the pros and cons of the most commonly used methods in cross-domain sentiment analysis to date.

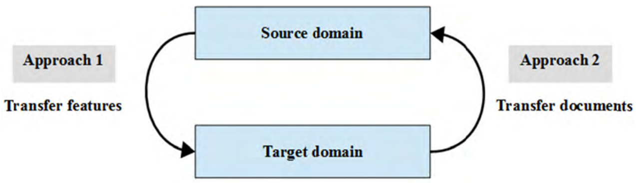

FIGURE 7. Two types of approaches for cross-domain sentiment analysis.

Based on our review of the methods that have been developed thus far for cross-domain sentiment analysis, it is clear that the majority of these methods can be classified into two main approaches (see Figure 7). The first is based on the transfer of source-domain features (training domain) to the target domain (test domain). Feature-based methods and thesaurus-based methods fall under this heading. The second is based on the transfer of some complete documents from the target domain to the source domain that then become part of the training documents for the source domain. Active learning methods fall into this group. Bearing in mind the advantages and disadvantages of all the above-discussed methods as well as their results, we need to propose a new method based on the benefits of the previous methods and that can make the most of the common features of both the source domain and the target domain to obtain the best sentiment classification possible.

There a still quite a few difficult challenges to overcome in the field of cross-domain sentiment classification. For instance, where there are different data distributions among domains and data sources this can lead to is domain and data discrepancy. Also, a solution has yet to be developed that can be used successfully on real-world industrial datasets that contain numerous domains. There are also cultural factors, linguistic variations, and noises in data and differing contexts that make it very difficult to perform cross-domain sentiment classification with a high level of accuracy. Such factors can lead to training data from the same domain (e.g. movie reviews) yielding completely different results.

VIII. CONCLUSION

Over the years, sentiment analysis has received a lot of attention from researchers working in the fields of natural language processing and text mining. However, there is a lack of annotated datasets that can be used to train a model for all domains, which is hampering the accuracy of sentiment analysis. Any research studies have been carried out to tackle this issue and to improve cross-domain sentiment classification. To aid researchers in their endeavors, we undertook a systematic literature review to highlight the main cross-domain challenges and techniques. From our review, it is clear that while there is no perfect solution as yet, much has been done to improve the accuracy of cross-domain sentiment analysis, which will pave the way for further enhancement in the future. A unified performance evaluation method could be designed to analyze these methods in as a future work of this research.

There are several potentially fruitful paths of research that could be followed by researchers in the coming years. The first path worthy of attention is that of multi-view learning in which some semi-supervised methods are utilized in combination to exploit various views of the same input data. The second concerns the use of models that can learn domain-independent features from the datasets of multiple source domains. A third avenue to follow would be to develop models that are able to adapt and use a domain-independent subjectivity lexicon by propagating polarity through the pivot domain, rather than from the source to the target domain. Fourthly, there is scope to further develop models that use new word embeddings and feature transfer techniques. Moreover, deep learning methods should be explored, as well as the use of heterogeneous feature spaces and multiple data sources. Finally, there is an emerging trend of combining word embedding techniques with deep neural learning, which, from the findings published thus far, seem to offer promising results. Great benefits could potentially be derived from investigating, designing, and implementing new models that combine word embedding with deep neural learning.

To conclude, when seeking solutions to the learning problems encountered in cross-domain sentiment analysis, we recommend that researchers:

- Develop methods for learning that consider the structure and nature of the reviews in the different domains from which data comes from (e.g., Twitter, YouTube, and Facebook) to gain a better understanding of the nature of the learning problem to be overcome;
- Develop methods for learning that take into account the distributional similarity (relatedness) between domains;
- Develop methods for learning that focus on improving the representation of pivot features and feature expansion; and
- Develop methods for learning that can overcome the reduced modeling accuracy that occurs due to domain discrepancy and find ways to better predict the best source domains.

REFERENCES

[1] A. Immonen, P. Pääkkönen, and E. Ovaska, "Evaluating the Quality of Social Media Data in Big Data Architecture," *IEEE Access*, vol. 3, pp. 2028–2043, 2015.
[2] D. Jiang, X. Luo, J. Xuan, and Z. Xu, "Sentiment computing for the news event based on the social media big data," *IEEE Access*, vol. 5, pp. 2373–2382, 2016.
[3] M. N. Injadat, F. Salo, and A. B. Nassif, "Data mining techniques in social media: A survey," *Neurocomputing*, vol. 214, pp. 654–670, Nov. 2016.
[4] T. A. A. Al-Moslmi, *Machine Learning and Lexicon-Based Approach for Arabic Sentiment Analysis*. Bangi, Malaysia: Fakulti Teknologi & Sains Maklumat/Institut, 2014.
[5] N. Omar, M. Albared, T. Al-Moslmi, and A. Al-Shabi, "A comparative study of feature selection and machine learning algorithms for arabic sentiment classification," in *Information Retrieval Technology*. Springer, 2014, pp. 429–443.
[6] M. Bouazizi and T. Ohtsuki, "A pattern-based approach for sarcasm detection on twitter," *IEEE Access*, vol. 4, pp. 5477–5488, 2016.
[7] F. Bertola and V. Patti, "Ontology-based affective models to organize artworks in the social semantic Web," *Inf. Process. Manage.*, vol. 52, no. 1, pp. 139–162, 2016.
[8] G. Vinodhini and R. M. Chandrasekaran, "A sampling based sentiment mining approach for e-commerce applications," *Inf. Process. Manag.*, vol. 53, no. 1, pp. 223–236, 2016.
[9] R. Piryani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000âĂŞ2015," *Inf. Process. Manag.*, vol. 53, no. 1, pp. 122–150, 2016.
[10] T. Al-Moslmi, S. Gaber, A. Al-Shabi, M. Albared, and N. Omar, "Feature selection methods effects on machine learning approaches in malay sentiment analysis," in *Proc. 1st ICRIL-Int. Conf. Inno. Sci. Technol. (IICIST)*, 2015, pp. 1–2.
[11] N. Omar, M. Albared, A. Al-Shabi, and T. Al-Moslmi, "Ensemble of classification algorithms for subjectivity and sentiment analysis of arabic customers 'reviews," *Int. J. Adv. Comput. Technol.*, vol. 14, no. 5, pp. 77–85, 2013.
[12] Z. Liu, S. Liu, L. Liu, J. Sun, X. Peng, and T. Wang, "Sentiment recognition of online course reviews using multi-swarm optimization-based selected features," *Neurocomputing*, vol. 185, pp. 11–20, Mar. 2016.
[13] T. Al-Moslmi, M. Albared, A. Al-Shabi, N. Omar, and S. Abdullah, "Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis," *J. Inf. Sci.*, vol. 23, p. 16555, Feb. 2017.
[14] F. Wu, Y. Huang, and Y. Song, "Structured microblog sentiment classification via social context regularization," *Neurocomputing*, vol. 175, pp. 599–609, Jun. 2016.
[15] K. Dashtipour *et al.*, "Multilingual sentiment analysis: State of the art and independent comparison of techniques," *Cognit. Comput.*, vol. 8, no. 4, pp. 1–15, 2016.
[16] S. Keele, "Guidelines for performing systematic literature reviews in software engineering," Dept. Comput. Sci., Univ. f Durham, Durham, U.K., Tech. Rep. Ver.2.3, 2007.
[17] J. Carrillo de Albornoz, L. Plaza, and P. Gervás, "A hybrid approach to emotional sentence polarity and intensity classification," in *Proc. 14th Conf. Comput. Natural Lang. Learn.*, 2010, pp. 153–161.
[18] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," *Decision Support Syst.*, vol. 53, no. 4, pp. 680–688, Nov. 2012.
[19] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in *Proc. 19th Int. Conf. World Wide Web*, vol. 10. 2010, p. 751.
[20] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification," in *Proc. ACL*, vol. 7. 2007, pp. 440–447.
[21] D. Bollegala, T. Mu, and J. Y. Goulermas, "Cross-domain sentiment classification using sentiment sensitive embeddings," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 398–410, Feb. 2016.
[22] J. Liang, K. Zhang, X. Zhou, Y. Hu, J. Tan, and S. Bai, "Leveraging latent sentiment constraint in probabilistic matrix factorization for cross-domain sentiment classification," *Proc. Comput. Sci.*, vol. 80, pp. 366–375, Mar. 2016.
[23] N. X. Bach, V. T. Hai, and T. M. Phuong, "Cross-domain sentiment classification with word embeddings and canonical correlation analysis," in *Proc. 7th Symp. Inf. Commun. Technol.*, 2016, pp. 159–166.
[24] Y. Zhang, X. Hu, P. Li, L. Li, and X. Wu, "Cross-domain sentiment classification-feature divergence, polarity divergence or both?" *Pattern Recognit. Lett.*, vol. 65, pp. 44–50, Jun. 2015.
[25] M. Franco-Salvador, F. L. Cruz, J. A. Troyano, and P. Rosso, "Cross-domain polarity classification using a knowledge-enhanced metaclassifier," *Knowl.-Based Syst.*, vol. 86, pp. 46–56, Jun. 2015.
[26] H. Hammer, A. Yazidi, A. Bai, and P. Engelstad, "Building domain specific sentiment lexicons combining information from many sentiment lexicons and a domain specific corpus," *Comput. Sci. Appl.*, vol. 456, pp. 205–216, Dec. 2015.
[27] G. Zhou, Y. Zhou, X. Guo, X. Tu, and T. He, "Cross-domain sentiment classification via topical correspondence transfer," *Neurocomputing*, vol. 159, pp. 298–305, Dec. 2015.
[28] R. Zhao and K. Mao, "Supervised adaptive-transfer PLSA for cross-domain text classification," in *Proc. IEEE Int. Conf. Data Mining Workshop*, Jan. 2014, pp. 259–266.
[29] C. Lin, Y. Lee, C. Yu, and H. Chen, "Exploring ensemble of models in taxonomy-based cross-domain sentiment classification," in *Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage.-(CIKM)*, 2014, pp. 1279–1288.
[30] Y. Tsai, R. T. Tsai, C. Chueh, and S. Chang, "Cross-domain opinion word identification with query-by-committee active learning," in *Technologies and Applications of Artificial Intelligence*. Cham, Switzerland: Springer, 2014, pp. 334–343.

[31] A. Tsakalidis, "An ensemble model for cross-domain polarity classification on twitter," in *Web Information Systems Engineering*. Cham, Switzerland: Springer, 2014, pp. 168–177.

[32] F. Bisio, P. Gastaldo, C. Peretti, R. Zunino, and E. Cambria, "Data intensive review mining for sentiment classification across heterogeneous domains," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, vol. 13. Aug. 2013, pp. 1061–1067.

[33] P. Yang, W. Gao, Q. Tan, and K.-F. Wong, "A link-bridged topic model for cross-domain document classification," *Inf. Process. Manage.*, vol. 49, no. 6, pp. 1181–1193, Nov. 2013.

[34] D. Bollegala, D. Weir, and J. Carroll, "Cross-domain sentiment classification using a sentiment sensitive thesaurus," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 8, pp. 1719–1731, Aug. 2013.

[35] 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.

[36] Z. Zhu, D. Dai, Y. Ding, J. Qian, and S. Li, "Employing emotion keywords to improve cross-domain sentiment classification," in *Proc. Chin. Lexical Semantics*, 2013, pp. 64–71.

[37] Y. He, C. Lin, W. Gao, and K.-F. Wong, "Dynamic joint sentiment-topic model," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 1, p. 6, 2013.

[38] S. Li, Y. Xue, Z. Wang, and G. Zhou, "Active learning for cross-domain sentiment classification," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 2127–2133.

[39] N. Ponomareva and M. Thelwall, "Semi-supervised vs. cross-domain graphs for sentiment analysis," in *Proc. RANLP*, 2013, pp. 571–578.

[40] B. Ohana, S. J. Delany, and B. Tierney, "A case-based approach to cross domain sentiment classification," in *Lecture Notes in Computer Science*, vol. 7466. Cham, Switzerland: Springer, 2012, pp. 284–296.

[41] R. Remus, "Domain adaptation using domain similarity- and domain complexity-based instance selection for cross-domain sentiment analysis," in *Proc.-12th IEEE Int. Conf. Data Mining Workshops*, Jun. 2012, pp. 717–723.

[42] N. Ponomareva and M. Thelwall, "Biographies or blenders: Which resource is best for cross-domain sentiment analysis?" in *Lecture Notes in Computer Science*, vol. 7181. Cham, Switzerland: Springer, 2012, pp. 488–499.

[43] N. Ponomareva and M. Thelwall, "Do neighbours help?: An exploration of graph-based algorithms for cross-domain sentiment classification," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn.*, Jul. 2012, pp. 655–665.

[44] S. D. Roy, T. Mei, W. Zeng, and S. Li, "SocialTransfer: Cross-domain transfer learning from social streams for media applications," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 649–658.

[45] Q. Wu and S. Tan, "A two-stage framework for cross-domain sentiment classification," *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14269–14275, 2011.

[46] Y. He, C. Lin, and H. Alani, "Automatically extracting polarity-bearing topics for cross-domain sentiment classification," in *Proc. 49th Annu. Meeting*, Jun. 2011, pp. 123–131.

[47] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 513–520.

[48] S. Li and C. Zong, "Multi-domain adaptation for sentiment classification: Using multiple classifier combining methods," in *Proc. Int. Conf. Natural Lang. Process. Knowl. Eng. (NLP-KE)*, 2008, pp. 1–8.

[49] G.-R. Xue, W. Dai, Q. Yang, and Y. Yu, "Topic-bridged PLSA for cross-domain text classification," in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2008, pp. 627–634.

[50] B. Wang, J. Tang, W. Fan, S. Chen, Z. Yang, and Y. Liu, "Heterogeneous cross domain ranking in latent space," in *Proc. 18th ACM Conf. Inf. Knowl. Manage.*, 2009, pp. 987–996.

[51] F. Zhuang *et al.*, "Collaborative dual-PLSA: Mining distinction and commonality across multiple domains for text classification," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 359–368.

[52] K. Lang, "NewsWeeder: Learning to filter netnews," in *Proc. 12th Int. Conf. Mach. Learn.*, 1995, pp. 331–339.

[53] D. D. Lewis. (1997). *Reuters-21578 Text Categorization Test Collection, Distribution 1.0*. [Online]. Available: http://www.research.att.com/lewis/reuters21578.html

[54] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *CS224N Project Rep., Stanford*, vol. 1, no. 2, pp. 12–18, 2009.

[55] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge, "Twitter polarity classification with label propagation over lexical links and the follower graph," in *Proc. 1st Workshop Unsupervised Learn. (NLP)*, 2011, pp. 53–63.

[56] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: A rating regression approach," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 783–792.

[57] A. K. McCallum, K. Nigam, J. Rennie, and K. Seymore, "Automating the construction of Internet portals with machine learning," *Inf. Retr.*, vol. 3, no. 2, pp. 127–163, 2000.

[58] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in *Proc. ACL-Conf. Empirical Methods Natural Lang.*, vol. 10. 2002, pp. 79–86.

[59] S. Baccianella, A. Esuli, and F. Sebastiani, "Multi-facet rating of product reviews," in *Advances in Information Retrieval*. Cham, Switzerland: Springer, 2009, pp. 461–472.

[60] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proc. Int. Conf. Web Search Data Mining*, 2008, pp. 219–230.

[61] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 120–128.

[62] S. Li and C. Zong, "Multi-domain sentiment classification," in *Proc. 46th Annu. Meeting Assoc. Comput. Linguistics Human Lang. Technol., Short Papers*, 2008, pp. 257–260.

[63] H. Guo, H. Zhu, Z. Guo, X. Zhang, X. Wu, and Z. Su, "Domain adaptation with latent semantic association for named entity recognition," in *Proc. Human Lang. Technol., Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2009, pp. 281–289.

[64] P. Wang, C. Domeniconi, and J. Hu, "Using wikipedia for co-clustering based cross-domain text classification," in *Proc. 8th IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2008, pp. 1085–1090.

[65] G. Paltoglou and M. Thelwall, "A study of information retrieval weighting schemes for sentiment analysis," in *Proc. 48th Annu. Meeting Assoc. Comput. Linguistics*, 2010, pp. 1386–1395.

[66] N. Jakob and I. Gurevych, "Extracting opinion targets in a single-and cross-domain setting with conditional random fields," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Oct. 2010, pp. 1035–1045.

[67] F. Huang and A. Yates, "Exploring representation-learning approaches to domain adaptation," in *Proc. Workshop Domain Adaptation Natural Lang. Process.*, 2010, pp. 23–30.

[68] Y. Bao, N. Collier, and A. Datta, "A partially supervised cross-collection topic model for cross-domain text classification," in *Proc. 22nd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2013, pp. 239–248.

[69] D. C. T. Hofmann, "The missing link-a probabilistic model of document content and hypertext connectivity," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 2001, pp. 430–436.

[70] R. Serafin and B. Di Eugenio, "FLSA: Extending latent semantic analysis with features for dialogue act classification," in *Proc. 42nd Annu. Meeting Assoc. Comput. Linguistics*, 2004, p. 692.

[71] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, 1977, pp. 1–38.

[72] T. Li, V. Sindhwani, C. Ding, and Y. Zhang, "Knowledge transformation for cross-domain sentiment classification," in *Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2009, pp. 716–717.

[73] L. Li, X. Jin, and M. Long, "Topic correlation analysis for cross-domain text classification," in *Proc. AAAI*, 2012, p. 12.

[74] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proc. ICML*, vol. 99. 1999, pp. 200–209.

[75] Q. Wu, S. Tan, H. Zhai, G. Zhang, M. Duan, and X. Cheng, "SentiRank: Cross-domain graph ranking for sentiment classification," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol.*, vol. 1. 2009, pp. 309–314.

[76] J. Yu and J. Jiang, "Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 236–246.

[77] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2003, pp. 127–134.

[78] J. Kranjc *et al.*, "Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the ClowdFlows platform," *Inf. Process. Manage.*, vol. 51, no. 2, pp. 187–203, 2015.

[79] B. Settles, *Active Learning Literature Survey*. Madison, WI, USA: Univ. Wisconsin, 2010.

[80] F. Olsson, "A literature survey of active machine learning in the context of natural language processing," in *Proc. SODA*, 2009, p. 36.
[81] D. Nozza, E. Fersini, and E. Messina, "Deep learning and ensemble methods for domain adaptation," in *Proc. IEEE 28th Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2016, pp. 184–189.
[82] M. Long, J. Wang, Y. Cao, J. Sun, and S. Y. Philip, "Deep learning of transferable representation for scalable domain adaptation," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 8, pp. 2027–2040, Feb. 2016.
[83] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, no. 1, pp. 177–196, Jan. 2001.
[84] P. Sanju and T. T. Mirnalinee, "Construction of enhanced sentiment sensitive thesaurus for cross domain sentiment classification using wiktionary," in *Proc. 3rd Int. Conf. Soft Comput. Problem Solving*, 2014, pp. 195–206.
[85] S. M. Jiménez-Zafra *et al.*, "Domain adaptation of polarity lexicon combining term frequency and bootstrapping," in *Proc. NAACL-HLT*, 2016, pp. 137–146.
[86] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI Commun.*, vol. 7, no. 1, pp. 39–59, 1994.
[87] X. Zhu, J. Lafferty, and R. Rosenfeld, *Semi-Supervised Learning With Graphs*. Pittsburgh, PA, USA: Carnegie Mellon Univ., 2005.
[88] X. Zhu and Z. Ghahramani, *Learning From Labeled and Unlabeled Data With Label Propagation*. Seattle, WA, USA: Semantic Scholar, 2002.
[89] A. B. Goldberg and X. Zhu, "Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization," in *Proc. 1st Workshop Graph Based Methods Natural Lang. Process.*, 2006, pp. 45–52.
[90] R. Navigli and S. P. Ponzetto, "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network," *Artif. Intell.*, vol. 193, pp. 217–250, Dec. 2012.
[91] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *Ann. Internal Med.*, vol. 151, no. 4, pp. 264–269, 2009.

TAREQ AL-MOSLMI received the B.Sc. degree (Hons.) in computer science from the University of Science and Technology, Yemen, in 2009, and the master's degree in computer science from the Universiti Kebangsaan Malaysia (UKM), Malaysia, in 2014. He is currently a Researcher with UKM. His research interest falls under natural language processing, machine learning, text and Web mining, and sentiment analysis.

NAZLIA OMAR received the B.Sc. degree (Hons.) from UMIST, U.K, the M.Sc. degree from the University of Liverpool, U.K., and the Ph.D. degree from the University of Ulster, U.K. She is currently an Associate Professor with the Center for AI Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia. Her main research interest is in the area of natural language processing and computational linguistics.

SALWANI ABDULLAH received the B.Sc. degree in computer science from the Universiti Teknologi Malaysia, the master's degree in computer science from the Universiti Kebangsaan Malaysia (UKM), and the Ph.D. degree in computer science with the University of Nottingham, U.K. She is currently a Professor in combinatorial optimization with the Faculty of Information Science and Technology, UKM. Her research interest falls under artificial intelligence and operation research, particularly in meta-heuristic algorithms in optimization areas that involve different real-world applications in static and dynamic optimization problems, such as university timetabling, job shop scheduling, nurse rostering, space allocation, and data mining tasks.

MOHAMMED ALBARED received the B.Sc. degree in computer science from Yarmouk University, Jordan, the master's degree in computer science from Yarmouk University, Jordan, and the Ph.D. degree in computer science from the Universiti Kebangsaan Malaysia, Malaysia. He is currently an Assistant Professor with Sana'a University. His research interest falls under natural language processing, machine learning, text and Web mining, and sentiment analysis.