

Received September 3, 2019, accepted September 21, 2019, date of publication October 3, 2019, date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945340

Challenges and Recommended Solutions in Multi-Source and Multi-Domain Sentiment Analysis

NOR ANIZA ABDULLAH^{ID1}, ALI FEIZOLLAH²,AININ SULAIMAN^{ID2},
AND NOR BADRUL ANUAR^{ID1}

¹Department of Computer System & Technology, Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

²UM Halal Research Center, University of Malaya, Kuala Lumpur 50603, Malaysia

Corresponding author: Nor Aniza Abdullah (noraniza@um.edu.my)

This work was supported in part by the Ministry of Education, Malaysia, through the Malaysian Higher Education Consortium of Halal Institute Grant, and in part by the University of Malaya for Providing the Facilities and Infrastructure to Conduct the Research under Grant MO001-2018.

ABSTRACT The massive availability of online reviews and postings in social media offers invaluable feedback for businesses to make better informed decisions in steering their marketing strategies towards users' interests and preferences. Sentiment analysis is, therefore, essential for determining the public's opinion towards a particular topic, product or service. Traditionally, sentiment analysis is performed on a single data source, for instance, online product reviews or Tweets. However, the need to develop a more precise, and more comprehensive result has steered the move towards performing sentiment analysis on multiple data sources. The use of multiple data sources for a particular domain of interest can increase the amount of datasets needed for training a sentiment classifier. Till now, the problem of insufficient datasets for training the classifier is only addressed by multi-domain sentiment analysis. Aiming to equip researchers with a thorough understanding on both multi-source and multi-domain sentiment analysis, this paper aims to identify the underlying challenges of multi-source and multi-domain sentiment analysis, and discuss the solutions applied by the researchers concerned. This paper also offers an insightful discussion of the findings derived from past studies, and based on these, propose some useful suggestions for the future direction of this research area. Findings derived from our review would be beneficial towards guiding researchers towards the future progress and advancement of multi-source and multi-domain sentiment analysis.

INDEX TERMS Sentiment analysis, multi-domain, multi-source, transfer learning, natural language processing.

I. INTRODUCTION

The rapid rise of the social media has populated the Internet with online reviews, and user generated contents. Internet users post their contents or views onto social media or websites, either for information sharing, or for personal expressions on various topics - from political issues to products, and services. Among the popular social media used as platforms are Facebook, Twitter, and Instagram. Since user-generated contents are usually produced based on the actual experiences of the users, their opinions are often perceived to be genuine, and reliable by the public in general. This is in accordance

The associate editor coordinating the review of this manuscript and approving it for publication was Bin Liu ^{ID}.

to [1] who pointed out that consumers rely on other consumers' experiences to validate the performance of certain products, and services. Previous consumers' experiences, recommendations, ratings, and comments about those products, or services, can influence new consumers' purchase decision. According to [2], online consumers' reviews are more influential than those reviews generated by professionals. This verdict is in line with [3] who said that 90% of the customers' purchase decision depends on online consumers' reviews and comments.

Based on the above, it is deduced that knowing the sentiment polarity of these online reviews, and comments of a product, is therefore, important for businesses. The information can be used to make better-informed decisions so as to

steer their marketing strategies towards users' interests, and preferences, thereby generating more profits [4]. This need has given rise to a new branch of information processing, typically known as sentiment analysis [5]. Reference [6] define sentiment analysis as "the computational process of recognizing and classifying various opinions (thoughts or judgements) expressed in texts". In accordance with the definition, the purpose of sentiment analysis is, thus to analyse online reviews and comments, and to examine their sentiments' polarity so as to obtain the public's opinion on an issue or a product, whether positive, negative or neutral [7].

However, most existing research works on sentiments analysis rely on a single data source, such as online review sites or Twitter. Such datasets can be bias, and also inadequate to represent the whole public's view on a topic, product, or service [4], [7], [8]. This has led to the recent progress in sentiment analysis which has transitioned from the use of single data sources to multiple data sources. The use of multiple data sources for a particular domain of interest can increase the amount of labelled datasets needed for training a sentiment classifier. Insufficient amount of labelled datasets is a common problem faced by sentiment analysis. Currently, the problem is addressed by conducting multi-domain sentiment analysis. In multi-domain sentiment analysis, a resource-rich domain will be used to transfer sentiment knowledge to a resource-limited domain. However, the involvement of different domains in the sentiment analysis process often incur many problems such as feature mismatch, polarity divergence, polysemy, and sparsity, as demonstrated in previous works [9]–[11], [13], [14]. The need to address these problems, has created the need to develop better techniques and solutions for managing sentiment analysis. Worth noting that the upside of using multi-source for sentiment analysis is that it may also be able to tackle the issue of inadequate labelled datasets for training a classifier, a commonly known problem in sentiment analysis when using the machine learning approach [9], [15]–[19].

Until today, review articles which focussed on the challenges faced in multi-source and multi-domain sentiment analysis has been scant. Most currently available reviews tend to include a survey on tasks, approaches, and applications of sentiment analysis, as can be seen in [20]. Thus far, only [21] had focussed on approaches in cross-domain sentiment analysis. Aiming to address this gap, this paper strives to review and examine past studies involving multi-source and multi-domain sentiment analysis. Specifically, this paper focuses on the challenges faced by past researchers and their recommended solutions to circumvent the challenges. The aim for doing this review article is therefore, to equip researchers with a reference or benchmark for future development in multi-source and multi-domain sentiment analysis.

The remainder of this paper is organised as follows. Section 2 gives the background to sentiment analysis, multi-source, and multi-domain sentiment analysis. Section 3 reports on the challenges and recommended solutions noted to ease the problems while Section 4 discusses

the findings. Section 5 suggests some future research directions in this area. Section 6 concludes the article.

II. BACKGROUND

A. SENTIMENT ANALYSIS

Sentiment analysis is the process of analysing text from online reviews or social media postings, with the aim of extracting the public's opinions on various topics [5], and then classifying them into a sentiment polarity which can be positive, negative, or neutral [18], [22]. There are various techniques to do a sentiment analysis. Generally, the techniques can be categorised into the following approaches: lexicon-based, machine learning-based, and a hybrid of both [5].

The lexicon-based approach obtains the sentiment of the analysed text by examining the frequencies, and the polarities of the negative and positive words used in the text [9], [18], [22], [23]. It requires a predefined dictionary of words, such as the SentiWordNet, which is annotated with three sentiment scores - positivity, negativity, and neutrality. The text captured from the social media, and review sites will first be pre-cleaned so as to filter the data, and to extract the relevant stem features. The semantic polarity of the stem will then be calculated, and classified, based on the sentiment scores defined in the sentiment dictionary. The approach behaves well when the texts are well-formed, and are grammatically correct [4]. The lexicon-based approach bears several benefits. First, this approach does not require a labelled training set for classifying the text [24]. Second, the approach defines the lexicons independently of the data, thereby preventing any instance of overfitting. Third, it is often used to perform sentiment analysis on multiple datasets [22]. The approach also has some disadvantages. For instance, the lexicons need to be predefined otherwise, the approach is unable to adapt the lexicon to specific forms of expressions which are associated with formal language. This approach is also unable to detect non-standard abbreviations, which are commonly used in posts published on the social media platforms, such as Twitter [4].

The machine learning-based approach of sentiment analysis caters to a standard classification problem [22]. Therefore, it is more suitable for extracting sentiments from unstructured contents, and less formal texts, such as tweets from Twitter. It also eliminates pre-defined lexicons, thereby providing room for greater flexibility to be applied in any domain [22]. However, as this approach entails the construction of a text classification model to train a sentiment classifier, large labelled training dataset are necessary for the sentiment analysis to be effective [18], [19]. Unfortunately, there are always instances of insufficiently labelled data, and manual annotating is laborious, costly, and time consuming [11], [19]. Moreover, a sentiment classifier trained for a specific domain may not produce accurate result when it is directly applied to different domains [17].

The third approach is the hybrid approach which involves using the lexicon database, and machine learning together.

[7] introduced a hybrid sentiment analysis which employed a machine learning theory, and a method based on polarity lexicon for analysing Chinese sentiment phrases. Another study [24] combined the lexicon-based approach with the machine-learning technique by implementing it on Facebook. The result of its sentiment analysis was found to be highly accurate, at 83.2% precision.

B. MULTI-SOURCE SENTIMENT ANALYSIS

Among the most common data source used for sentiment analysis are online reviews and tweets [22]. In the context of this paper, multiple data sources mean data extracted from more than one sources. Among the common data sources are online product reviews, blogs, e-news, Twitter, Facebook, Instagram, and Google data. Most approaches used in sentiment analysis rely on a single data source. Few have ventured into using multiple data sources [25]. The dependency on a single data source is not healthy because it could lead to biased conclusions. The process of performing a sentiment analysis on datasets obtained from multiple data sources is known as multi-source sentiment analysis or cross-source sentiment analysis [4].

Multi-source or cross-source sentiment analysis has been commercially used for the monitoring of products and brands in social media channels [4]. In [4], a decision support system was developed so as to manage the promotion of products on multiple social media channels (Facebook, Twitter, and Instagram). The support system embodies a social media listener that contains a sentiment analysis engine which monitors multiple social channels. The engine gathers data from the social media users' postings that are related to the promotions and marketing campaigns. It then performs the sentiment analysis of the collected data so as to determine the users' opinions about the campaign.

The importance of performing sentiment analysis on multiple data sources has been emphasised in previous studies [7], [8]. For instance, [7] performed sentiment analysis on a combined dataset comprising online reviews, and tweets so as to develop an improved performance of sentiment predictions. Reference [8] conducted a sentiment analysis of multiple data sources comprising online news, Google search, and Twitter. The aim was to obtain a better event prediction of the stock market. Similarly, [27] also performed a sentiment analysis on online news, tweets from Twitter, and quantitative historical data so as to better predict the stock market. In their study, [28] noted that event detections using multiple data sources were more meaningful, and they also bore valuable outcomes. All these aforementioned studies have provided concrete evidence which proved that using multi-source or cross-source sentiment analysis provides a better prediction outcome than using just a single data source.

C. MULTI-DOMAIN SENTIMENT ANALYSIS

Sentiment analysis is a domain-dependent process [17] because different domains express sentiments differently [2], [9], [15], [18], [19]. For instance, the word 'easy' in

the Electronics domain is normally positive, but in the Books domain, it carries a negative sentiment. The easiest way to address this problem is thus, to prepare enough labelled datasets for each domain for training a classifier [2]. The disadvantage of this method is that it is costly to manually label sufficient samples in every possible domain of interest [2], [19]. Moreover, a domain dependent classifier also requires a large amount of labelled datasets in order to be accurate [7], [15], [17], [22].

A better solution to resolve this issue is by considering resource-rich labelled data from other domains, and by designing a more robust sentiment classifier which can work across different domains [15], [17], [19], [22]. This scenario is called the multi-domain or cross-domain sentiment analysis [19], [21], [22]. In [22], the cross-domain sentiment analysis was described as beneficial when labelled datasets were difficult to obtain.

A multi-domain or cross-domain sentiment analysis is normally performed when there are similarities in features between the two domains [7]. In a typical domain adaptation process, the sentiment knowledge from a domain with sufficient annotated data (i.e., source domain) is transferred to a new domain with limited, or no labelled data (i.e., target domain) [15], [17]. Prior to the knowledge transfer, analysis of the similarity between the source domain, and the target domain must be performed so as to reduce the differences in the feature distributions between the domains [17].

In the following section, the findings of the review on challenges in multi-source and multi-domain sentiment analysis, followed by the recommended solutions is further elaborated.

III. CHALLENGES AND SOLUTIONS

With the knowledge one has about the public's sentiment towards a product or service, one can then make a decision on whether to purchase or not to purchase a product or service. The polarity of the sentiment can be obtained by processing, and analysing the public's sentiments gathered from online reviews, and social media. Literature [4], [7], [8], [26], [27] has shown that sentiment analysis resulting from multiple data sources tend to produce a more concrete prediction. Reference [4] highlighted that the use of multiple data sources has generated an overwhelming amount of data for analysis, hence it is quite demanding to identify the relevant opinions from the irrelevant ones. They also pointed out that data coming from various sources also display different characteristics. This is another challenge that needs to be properly addressed [4], [28]. Multi-domain sentiment analysis, on the other hand, focuses more on the challenges of adapting to the different domains [15], [17]–[19], [26], and in enabling computers to process issues related to the natural language which is embodied in those different domains as demonstrated in [9]–[15]. These challenges linked to natural language processing (NLP), as stated by Al-Moslmi *et al.* [21], encompass feature divergence, polarity divergence, sparsity, and polysemy. Based on this understanding, this review paper will present a classification of the basic challenges

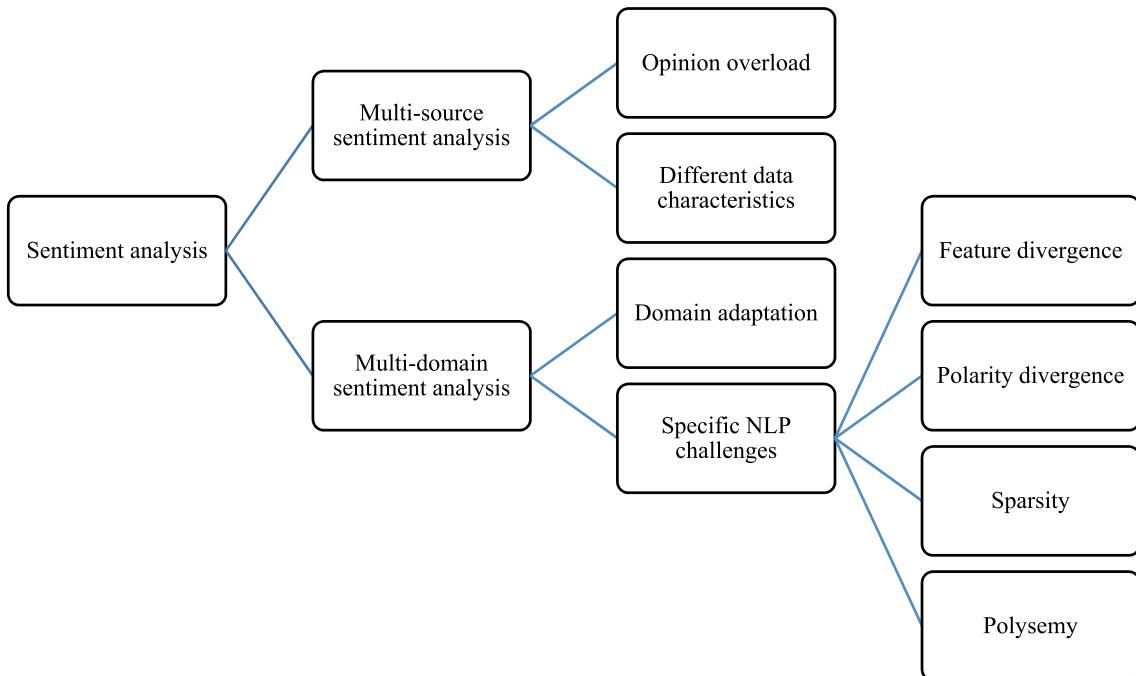


FIGURE 1. Classification of challenges in multi-source and multi-domain sentiment analysis.

noted in multi-source and multi-domain sentiment analysis. Figure 1 illustrates.

A. MULTI-SOURCE SENTIMENT ANALYSIS

1) OPINION OVERLOAD

When multiple data sources are used in sentiment analysis, the amount of data that needs to be processed, and analysed also increases. This can be an overwhelming, and difficult process [4]. Therefore, it is crucial to extract only the relevant opinions from the massive online reviews, or social media posts. This means that the extracted reviews or posts coming from multiple data sources must be syntactically bound by a common subject of interest.

The traditional way to detect a common subject of interest from a pool of information is by describing a topic as a set of keywords. These are mostly composed of verbs, or nouns. A common way to detect a topic in a tweet, for example, is via hashtags. Reference [29] used tweets with hashtags, and emoticons, and a non-parametric supervised model which consists of multiple features, such as the tweet volume feature, and sentiment variation features, to detect trending topics on Twitter. They used the MapReduce to accelerate the hashtag extractions, and their analysis. The correct interpretation of the hashtag makes it possible to link the related contents gathered from other sources, such as online product review websites, or online news articles. On the other hand, [40] evaluated the data mining techniques which were used to detect topics, or events from the input data streams of several online review websites. The techniques were divided into supervised learning, and

unsupervised learning approaches. The results indicated that keyword clustering provided the best accuracy for detecting topics, when compared to other approaches.

In another study [28], a microblog theme crawler named Sina was used to find the associated microblogs. The crawler extracts events related microblogs by searching for the events' descriptions from online news. The text similarity between the microblogs and the events' descriptions is then calculated. This helps to remove the irrelevant microblogs data. Reference [8] used three different data sources to predict the public's opinions of the stock market. Their data sources were Bloomberg's online news article, Google search data, and Twitter. The researchers first determined the relevant economic topics from the news articles by modelling the topic of interest in the news by using the Latent Dirichlet Allocation (LDA) algorithm. Upon the execution of the LDA, 30 most interesting topics were identified. Each topic was then manually analysed so as to select the topics that were most relevant. These were traced to the finance and economic domains. News articles which contained any of the selected topics were then considered as relevant. A natural language processing tool, and a customised economics feature dictionary were then used to derive the sentiment of the selected news articles. Data derived from the Google search volumes were then processed with Lasso regression. The aim was to select the most informative features that represented the public's sentiment towards certain stock indices. The next step was to detect burst events on the tweeter data so as to select the features that represent any dramatic movement in society with regards to financial market concerns.

2) DIFFERENT DATA CHARACTERISTICS

Multi-source sentiment analysis contains various characteristics of the data which come from different data sources, such as online reviews, online news, and social media. For instance, Twitter messages are short and noisy, and they use casual language that often contains massive user-invented acronyms, emoji expressions, users' opinions, and sarcasms [18], [28]. Each tweet contains only 140 characters, or less when compared to other media, such as Facebook or online reviews. In contrast, online reviews are lengthy, use more formal language, and appear more authoritative [28]. This was verified by [30] who explained that data from different sources have different feature distributions, therefore, simply merging the data from these sources may not yield optimal results [30]. This was also iterated by [28] who stressed that the differences in data characteristics must be properly addressed when merging users' opinions, or comments gathered from multiple data sources.

The review of literature conducted for this paper noted that there were several ways to address this challenge. For instance, in order to obtain a warning information for an event, [28] used the Analytic Hierarchy Process (AHP) to analyse information and sentiments extracted from the news articles, and microblogs on food safety, and weigh each indicators. These indicators were used to represent, or to describe the target event, such as the food safety event. The AHP is a technique that is used for structuring related information so as to analyse the complex problems, prior to making decisions [31]. The weighted sum of these indicators would imply the event's level of importance and its sentiment. If the event was identified as very important, and its sentiment value was negative, then an alert on the possible issue on food safety would be delivered to the target users.

In [8], a data integration and prediction model, based on Delta Naïve Bayes, was proposed for predicting financial market. They combined features from Google search, Bloomberg news, and Twitter together. The Google search volumes were used to obtain the public's sentiment towards a particular stock index, and the futures of the financial market. The online financial news articles were used to determine the public's opinions on the stock market indices in various countries. Analysis was performed through natural language processing, using a customized economic feature dictionary. Following this, recent dramatic societal movements of the stock market have been determined through the analysis of Twitter bursts, and contents. Each feature was assumed independent of the other, and that each feature must be processed separately, and normalized, before each can be combined to determine the prediction. In another study that also aimed to predict stock markets, [27] proposed the Multi-source Multiple Instance framework as a means to integrate the historical trading data, investors' sentiment, with the economic news events that were extracted from three separate sources. The framework extended the use of the Multiple Instance Learning paradigm, based on the supervised learning made on groups of instances.

A simpler approach to process and merge data from multiple sources was conducted by [4], and [7]. Reference [4] built a uniform data structure to represent the users' posts, and comments which were extracted from multiple social media networks. The data structure was made up of a string of texts, and a set of features that represent the users, and their interactions with the posted comments. Data collected through the uniform data structure were then pre-processed, according to the NLP procedure. Following this, [7] created a training dataset of instances which were obtained from a combination of tweets (the sentiment140 tweet corpus), and online product reviews (i.e. Amazon reviews datasets). A classifier was then trained on the combined datasets.

B. MULTI-DOMAIN SENTIMENT ANALYSIS

1) DOMAIN ADAPTATION

Domain adaptation is the process of training a sentiment classifier for one domain by using information taken from another similar domain [11]. The main purpose for doing this is to provide sufficient datasets for the trained sentiment classifier to deliver accurate analysis. Even though domains involved in the adaptation process shared some similar features, discrepancy among them may still occur due to the variety of expressions used to convey opinions. Therefore, the process of adapting sentiment knowledge from two different domains is a non-trivial and challenging task. Failure to correctly adapt the knowledge would result in a negative transfer [19]. In that regard, domain adaptation is presented as one of the challenges in multi-domain sentiment analysis. As a result of this, our literature review can focus on the existing solutions used in adapting the sentiment knowledge obtained from different domains. To aid understanding of the existing solution for domain adaptations, we thus categorised them based on the techniques that they used.

a: TRANSFER LEARNING

In the process of domain adaptation, there must exist at least one source domain, and one target domain [9], [15], [17]. The source domain is assumed to contain sufficient amounts of labelled datasets containing the sentiment knowledge as compared to the target domain. With this assumption, a transfer learning model can be developed. This outlines a way for the sentiment knowledge to be transferred from the source domain to the target domain. Therefore, a domain adaptation method which is based on the transfer learning approach must seek to identify features which are shared between the domains. This helps to bridge the source domain with the target domain.

Among the earliest domain adaptation methods used is structural correspondence learning (SCL) by [15]. The SCL uses pivot features on unlabelled datasets from both domains to represent words that occur frequently, and behaved similarly in the two domains. By modelling their correlations with the non-pivot features from both domains, the

correspondence features among the pivot and non-pivot features can be deduced for training a classifier. For their experiment, Blitzer *et al.* [15] produced a benchmark dataset derived from online product reviews extracted from Amazon in the following domains: Books, DVDs, Electronics, and Kitchen appliances. Their result showed that the SCL outperformed other methods which were based on supervised and semi-supervised learning. Another notable work in this area was that of [16] who proposed spectral feature alignment (SFA) algorithms. The SFA aims to reduce the discrepancy between a source domain and a target domain by aligning the domain-specific and domain-independent features of both domains. The co-occurrence between these features was then modelled through a bipartite graph so as to produce clusters of connected features, for training a classifier. The proposed SFA outperformed other baseline methods during the experiment involving real-world datasets: the datasets by [15], the Amazon datasets on video games, electronics and software, Yelp dataset (hotel reviews), and Citysearch dataset (hotel reviews).

In [12], an algorithm called topical correspondence transfer (TCT) was proposed. This approach worked on two fundamental assumptions: a) some shared topics exist between the source domain and the target domain, and b) every domain must have domain-specific topics. Therefore, each review document must be represented by one matrix of shared topics, and another matrix of domain-specific topics. This makes it easier to cross reference the matrices when matching similar sentiment features from the source domain and the target domain. The TCT is different from the SCL, and the SFA in that it utilizes the correspondence between shared topics, and domain-specific topics for both domains. Reference [12] tested their method on the Amazon reviews datasets which were produced by [15]. They found that their method outperformed other similar methods used in cross domain sentiment analysis.

Most of the earlier methods used in domain adaptations mainly transferred sentiment knowledge from one source domain to a target domain [9], [17]. Even though the methods were proven successful, their adaptation performance tended to decline when there was a significant difference in distribution of features between both domains [9], [17]. In order to reduce this problem, multiple source domains were used in multi-domain sentiment analysis. For instance, [9] used labelled datasets from several source domains, and unlabelled datasets from both the source domain and the target domain, to create a sentiment sensitive thesaurus. The thesaurus contains different words which express the same sentiment for different domains. To reduce the differences in the feature distributions between both the source and target domains, related words from the thesaurus were used as additional features to represent both domains. The proposed method yielded an average accuracy of 80.9% when experimented on the datasets produced by [15]. It outperformed both the SCL, and the SFA, in many application domains.

b: MULTI-TASK LEARNING

Most multi-domain adaptation methods tend to build a transfer learning model so as to adapt the polarity of the terms' sentiments derived from the different domains. Even though the methods were able to adapt the relevant sentiment features between different domains, the transfer learning approach imposes the necessity to build a new transfer model, each time a new domain needs to be analysed. This limits its generalization's capability [14], [36].

Therefore, instead of using the transfer learning approach to adapt a source domain to a target domain, [33], [17] applied the multi-task learning approach so that both the general- and domain-specific sentiment knowledge from several source domains, can be learned simultaneously, without the need for a transfer learning model. They thus proposed a classifier which comprised a general, and a domain-specific classifier. The former captures common sentiment knowledge in the different domains while the latter captures the domain-specific sentiment knowledge of every domain. A domain similarity graph was then used by the domain specific classifier to determine the relatedness among the domains. The graph was built based on the terms' distribution (similarity in textual content), and the sentiment words' distribution (similarity in sentiment words) between each domain. The result of the case study in [33], [17] denoted that the proposed classifier can accurately capture global sentiment words as well as maintain the consistency of their sentiment polarities in different domains. The experimental result of the Amazon datasets [15], and Sanders' Twitter Sentiment datasets showed that approach by [33], [17] had significantly outperformed other baseline multi-task learning method.

In [19] a domain attention model was proposed, based on a multi-task learning model. The domain attention model uses neural network to simultaneously compute individual parts of a sentence so as to produce an output that can determine the most discriminative feature in a review text. The model has a domain module and a sentiment module. The domain module predicts which domain a word belongs to. As it works on multiple domains, it can also identify the common features, and the domain-specific features for each domain. The sentiment relatedness of the features between the different domains was then determined by the sentiment module. Their experimental results derived from each of the Amazon datasets produced by [15] showed that their approach had outperformed all the other baseline methods.

c: WORD EMBEDDINGS

Another attempt to address the limitation of using a transfer learning approach is traced to the works of [14]. Instead of using the transfer learning model, they proposed a neural word embeddings approach which exploits information overlaps between domains, and words were mapped to vectors. A deep learning approach was then used to classify the word vectors into two outputs. The first output represents the overlap degree between the word vector sequence of a

review document, and the domain itself. The second output represents the polarity value of a review document with regards to each domain. The proposed approach was then implemented in the NeuroSent. The tool used the Dranziera datasets of online reviews obtained from twenty different domains, known as in-model domains (IMD), for training and testing the classifier. Another set of online datasets were derived from seven different domains called the out-model domains (OMD). These were mainly deployed for testing. The experimental evaluation of the NeuroSent on both the IMD and OMD datasets showed that it outperformed most of the baseline methods, including the Support Vector Machine, the Naive Bayes, and Maximum Entropy.

Further, [36] used fuzzy logic to address any uncertainties when assigning polarity values of concepts belonging to different domains. Reference [34], likewise, addressed the domain adaptation challenge by using word embeddings. They implemented canonical correlation analysis (CCA) so as to establish the correlations between the domain-independent words with the domain-dependent words. These were derived from the source domain, and the target domain. A learning method was then applied on the combination of word embedding features, the CCA features, and the raw features so as to produce a classification model. Their proposed classification model yielded an average accuracy of 77.8% when experimented on the datasets produced by [15].

d: SENTIMENT LEXICONS

Other interesting efforts which employed the lexicon-based approach to address the challenge of adapting different domains came from [35], [23], and [2]. Reference [35] supported domain adaptation by creating the ontological lexicon to achieve the contextualized, cross-domain lexicons. The datasets used were from online product reviews obtained from Amazon, hotel reviews obtained from TripAdvisor.com, and movie reviews obtained from the Internet Movie Database. Their results showed that a contextualized lexicon that is trained on three different domains yielded better results than that trained on a single domain [35]. Their result also verified that contextual lexicons enhanced the performance of lexicon-based sentiment analysis.

Additionally, the combination of the contextualized lexicon, and the terms' knowledge from the WordNet, was also able to classify ambiguous terms in online reviews into positive or negative polarities. In another study, [23] adapted the differences among the different domains by integrating several sentiment dictionaries, such as WordNet, SentiWordNet, SentiNet, SentiSense and Opinion Lexicon. They also proposed an algorithm to remove and/or shift the polarity of a sentiment word. The datasets used for the experiment were online product reviews from Amazon.com, including the Smartphones, Movies, and Books domains. The experimental result denoted an accuracy of 82.6%, 80.1%, and 81.8% respectively, for Smartphones, Movies, and books. In another study, [2] addressed domain differences on online product reviews in the Hindi language by creating a sentiment

aware dictionary for the language. The datasets by [15] were also used in their experiment. Their results showed that the dictionary can accurately classify the unlabelled and unseen reviews into positive and negative polarities. In the works of [18], the general-purpose sentiment lexicons was proposed for domain adaptation. The justification for using the general-purpose sentiment lexicon was to reduce the dependency on the labelled datasets taken from multiple source domains.

The general-purpose sentiment lexicon also has a better generalization capability when compared to the sentiment classifier trained in a source domain [18]. This general-purpose sentiment lexicon was developed based on the multi-level contextual sentiment relations that came from unlabelled datasets of a target domain. The sentiment knowledge, and polarity relations were derived from the target-, phrase-, sentence-, and document-levels. The extracted sentiment knowledge, and relations were then used to train the classifier. The Amazon datasets produced by [15] were likewise, used in their experiment. Their results depicted an improvement in the sentiment domain adaptation performance as compared to other baseline methods.

Similarly, [26] proposed a domain adaptation method that combined the sentiment knowledge obtained from the general-purpose sentiment lexicon, the labelled data of the target domain, the sentiment knowledge of several source domains, and the domain-specific sentiment relations. This sentiment knowledge was able to produce an improved domain adaptation result when evaluated using the [15].

2) SPECIFIC NATURAL LANGUAGE PROCESSING CHALLENGE

In performing sentiment analysis across multiple domains, challenges related to specific natural language processing (NLP) such as feature divergence, polarity divergence, sparsity, and polysemy, can influence the analysis [21]. Some of these challenges may also be encountered in multi-source sentiment analysis. The following sub-sections present the approaches observed in the literature review which specifically address each of the challenges noted.

a: FEATURE DIVERGENCE

Feature divergence refers to the mismatch in sentiment polarity values between the domain-specific features of different domains [21]. In the Electronics domain, for example, words such as “compact” and “sharp” were used to signify positive sentiments while the word, “blurry” signifies a negative sentiment. However, in the Video Games domain, words like “hooked”, and “realistic” were used to signify positive sentiments whereas “boring” was used to signify a negative sentiment [16]. When a comparison of these domain-specific features was made, there was no similarity between both domains, hence the data distribution between both domains was very different [11]. This means that words which appeared in the reviews of the target domain do not always appear in the trained model [9], thereby resulting in the poor performance of the sentiment classifier when it was applied to a domain that was unlike the one it was trained in.

Therefore, most research looking at cross-domain sentiment analysis tended to focus on addressing problems which were associated with feature divergence [9], [11], [15], [16]. Reference [9] tackled this problem by building a sentiment sensitive thesaurus that can identify the connections among the words in different domains. Another approach used new feature representations where the independent features were regarded as a bridge between the source domain, and the target domain. This has been noted in the work by [11], [15], and [16].

b: POLARITY DIVERGENCE

Polarity divergence represents a situation in which some features are positive in one domain, but are negative in another domain [11]. As an example, “easy” was a common word used in the Electronics domain to signify a positive sentiment, “*this smartphone is easy to use.*” However, “easy” was frequently used in the Movies domain to represent a negative sentiment, “*the ending of this movie is easy to guess*” [11]. It is therefore, a challenge to infer the polarity of texts belonging to domains which are different from those used for building the classification model [36].

To address polarity divergence, [10] proposed an ensemble model which combines sentiment analysis outputs from various algorithms, such as the combination of hybrid machine-learning classification approaches with the lexicon-based approach, in a weighted scheme, so as to classify tweets. The proposed ensemble model can achieve an average increase of 10.22% of accuracy over the main baseline model. Reference [11] then proposed an algorithm that transfers the polarity of features from a source domain to a target domain by using the independent features as a bridge. Independent features are features that are predictive in both domains even though they may have different polarities. Reference [13] addressed polarity divergence by considering the polarity strengths of the features rather than their absolute positivity or negativity. Their method consisted of two parts. First, they combined the score of the features taken from the sentiment lexicons. Then they used this score data to predict the domain-specific lexicon score. Similarly, [7] also considered the polarity strengths of the features. However, the strength was measured through the probability value instead.

On the other hand, [36] tackled polarity divergence with an approach that computed the polarity of texts which belonged to the domains that were different from the one used to train the classifier. The approach first captured the linguistic overlaps that occurred between the domains by using word embeddings, and deep learning. It then resolved any uncertainties in the sentiment polarities between different domains through the fuzzy logic.

c: SPARSITY

Sparsity refers to a situation where the target domain contains words or phrases that do not, or rarely appear in the source domain [21]. Sparsity can reduce the performance of the domain dependent sentiment classifier [9].

One way to overcome sparsity is to use a feature expansion method, and sensitive thesaurus, whereby a feature vector is augmented with additional related features taken from a sentiment-sensitive thesaurus [9]. In [37], the effect of data sparsity in the Hindi language was minimised by using bilingual word embeddings, and the deep learning approach. By applying the bilingual word embeddings on the English-Hindi and English-French language pairs, the language barrier between a resource-rich, and a resource-poor language is bridged in the shared vector space. Any hidden feature between the two languages was then learned through the deep learning approach. Their experimental results on datasets on Restaurant and Laptop domains, shows that their approach outperforms other approaches for multi-linguality and cross-linguality sentiment analysis. Noteworthy that the work by [37] is an example of multi-linguality and cross-linguality in sentiment analysis.

d: POLYSEMY

Polysemy is the coexistence of many possible meanings for a word, both in the source and the target domains due to the context of the respective domains [21]. For instance, the word, “lie” means put oneself in a resting position in the Exercise domain whereas “lie” means making an untrue statement in the Politics domain. To the best of our knowledge, not many studies have addressed the issue of polysemy in cross-domain sentiment analysis. The only work we detected was performed by [38], who attempted to tackle polysemy by using the meta-learning approach which combines different classical classifiers with knowledge-based classifiers. The knowledge-based classifiers were word sense disambiguation (WSD) based classifier, and vocabulary expansion-based classifier. The WSD-based classifiers were trained by using knowledge graphs which were built when implementing the WSD on the BabelNet, a multilingual semantic network [39]. This network classified a review document into sets of disambiguated words (nouns, adjectives, verbs and adverbs). These disambiguated words were then used by the vocabulary expansion-based classifier to identify the semantically-related concepts, so as to expand the vocabulary. It also used the BabelNet knowledge graph to facilitate the process.

IV. DISCUSSION

There is a growing interest among researchers to embark on multi-source or multi-domain sentiment analysis for an improved conclusion, and sentiment outcome. In this study, we reviewed and analysed 26 studies which were related to multi-source and multi-domain sentiment analysis. A quick identification of these studies is projected in Table 1 while Table 3 and Table 4 provide a summary of the reviewed studies, highlighting their approaches in recommending solutions for the identified challenges. To aid understanding of the recommended solutions, the approaches were further categorised according to the techniques used. These findings are

TABLE 1. Quick identification of multi-source and multi domain sentiment analysis challenges addressed by current research.

Author	Year	Multi source	Multi domain	Opinion Overload	Different Data Char.	Domain Adapt.	NLP Challenge			
							FD	PD	S	P
[4]	2019	✓			✓					
[19]	2018		✓			✓	✓			
[27]	2018	✓			✓					
[37]	2018		✓			✓				✓
[2]	2017		✓		✓	✓	✓	✓	✓	
[8]	2017	✓		✓						
[36]	2017		✓			✓		✓		
[26]	2016		✓		✓					
[7]	2016	✓			✓					
[40]	2016	✓		✓						
[14]	2016		✓			✓	✓	✓		
[17]	2016		✓			✓	✓			
[18]	2016		✓			✓	✓			
[34]	2016		✓			✓	✓			
[11]	2015		✓			✓	✓	✓		
[12]	2015		✓			✓	✓			
[13]	2015		✓			✓		✓		
[28]	2015	✓		✓	✓					
[33]	2015		✓			✓	✓			
[38]	2015		✓			✓				✓
[23]	2014		✓			✓	✓	✓		
[10]	2014		✓			✓		✓		
[35]	2013		✓	✓		✓	✓	✓		
[9]	2012		✓			✓	✓			✓
[16]	2010		✓			✓	✓			
[15]	2006		✓			✓	✓			

Note: FD: Feature Divergence, PD: Polarity Divergence, S: Sparsity, P: Polysemy

further elaborated so as to feature the contributions of the reviews.

A. PROGRESS IN MULTI-SOURCE AND MULTI-DOMAIN SENTIMENT ANALYSIS

We had reviewed 26 recent studies related to multi-source and multi-domain sentiment analysis, commencing year 2013 onwards until the present, with the exception of notable studies in cross domains sentiment analysis, such as [14], [10] and [9]. Table 1 shows that there is more research work done in the multi-domain sentiment analysis (20 studies) as compared to the multi-source sentiment analysis (6 studies). The results also showed that there were consistent research efforts made in the multi-domain sentiment analysis throughout the years. However, studies in sentiment analysis which used multiple data sources had only started to grow in recent years. This may be attributed to the widespread use of social media among the public.

An in-depth look into those studies (refer to Tables 3 and 4) showed that most of the data source were traced to online reviews and Tweets. Noteworthy from the 26 studies that we reviewed was the fact that only one study by [4]

had included Facebook, and Instagram posts for sentiment analysis. Majority of the studies in multi-domain sentiment analysis had used the Amazon product reviews datasets produced by [14]. The recent progress in multi-source sentiment analysis shows a growing trend among researchers in using other sources of datasets, such as economics online news, Google search volume data, and historical quantitative data.

As the datasets of [14] had been commonly used in most studies, we can thus infer that the most analysed domains used for evaluating public sentiments were the Books, DVDs, Electronics, and Kitchen appliances domains. During the domain adaptation process, multi-domain sentiment analysis which used a single data source (e.g. Amazon online product reviews) can use either one domain (e.g. Book), or more than one domain (e.g. Book and Movies) as its source domain. Initial works in the multi-domain sentiment analysis, such as [14], and [10] had used a single domain as the source domain. However, from 2013 onwards, it was observed that majority of the researchers in this area were using multiple domains as their source domain. This effort was initiated by [9].

TABLE 2. Advantages and disadvantages of the recommended solutions for domain adaptation challenges.

Approaches	Advantages	Disadvantages
Transfer learning model	A well-established approach for cross domains sentiment analysis.	Require a new transfer model for every new domain, hence limits generalization.
Multi-task learning	It does not require to build a model prior to analysing new domain.	If the number of tasks is very huge, optimization algorithm is required to speed up the learning process.
Word embeddings	It provides explicit representation of words in regards to the context they appear, in which words with common contexts appear near to one another in the vector space.	The word-to-vector representation does not take into account the semantic meaning of words.
Sentiment lexicon	Does not need to build a transfer model prior to analysing any new domain.	A predefined set of sentiment lexicon is required.

TABLE 3. Summary of multi-source sentiment analysis' challenges, recommended solutions, and datasets used.

Challenge	Solution Category	Ref.	Approach	Data Source	Dataset	Domain
Opinion overload	Event detection	[28]	Introduce Sina, a microblog theme crawler to find associated microblogs based on event in online news	Online news, microblog	China Online news and microblogs	Food safety
	Topic detection	[8]	Use LDA, a topic modelling algorithm to identify 30 most interesting topics from the news articles. Use burst detection and burst event grouping on Tweeter data to identify dramatic social movement and emotion.	Online news, tweeter, Google search trends	Bloomberg news, Google data, tweets	Financial market
	Topic detection	[40]	Use unsupervised learning and supervised learning approach to detect predefined and unknown topics from unstructured user reviews	Online reviews from multiple hotel sites	Swedish Copperhill Mountain Lodge and Holiday Club Åre	Hotel
Different data characteristics	Multi-source data integration model	[28]	Used Analytic Hierarchy Process (AHP) to combine features/indicators from multiple data sources	Online news, microblog	China Online news and microblog	Food safety
	Multi-source data integration model	[8]	Create Delta Naïve Bayes model to combine features from multiple data sources for predicting financial market	Online news, tweeter, Google search trends	Bloomberg news, Google data, tweets	Financial market
	Multi-source data integration model	[27]	Extend Multiple Instance Learning (MIL) with Multisource capability for predicting stock market index movement.	China financial info service provider, social media	Historical quantitative data and news from Wind, and posts from Xueqiu (investor social network in China)	Stock market
	Combined training datasets	[7]	Create a training dataset that combine instances from both data source	Online reviews, tweeter	Amazon review, sentiment140 corpus and combined	No specific domain
	Uniform data structure	[4]	Represent posts and comments in a uniform data structure that comprises of text string and features.	Tweeter, Facebook and Instagram	Tweeter, Facebook and Instagram	Restaurant/ Electronic

A further analysis of the reviews, as shown in Table 3, provided insights into the evolution of the research done in sentiment analysis, i.e., from just focussing on finding the best technique to producing accurate sentiment result, to focusing on finding the best approach to implement

multi-source sentiment analysis for predicting future events. This trend was also noted in previous works [8] and [27] which had exploited sentiment analysis for predicting future trends in financial stock markets. Additionally, in [28], the contribution of sentiment knowledge was derived from

TABLE 4. Summary of multi-domain sentiment analysis' challenges, recommended solutions, and datasets used.

Challenge	Solution Category	Ref.	Approach	Data Source	Dataset	Domain
Domain Adaptation	Transfer learning	[15]	Introduce SCL to determine features corresponding between a source domain and a target domain.	Online reviews	Amazon product reviews (Blitzer datasets)	Book, DVDs, electronics, kitchen appliances
	Transfer learning	[16]	Introduce SFA to reduce discrepancy between a source domain and a target domain by exploiting co-occurrence relationship between the domain-specific and domain independent features.	Online reviews	Amazon product reviews Yelp and Citysearch hotel reviews	Book, DVDs, electronics, kitchen appliances; video game, electronic and software
	Transfer learning	[12]	Proposed the TCT to exploit the relationship between domain-specific and shared topics among domains.	Online reviews	Amazon product reviews (Blitzer datasets)	Book, DVDs, electronics, kitchen appliances
	Transfer learning	[9]	Create a sentiment sensitive thesaurus for cross-domain sentiment classification of user reviews.	Online reviews	Amazon product reviews (Blitzer datasets)	Book, DVDs, electronics, kitchen appliances
	Multi task learning	[17]	Use multi-task learning to extract the global and domain-specific sentiment models from the multiple source domains.	Online reviews	Amazon product reviews (Blitzer datasets),	Book, DVD, Electronics and Kitchen
	Multi-task learning	[19]	Introduce the domain attention model based on a multi-task learning approach and neural network.	Online reviews,	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen
	Multi-task learning	[33]	Use multi-task learning to learn the general- and domain-specific sentiment knowledge of multi-domains simultaneously.	Online reviews, Twitter	Amazon product reviews (Blitzer datasets), Sanders Twitter Sentiment Dataset	Books, DVD, Electronics and Kitchen Tweets related to Apple, Google, Microsoft, Tweeter companies
	Words embeddings	[14]	Introduce an approach based on word embeddings with deep learning.	Online reviews	Dranziera datasets (includes Amazon Product reviews)	IMD: Amazon Instant Video, Automotive, Baby, Beauty, Books, Clothing Accessories, Electronics, Health, Home Kitchen, Movies TV, Music, Office Products, Patio, Pet Supplies, Shoes, Software, Sports Outdoors, Tools Home Improvement, Toys Games and Video Games. OMD: Cell Phones Accessories, Gourmet Foods, Industrial Scientific, Jewelry, Kindle Store, Musical Instruments, and Watches
	Words embeddings	[34]	Use word embeddings and canonical correlation analysis.	Online reviews	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen
	Words embeddings	[36]	Introduce an approach based on word embeddings with deep learning, and fuzzy logic is used to remove uncertainties in polarity sentiment	Online reviews	Dranziera datasets (includes Amazon Product reviews)	IMD: same with the above OMD: same with the above
Sentiment knowledge/ontological lexicon	Sentiment knowledge/ontological lexicon	[35]	Create a contextualized, cross-domain lexicon using ontology. The approach detects ambiguous sentiment terms, collects context terms for each, and then uses these context terms to improve sentiment analysis.	Online reviews	Amazon product reviews/ TripAdvisor/ movie review corpus	Products, hotels, movies
	Sentiment knowledge/thesaurus	[9]	Create a sentiment sensitive thesaurus for cross domains sentiment classifications. It uses labelled and unlabelled data from multiple source domains, and unlabelled data from a target domain.	Online reviews	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen
	Sentiment knowledge/dictionaries	[23]	Integrate multiple sentiment dictionaries for cross domains sentiment classifications	Online reviews	Amazon product reviews	Smartphones, movies and books

TABLE 4. (Continued.) Summary of multi-domain sentiment analysis' challenges, recommended solutions, and datasets used.

	Sentiment knowledge/dictionary	[2]	Create a sentiment aware dictionary using labelled data from multiple source domains and unlabelled data from both source and target domains. The dictionary can label the unlabelled reviews from the target domain.	Online reviews	Amazon product reviews	Kitchen Appliances, DVDs, Electronics and Books.
	Sentiment knowledge	[26]	Perform domain adaptation based on collaboration of four kinds of sentiment knowledge i.e. general sentiment knowledge in sentiment lexicons, labelled data in target domain, sentiment information in multiple source domains, and domain-specific sentiment relations among words mined from the unlabelled data of target domain	Online reviews	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen
	Transfer learning	[18]	Built and use a general purpose sentiment lexicon based on multi-level contextual sentiment relations, for adapting different domains	Online reviews	Amazon product reviews (Blitzer datasets),	Book, DVD, Electronics and Kitchen
Specific NLP Challenge	Feature divergence	[9]	Address feature divergence problem by using a thesaurus to group different features expressing the same sentiment.	Online reviews	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen
	Feature divergence	[16]	Use SFA algorithm to address feature divergence between a source domain and a target domain.	Online reviews	Amazon product reviews Yelp and Citysearch hotel reviews	Book, DVDs, electronics, kitchen appliances; video game, electronic and software
	Feature divergence	[15]	Use SCL algorithm to address feature divergence between a source domain and a target domain.	Online reviews	Amazon product reviews (Blitzer datasets)	Book, DVDs, electronics, kitchen appliances
	Feature divergence	[11]	Address feature divergence problem by introducing an algorithm that solves both feature divergence and polarity divergence.	Online reviews	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen
	Feature divergence	[12]	Address feature divergence by creating a shared feature of domains as well as domain-specific features.	Online reviews	Amazon product review	Books, DVD, Electronics and Kitchen
	Polarity divergence	[11]	Transfers the polarity of features from a source domain to a target domain by using independent features as a bridge.	Online reviews	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen
	Polarity divergence	[13]	Address polarity divergence by considering sentiment score of a feature rather than pure polarity	Online review	www.komplett.no , and mpx.no (Norwegian sites)	Digital product review (Norwegian product reviews)
	Polarity divergence	[10]	Address polarity divergence by combining different representations of their dataset and using ensemble method	Twitter	Tweets containing happy and sad emoticons	General tweets
	Polarity divergence	[14], [36]	Introduce an approach based on word embeddings with deep learning to compute the polarity of texts [14]. Further, in [36], fuzzy logic is used to resolved any uncertainties associated with the polarity of terms from different domains	Online reviews	Dranziera datasets (includes Amazon Product reviews)	IMD, OMD
	Sparsity	[9]	Address sparsity problem by expanding feature vector with related features from sentiment sensitive thesaurus.	Online reviews	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen
	Sparsity	[37]	Address data sparsity in low-resource language (Hindi) by leveraging the word embedding created on resource-rich language such as English.	Online reviews	Amazon product reviews (English-Hindi and English-French)	Restaurant, Laptop
	Polysemy	[38]	Address polysemy problem by using BabelNet and WSD, so as to derive the meaning of words in context	Online review	Amazon product reviews (Blitzer datasets)	Books, DVD, Electronics and Kitchen

the public. This was used to monitor issues related to food safety, in which alert messages will be automatically disseminated to the public, in case of any violations. Current progress in sentiment analysis is believed to be imperative for a more impactful outcome. Obviously, the progress may not be possible if the process does not take into account datasets from various sources.

B. CHALLENGES AND RECOMMENDED SOLUTIONS IN MULTI-SOURCE SENTIMENT ANALYSIS

The use of multiple data sources allows for a thorough coverage of opinions on a particular topic, or product, hence providing a comprehensive and inclusive conclusion. The inclusion of multiple data sources, however, also poses several challenges to the process. First, with the overwhelming

amount of data, and opinions coming from numerous sources [4], a situation termed as information overload can occur. When this happens, the analysis process is slowed down, thereby producing a delayed or inaccurate sentiment outcomes. Therefore, the right filtration, and selection mechanisms is necessary for filtering the relevant data, or opinion from the irrelevant ones, and to select the most relevant ones for further analysis. Most of the studies that we reviewed had adopted the two-layer mechanism. The first layer was to determine the topic, product, or event of interest by using the topic detection, or the event detection methods. The second layer was to fetch data related to the identified topic, product, or event of interest from all the data source involved, and then to classify the relevant opinions, based on the identified topics, products or events. For instance, [28] first identified the events from the online news, then based on the identified events, they built a microblog theme crawler to find the associated microblogs. Reference [8] also initially used a topic modelling algorithm to identify the interesting topics about the stock market from the economic news articles. Subsequently, for each topic identified, they used the burst detection, and burst event grouping algorithms to classify the related tweet messages into useful patterns. These were then used to represent the public's social movements and sentiments of the stock market.

The second challenge that has to be addressed by the multi-source sentiment analysis is the presence of various characteristics of the data in the data sources. For instance, tweet messages are shorter than online reviews, and the language used is less formal. Our review noted that the divergence between these data had not been thoroughly studied, or discussed. Data can also be in a form of Google search volumes, or bursts of tweets over a duration of time [8], or in the form of news article, or quantitative data such as historical economic data [27]. With these differences, it is impossible to simply merge them for analysis. Our review also found that almost all the studies processed, and analysed data from the different sources independently, and the resulted outputs were then combined in a data integrated model for further analysis so as to determine the final conclusion, or prediction [8], [27], [28]. Reference [4] addressed the differences of the data characteristics by creating a uniform data structure to represent the data which were obtained from different sources. The uniform data structure was then created to represent data obtained from Twitter, Facebook and Instagram. Unfortunately, the work did not further describe the design of the uniform data structure.

C. CHALLENGES AND RECOMMENDED SOLUTIONS IN MULTI-DOMAIN SENTIMENT ANALYSIS

Generally, the recommended solutions for adapting the different domains found in our review could be divided into four categories, based on their techniques. They are: transfer learning, multi-task learning, word embeddings, and sentiment lexicon. The transfer learning approach had been widely used for domain adaptations [9], [14], [17]. The SCL [15]

and the SFA [10] were among the earliest and notable solution used for domain adaptations. These adaptations were based on the transfer learning approach. Both adaptations demonstrated the transferring of sentiment knowledge from a source domain to a target domain by using a new feature representation termed as a bridge. Subsequently, [9], and [17] recommended the use of the multiple source domain to improve the adaptation performance of the sentiment classifier, across different domains. Despite the success of the transfer learning approach in domain adaptations, the approach requires a new transfer learning model to be developed each time a need to analyse a new domain arises. To tackle this limitation, several alternative solutions had been proposed, including those that were based on multi-task learning [17], [19], [33], word embeddings [14], [34], [36], and sentiment lexicon [9], [18], [22], [23]. The advantages and disadvantages of these approaches are summarised in Table 2.

In the context of sentiment analysis, the vocabularies of each domain would contain words that can be categorised into domain-specific words, or domain independent words. The domain-specific words are predictive in one domain, but not in another domain, as exemplified by the word, 'read' in the Books domain and 'play' in the Video Games domain. This situation would lead to a feature divergence problem when predictive features in a target domain cannot match the classifier built by the predictive features of the trained domain.

On the other hand, the domain independent words were predictive features in both the source and target domains. They can either represent the same polarity or the different polarity in both domains. For instance, 'good' is used to represent the positive polarity in many domains including the Books and Electronics domains whereas 'easy' signifies a positive polarity in the Books domain, but a negative polarity in the Movies domain. This condition is known as polarity divergence. Besides feature divergence, and polarity divergence, other components of the NLP challenge which can influence multi-domain sentiment analysis are sparsity and polysemy. The effectiveness of a sentiment classifier often depends on its ability to address these components of the NLP challenge [21]. Our review showed that among the components of the NLP challenge, polysemy and sparsity, were two areas least explored. Our review also showed that most of the methods used to address feature divergence, and polarity divergence were based on transfer learning, or a sentiment sensitive thesaurus.

V. FUTURE DIRECTION

Our review has shown that there is a growing need to implement sentiment analysis by using data from multiple data sources. Additionally, there has also been a continuous effort to improve the accuracy of sentiment analysis across multiple domains. Nonetheless, a comparative empirical study on multi-source sentiment analysis with or without multi-domain is yet to be implemented. The availability of such

a study would pave the way for a better evaluation of the following:

- 1) The advantages and disadvantages of implementing multi-source sentiment analysis with multi-domain,
- 2) The advantages and disadvantages of implementing multi-source sentiment analysis without multi-domain,
- 3) The significant impact of multi-domain on the accuracy and efficiency of sentiment analysis when implementing (i) and (ii), and the reasons.

It appears that even though benchmark datasets like the Amazon product reviews produced by [15] datasets were used, the evaluation procedures were still proprietary. Based on this, it is difficult to perform benchmarking among the recommended approaches for multi-source and multi-domain sentiment analysis. Moreover, there is no work done on the unified performance evaluation model yet, in order to benchmark the accuracy of the recommended solutions, even though it has been suggested by [21]. In this regard, it is important to perform a comparative empirical study on the existing multi-source data integration models so as to understand the factors that influenced their designs, and performances.

Also worthy of mention is that even though research works on multi-domain sentiment analysis had been done since a decade ago, the datasets used were still confined to online reviews, particularly the Amazon product reviews. Therefore, research in multi-domain sentiment analysis should consider using reviews from multiple data sources, such as online news, blogs, and social media channels. These can serve as the alternative approach for addressing the problem of insufficient datasets for training a sentiment classifier. Finally, more research work should be carried out so as to address the polysemy and scarcity issues noted in multi-domain sentiment analysis. This could provide a more accurate and humanized outcome for sentiment analysis.

VI. CONCLUSION

Sentiment analysis which relies on single data source is prone to biased conclusions, and inaccurate prediction outcomes. Realising this, recent studies in sentiment analysis has slowly transitioned from the use of one data source to multiple data sources. In this paper, we have reviewed, and analysed the challenges associated with this new progress. For each of the challenges found in both the multi-source and multi-domain sentiment analysis, we also discussed the approaches taken by the researchers in their attempts to provide solutions to the challenges, which encompass the types of data sources and datasets used. This review has classified the challenges found in multi-source and multi-domain sentiment analysis into four main categories: opinion overload, different data characteristics, domain adaptations, and the NLP challenge. For ease of understanding, the recommended solutions derived from the literature for each of the challenges were then classified according to the techniques used. We believe that the findings derived from our review of 26 studies can serve as a useful guide for future research to facilitate further progress, and advancement in multi-source and multi-domain sentiment analysis.

REFERENCES

- [1] N. Hajli, "Social commerce constructs and consumer's intention to buy," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 183–191, 2015.
- [2] V. Jha, R. Savitha, P. D. Shenoy, K. Venugopal, and A. K. Sangaiah, "A novel sentiment aware dictionary for multi-domain sentiment classification," *Comput. Elect. Eng.*, vol. 69, pp. 585–597, Jul. 2018.
- [3] L. Peng, G. Cui, M. Zhuang, and C. Li. (2014). *What Do Seller Manipulations of Online Product Reviews Mean to Consumers*. Accessed: Jul. 2019. [Online]. Available: <http://commons.ln.edu.hk/hkibswp/70>
- [4] P. Ducange, M. Fazzolari, M. Petrocchi, and M. Vecchio, "An effective Decision Support System for social media listening based on cross-source sentiment analysis models," *Eng. Appl. Artif. Intell.*, vol. 78, pp. 71–85, Feb. 2019.
- [5] A. B. Eliacik and N. Erdogan, "Influential user weighted sentiment analysis on topic based microblogging community," *Expert Syst. With Appl.*, vol. 92, pp. 403–418, Feb. 2018.
- [6] R. Ireland and A. Liu, "Application of data analytics for product design: Sentiment analysis of online product reviews," *CIRP J. Manuf. Sci. Technol.*, vol. 23, pp. 128–144, Nov. 2018.
- [7] B. Heredia, T. M. Khoshgoftaar, J. Prusa, and M. Crawford, "Integrating multiple data sources to enhance sentiment prediction," in *Proc. IEEE 2nd Int. Conf. Collaboration Internet Comput. (CIC)*, Pittsburgh, PA, USA, Nov. 2016, pp. 285–291.
- [8] F. Jin, W. Wang, P. Chakraborty, N. Self, F. Chen, and N. Ramakrishnan, "Tracking multiple social media for stock market event prediction," in *Proc. Ind. Conf. Data Mining*, Ney York, NY, USA, 2017, pp. 16–30.
- [9] 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.
- [10] A. Tsakalidis, S. Papadopoulos, and I. Kompatiari, "An ensemble model for cross-domain polarity classification on twitter," in *Proc. Int. Conf. Web Inf. Syst. Eng.*, Thessaloniki, Greece, 2014, pp. 168–177.
- [11] 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, Nov. 2015.
- [12] 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, Jul. 2015.
- [13] 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," in *Proc. IFIP Int. Conf. Comput. Sci. Its Appl.*, Saida, Algeria, 2015, pp. 205–216.
- [14] M. Dragoni and G. Petrucci, "A neural word embeddings approach for multi-domain sentiment analysis," *IEEE Trans. Affect. Comput.*, vol. 8, no. 4, pp. 457–470, Oct./Dec. 2017.
- [15] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc. Conf. Empirical Methods Natural Lang. Process. Assoc. Comput. Linguistics*, Sydney, NSW, Australia, 2006, pp. 120–128.
- [16] 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*, Raleigh, NC, USA, 2010, pp. 751–760.
- [17] F. Wu and Y. Huang, "Sentiment domain adaptation with multiple sources," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, Berlin, Germany, 2016, pp. 301–310.
- [18] F. Wu, S. Wu, Y. Huang, S. Huang, and Y. Qin, "Sentiment domain adaptation with multi-level contextual sentiment knowledge," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Indianapolis, IN, USA, 2016, pp. 949–958.
- [19] Z. Yuan, S. Wu, F. Wu, J. Liu, and Y. Huang, "Domain attention model for multi-domain sentiment classification," *Knowl.-Based Syst.*, vol. 155, pp. 1–10, Sep. 2018.
- [20] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.
- [21] T. Al-Moslmi, N. Omar, S. Abdullah, and M. Albared, "Approaches to cross-domain sentiment analysis: A systematic literature review," *IEEE Access*, vol. 5, pp. 16173–16192, 2017.
- [22] G. Katz, N. Ofek, and B. Shapira, "ConSent: Context-based sentiment analysis," *Knowl.-Based Syst.*, vol. 84, pp. 162–178, Aug. 2015.
- [23] H. Cho, S. Kim, J. Lee, and J. S. Lee, "Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews," *Knowl.-Based Syst.*, vol. 71, pp. 61–71, Nov. 2014.

- [24] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Comput. Hum. Behav.*, vol. 31, pp. 527–541, Feb. 2014.
- [25] W. He, Z. Zhang, and V. Akula, "Comparing consumer-produced product reviews across multiple Websites with sentiment classification," *J. Organizational Comput. Electron. Commerce*, vol. 28, no. 2, pp. 142–156, 2018.
- [26] W. Fangzhao, H. Yongfeng, and Y. Zhigang, "Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources," *Inf. Fusion*, vol. 35, pp. 26–37, May 2016.
- [27] X. Zhang, X. Zhang, S. Qu, J. Huang, B. Fang, and P. Yu, "Stock market prediction via multi-source multiple instance learning," *IEEE Access*, vol. 6, pp. 50720–50728, 2018.
- [28] F. Li, Y. Lv, Q. Zhu, and X. Lin, "Research of food safety event detection based on multiple data sources," in *Proc. Int. Conf. Cloud Comput. Big Data (CCBD)*, Marrakesh, Morocco, Nov. 2015, pp. 213–216.
- [29] B. Peng, J. Li, J. Chen, X. Han, R. Xu, and K.-F. Wong, "Trending sentiment-topic detection on twitter," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*, Cairo, Egypt, 2015, pp. 66–77.
- [30] B. Tan, E. Zhong, E. W. Xiang, and Q. Yang, "Multi-transfer: Transfer learning with multiple views and multiple sources," in *Proc. SIAM Int. Conf. Data Mining*, Texas, TX, USA, 2013, pp. 243–251.
- [31] T. L. Saaty, "Decision making with the analytic hierarchy process," *Int. J. Services Sci.*, vol. 1, no. 1, pp. 83–98, 2008.
- [32] J. Zhou, J. Chen, and J. Ye, "Malsar: Multi-task learning via structural regularization," Ph.D. dissertation, Arizona State Univ., Tempe, AZ, USA, 2011.
- [33] F. Wu and Y. Huang, "Collaborative multi-domain sentiment classification," in *Proc. IEEE Int. Conf. Data Mining*, New Jersey, NJ, USA, Nov. 2015, pp. 459–468.
- [34] 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.*, Ho Chi Minh, Vietnam, 2016, pp. 159–166.
- [35] A. Weichselbraun, S. Gindl, and A. Scharl, "Extracting and grounding contextualized sentiment lexicons," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 39–46, Mar./Apr. 2013.
- [36] M. Dragoni and G. Petruccia, "A fuzzy-based strategy for multi-domain sentiment analysis," *Int. J. Approx. Reasoning*, vol. 93, pp. 59–73, Feb. 2018.
- [37] M. S. Akhtar, P. Sawant, S. Sen, A. Ekbal, and P. Bhattacharyya, "Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol.*, Louisiana, LA, USA, 2018, pp. 572–582.
- [38] M. Franco-Salvador, F. L. Cruz, J. A. Troyano, and P. Rosso, "Cross-domain polarity classification using a knowledge-enhanced meta-classifier," *Knowl.-Based Syst.*, vol. 86, pp. 46–56, Sep. 2015.
- [39] 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.
- [40] T. Menner, W. Höpken, M. Fuchs, and M. Lexhage, "Topic detection: Identifying relevant topics in tourism reviews," in *Information and Communication Technologies in Tourism 2016*, A. Inversini and R. Schegg, Eds., Springer, 2016. doi: [10.1007/978-3-319-28231-2_30](https://doi.org/10.1007/978-3-319-28231-2_30).



NOR ANIZA ABDULLAH received the master's degree in interactive multimedia from Westminster University, London, and the Ph.D. degree in computer science from Southampton University, U.K. She is currently an Associate Professor with the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. She has published articles in both international and local journals, book chapters, and conferences. Her research interests include in personalized information retrieval, recommender systems, adaptive learning, sentiment analysis, and multimedia content-based retrieval and big data. She serves as a Reviewer for several ISI-indexed journals.



ALI FEIZOLLAH received the bachelor's degree in information system (IS) from the Ajman University of Science and Technology (AUST), Ajman, UAE, in 2010, the master's and Ph.D. degrees in computer science with the University of Malaya, Kuala Lumpur, Malaysia, in 2011 and 2017, respectively, where he is currently a Postdoctoral Research Fellow. His research interests include mobile malware, intrusion detection systems, sentiment analysis, and machine learning.



AININ SULAIMAN has published articles in *Information and Management*, *Computers in Human Behavior*, *Telematics and Informatics*, *Behavior and Information Technology*, the *American Journal of Scientific Research*, *Management Decisions*, the *International Journal of Mobile Communications*, and *Government Information Quarterly*. Her research interests include management information systems, technology diffusion, e-commerce, and green information technology.



NOR BADRUL ANUAR received the master's degree in computer science from the University of Malaya, Kuala Lumpur, Malaysia, in 2003, and the Ph.D. degree in information security from the Centre for Security, Communications and Network Research, Plymouth University, U.K., in 2012. He is currently an Associate Professor with the Faculty of Computer Science and Information Technology, University of Malaya. He has published a number of conference and journal papers locally and internationally. His research interests include information security (i.e., intrusion detection systems), data sciences, artificial intelligence, and library information systems.