IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 10, 2020, accepted October 13, 2020, date of publication October 16, 2020, date of current version October 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3031588

A Multiclassification Model of Sentiment for E-Commerce Reviews

SHAOZHONG ZHANG[1], DINGKAI ZHANG[2], HAIDONG ZHONG[3], AND GUORONG WANG[1]

[1]College of Information and Intelligence Engineering, Zhejiang Wanli University, Ningbo 315100, China
[2]Department of Computer Information Technology and Graphics, Purdue University Northwest, Hammond, IN 46323, USA
[3]Institute of Digital Industry Research, Zhejiang Wanli University, Ningbo 315100, China

Corresponding authors: Shaozhong Zhang (dlut_z88@163.com) and Dingkai Zhang (zhan3385@purdue.edu)

This work was supported in part by the Public Technology Research Project of Zhejiang under Grant LGF20F020004, in part by the Humanities and Social Sciences Research Project of the Ministry of Education of China under Grant 20YJAZH130, in part by the Zhejiang Philosophy and Social Science Planning, Zhijiang Youth Project, under Grant 19ZJQN20YB, in part by the Zhejiang Social Science Planning Project for Zhijiang Youth Cultivation under Grant G306, in part by the Ningbo Natural Science Foundation Key Project under Grant 2019A610046, and in part by the Ningbo Philosophy and Social Sciences Leader Training Project, Role and Development Strategy of Cross-Border E-Commerce in Ningbo's Transition from a Big Foreign Trade City to a Strong Foreign Trade City.

ABSTRACT Consumer reviews are important information that reflects the quality of E-commerce goods and services and their existing problems after shopping. Due to the possible differences in consumers' experiences with goods and service quality, consumer reviews can involve multiple-aspect expressions of emotions or opinions. This may result in attitudes expressed by a consumer in the same review sometimes having a variety of emotions. We introduce a sentiment multiclassification method based on a directed weighted model. The model represents the sentiment entity vocabulary as the sentiment nodes and represents the relation between nodes as the directed weighted link. The sentiment entity vocabulary is the entity with attributes, which can express sentiment meaning in related reviews. Directed weighted links represent the sentiment similarity between two nodes of entities with attributes and determined by the direct correlation calculation between them. The paths are all connected directed links from one node to another, which are composed of several nodes and links with close sentiment similarity. Then, we can establish a directed weighted model concerning the sentiments. Directed weighted links having similar sentiment relations with each other may constitute a directed weighted path. There are several directed weighted paths from a start node to the end nodes of the sentiment entity vocabulary in the directed weighted model. Each different path is a different sentiment expression, which represents a different sentiment type. The different sentiment classifications can be obtained through the restriction of path length. Experiments and analysis of the results show that the sentiment multiclassification model based on the directed weighted model proposed in this paper can classify the review sentiments according to different limited threshold rules. Comprehensive analysis indicates the classification results have good accuracy and high efficiency.

INDEX TERMS E-commerce reviews, sentiment multiclassification, directed weighted model, directed weighted link, directed weighted path.

I. INTRODUCTION

With the rapid development and application of WEB 2.0 technology, sharing reviews in E-commerce have become an important way for consumers to express their opinions and exchange experience online. The information of E-commerce reviews contains consumers' various evaluations and comments on commodities and services, including the performance, quality and price of commodities. The reviews involved in the attitude and quality of service, the situation of the logistics or the express delivery. All of these are objects of reviews. The review information on an object is published and updated by a large number of consumers, and accumulated over time into a multiclassification data set with complex structure, diverse content, sentiments and emotions. This kind of data includes many types of information, including the users' evaluations, attitudes, and behaviors which are determined to certain things, such as events, commodities

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

and services. It is very important to extract this valuable knowledge from this huge amount of data to provide valuable services for enterprises, institutions and individuals.

It is a significant way for online shopping consumers to share their personal experiences by writing related reviews which could reflect the quality of goods and services and the problems of consumers in E-commerce. Consumers can express their attitudes towards the quality of goods and services, personal feelings and so on through reviews. Since different consumers have different concerns about quality of products and services, they will have different feelings about their online shopping experiences. Then, their reviews may express various emotions, sentiments and attitudes. In addition, consumer's reviews may involve multiple-aspect expressions of emotions or opinions, which may result in the fact that the sentiments expressed by consumers in a same review are sometimes not singular. In many cases, consumers may agree with some opinions and disagree with others in one review rather than affirm or deny the entire review. Therefore, an E-commerce review presents a complex and multiple sentiment state.

Sentiment analysis is a branch of Natural Language Processing(NLP) that studies the meanings and attributes of a given text through the analysis of text terms. The task of sentiment analysis is to detect and extract all sentiment entities and their attributes in the context of a review document [1]. In recent years, with the rapid development of social media and online review applications, the review information has spread all over social, economic, political, business and other aspects. Especially in recent years, with the rapid development of E-commerce, there is an increasing number of reviews related to commodities and merchants in E-commerce, which is an important factor affecting the buying tendency of customers. Therefore, it is of great significance to study the sentiment analysis of reviews in E-commerce.

Sentiment of a context generally consists of five parts, which are, (1) object, (2) attribute of the object, (3) sentiment meaning to the object, (4) holder of sentiment, and (5) time of sentiment expression [2]. The existing studies usually use entities to describe the goals for sentiment analysis [3]. This type of entity can be a product, a service, a topic, a person, an organization, or an event. The relation between entities is a hierarchical structure described by a set of attribute values, and the entities at different layers have their own attribute values [4]. Sentiment meaning of an object refers to the sentiments of an entity in a certain attribute, which can usually be expressed as positive, negative, or neutral or described by a set of different strength levels [5]–[7].

The existing approaches to sentiment analysis fall into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches [8]. Corresponding to the components of sentiment, the first task of sentiment analysis in knowledge-based techniques is to extract entities, which are generally implemented through entity name recognition [9], [10]. Based on entity name recognition, it is also necessary to classify similar entities [11], [12]. Entity attributes and their classification use attributes to represent the characteristics of an entity, besides, different attributes have different names. Sentiment classification is a key problem of studying and analyzing the overall emotional tendency in a document [13], [14]. Standard machine learning methods could be used for sentiment classification by statistical methods, such as Support Vector Machines, Maximum Entropy, and Naive Bayesian classifier [15]–[17]. All of these methods could be used in topic-based sentiment classification research [18]. The hybrid approaches exploit both knowledge-based techniques and statistical methods to perform tasks such as emotion recognition and polarity detection from texts [19].

However, online reviews are increasingly expressing a trend of uncertainty. These reviews often exist in the form of a short text, and they are highly random in their attitudes towards a certain entity. These reviews are not strictly abided by the integrity of the information. Besides, there are usually several entities used to describe an object in a review. People may have different sentiments towards different entities. Thus, these sentiments exhibit several emotions that differ from one entity to another for the same object. This results in multisentiments to an object, which means, there may be one sentiment for one entity and another sentiment for another within a review of the same object. It shows that the sentiment for an object with different entities in a review may have different sentiments. This type of different sentiments is multisentiment. Therefore, how to distinguish the different sentiments classification entities in a same review is the focus of the study [20]. This is the multiclassification problem in the sentiment research.

The research in this paper takes the data of consumer reviews in E-commerce as the target, which generally come from the consumer comments and evaluations on E-commerce websites, and descriptions related to topics on blogs, microblogs, and Twitter, etc. We introduced a directed weighted model for sentiment multiclassification research. Our proposal is to combine the directed weighted model with sentiment analysis, extract sentiment keywords as nodes from entities and their attributes, and build directed weighted links between two nodes to satisfy a particular condition. Directed weighted links represent sentiment similarity between nodes. Then, a directed weighted path is proposed to search the similarity feature nodes of sentiments and implement sentiment classification analysis. A type of sentiment classification is a set of nodes and directed weighted links, where the nodes are on the directed weighted path and the paths are based on a series of closely related directed weighted links.

The rest of the paper is organized as follows. The related works section discusses the related work of our study, including the existing studies in sentiment analysis and opinion mining, as well as the issue of the sentiment classification technique, including supervised, semi-supervised, and unsupervised methods. Based on summarizing the existing sentiment classification research, the new problems of sentiment

multiclassification are presented. In the definition section, several related definitions are given to describe sentiment multiclassification and the directed weighted model. The next is the general approach to computing the framework, which is presented to study the methods and processes of sentiment multiclassification based on the directed weighted model. In the computational methodology section, an MDK-LDA method is used to address the exploitation of objects and topics. The mutual information is adopted to extract the entities with attributes. The formula is put forward to calculate the links weights and paths lengths between nodes, which are entities with attributes. Then, based on the shortest path theory, a sentiment multiclassification algorithm is proposed. The experiment and analysis section presents the experimental analysis of the proposed model and algorithms based on a dataset of publicly available online E-commerce reviews. The last section concludes the paper and discusses possible points that need to be furthered in the future.

II. RELATED WORKS

A. SENTIMENT ANALYSIS AND ITS CLASSIFICATION

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards objects or topics, such as products, services, organizations, individuals, issues, events, and their attributes [2]. In the last ten years, many researchers have carried out some productive studies in sentiment analysis and opinion mining for the review text contents online. The purpose of sentiment analysis and opinion mining is to analyze and identify the views, attitudes and emotions of people regarding a particular object or topic [21]. Those objects and topics can be persons, events, and commodities. Besides, an object or topic may be composed of multiple parts, each part can express an actual sentiment meaning, and these entities with independent sentiments all have abilities to express the user's emotions [22]. Therefore, the entities may represent some aspects of the object, and an object may include several entities. Sometimes these entities express the same type of emotional tendency, but sometimes they express the different.

Currently, the existing studies have mostly focused on the research of sentiment classification of text data [20]. Sentiment classification is one of the important parts of sentiment analysis; it separates subjective sentences from objective ones and then identifies the polarity (negative, neutral or positive) of the sentiments that expressed in the subjective sentences [4]. In the existing sentiment classification research, the main methods can be divided into three types, supervised, unsupervised, and semi-supervised classifications. Venugopalan and Gupta [23] presented a machine learning algorithm based on distance monitoring. It divided Twitter information into two types, positive and negative. It presented the results of machine learning algorithms on classifying the sentiment of Twitter messages using distant supervision method. The team of Li and Liu [24] applied a TF-IDF weighting method to a mechanism for voting and importing term scores, and then obtained an acceptable and stable clustering result. It also commits to the direction of positive and negative polarity classification. Liu, etc. Reference [25] proposed a rule-based sentiment polarity calculation method which may be used for extracting sentiment features from Chinese reviews; it is based on a sentiment word lexicon to calculate the basic polarity of the sentiment features, meanwhile, a dynamic sentiment word's polarity is judged and adjusted according to the context information. A method for collecting a corpus with positive and negative sentiments, and a corpus with objective texts has been presented by Pak's team [26]. This method can collect both negative and positive sentiments without human efforts. In addition, the objective texts are automatically collected, and the size of the collected corpora can be arbitrarily large. It performs statistical linguistic analysis of the collected corpus. Based on the collected corpora, a sentiment classification can be built. Kao and Lin [27] established a sentiment analysis system for a review of Chinese sentiment orientation analysis. The system analyzes the problems of tendency with sentiment contents of the reviews according to some certain characteristics and some views of a particular category. It proposed the use of the concept of dependencies to identify review sentiment orientation. Zhang, etc. Reference [28] aimed to reduce the annotation effort for multi-modal sentiment classification via semi-supervised learning method. Its key idea was to use the semi-supervised variational autoencoders to detect more information from unlabeled data for multi-modal sentiment analysis.

B. MULTICLASSIFICATION OF SENTIMENT

Sentiment classification can be carried out by existing techniques for polarity identification. This type of polarity generally has three classes which are positive, negative, and neutral [29]. Other proposals include several classes or subclasses (e.g., "very positive," "positive," "mostly positive," and "very negative," "negative," and "mostly negative") [30], [31]. In a related context, most of the state-of-the-art works and researches on the automatic sentiment analysis and opinion mining of texts are collected from social networks and microblogging websites, which are oriented towards the classification of texts into positive and negative [32].

In recent years, studies related to sentiment multiclassification have also been received some attention. The studies [33]–[35] proposed a novel approach for the classification of texts collected from Twitter Which can classify these tweets into multiple sentiment classes in addition to the tasks of binary and ternary classification. They limited their scope to seven different sentiment classes. The proposed approach is scalable and can be run to classify texts into more classes. The study [30] proposed a multilabel classification-based approach for sentiment analysis. This work is the first research that tried to propose the use of multilabel classification for sentiment classification of microblogs. The prototype they proposed has three main components, text segmentation, feature extraction, and

multilabel classification. They conducted a detailed empirical study of different multilabel classification methods for sentiment classification to compare the classification performances. In addition, the study [36] designed a more complex multilabel ABSA method that could predict one or multiple aspect-sentiment labels from the text. Also the study [37] promoted a Recurrent Neural Network (RNN) language model based on Long Short-Term Memory (LSTM) networks to implement multiclassification for texts sentiment. Their method can help people to get complete sequence information effectively. Their results showed that compare with the traditional RNN language model, LSTM is better for analyzing the emotion of long sentences. Moreover, as a language model, LSTM is mainly applied to achieve multiclassification tasks for understanding text emotional attributes. Other studies have improved LSTM and proposed the Bidirectional Long Short-Term Memory (Bi-LSTM) method, which has higher efficiency in identify the emotional polarity of product reviews [38], [39].

Other studies have carried out the sentiment analysis at different levels: the document level, the sentence level, and the entity and feature level [40]. The document-level sentiment analysis judges the overall emotional tendency of the documents, and its classification results are generally divided into fixed types, positive, negative and neutral types, etc. In the sentence level analysis, it considers whether sentences can express any opinion [41], [42]. In the entity and feature level analysis, it looks directly at the opinion itself based on the idea that an opinion consists of a sentiment [27].

C. PROBLEMS AND SOLUTIONS

The proposed state-of-the-art approaches have signification achievements in sentiment analysis and classification. However, most of the existing methods are mainly focused on exploring binary and ternary sentiment classification for reviews. Although some research studies have divided sentiment into subclasses or several levels, these subclasses and levels describe the different degrees of the same sentiments only. It is impossible to divide the sentiment of reviews into more detailed categories.

(1) The existing research studies on sentiment classification mostly do not consider the various attitudes of each entity involved in a review. Most of them divide the sentiments into several stable types, such as positive, negative, and neutral types. Some of them divide the emotions into more classes, such as happiness, anger, disgust, surprise, fear, sadness, etc. However, since the review of an object often contains several different parts, and each part may contain a different entity of the object. People may have different attitudes towards these entities, which may result in different sentiments for the whole object. As a result, the sentiment classification of reviews cannot be built for simple certain types only.

For example, there is a review of a cell phone by userID "ABC" from Amazon.com: "The cell phone looks just like the picture (a). They are stickers (b) and it work well (c). I just do not like the rounded shape because I was always bumping (d) it and Siri kept popping up (e). The battery life is also long (f). However, my wife thinks it is too heavy for her (g). I just won't buy a product like this again (h)."

There are several sentences to describe the object "cell phone." The main entities that describe the phone in these sentences are "picture," "work," "shape," and "battery." The customer expressed different emotions about different entities. It is difficult to classify the sentiment types of the review using existing sentiment analysis methods.

(2) Besides, the existing sentiment classification studies are mainly based on the use of statistics. However, it is difficult for the statistics-based approach to distinguish the sentiment relations between words.

In the example above, (a), (b), (c) and (f) can be thought of as positive emotions, (d), (g), and (h) can be negative emotions, and (e) can be neutral emotions. Nevertheless, what is the sentiment of the review as a whole? Obviously, we cannot accurately classify those sentiments through a simple statistics of sentiment vocabulary.

In this paper, we propose a directed weighted model for sentiment classification of reviews. The model is composed of nodes and directed weighted links. The nodes are extracted from the reviews, which represent the entity with attributes, and the directed weighted links are edges from one node to another. The directed weighted paths are sets with nodes and links from a start node to all end nodes which are connected by a series of directed weighted links. Each directed weighted path represents a type of sentiment classification. The similarity relation between nodes in the model is utilized to analyze sentiment relations between entities with attributes. The link weight is used to indicate the strength of the relation of them. The path is made up of several links and is used to divide sentiment categories of reviews. The path can connect all relevant sentiment nodes into a directed weighted path. We can identify a plurality of sentiment classification by computing the weight of the different paths. To some extent, the methodology of computing the strength of the relation between nodes can reduce some disadvantages of the existing methods in the previous. Moreover, we can determine the number of multiclassification sentiments of reviews based on a threshold on path length. Different thresholds may cause different classifications of sentiments. The directed weighted model proposed in the paper could solve the problem of multiclassification of review sentiments.

III. SENTIMENT MULTICLASSIFICATION AND DIRECTED WEIGHTED MODEL-RELATED DEFINITIONS

A. SYMBOLS AND DESCRIPTION

To facilitate the description, we list some necessary nomenclatures involved in the paper.

(1). S and $S(O, H, T)$: S is the sentiment of a target, and O, H, T are the object, holder, and time, respectively.
(2). $R_i \in C_k \wedge Label_i \in C_k$ is the *ith* review text.
(3). e is an entity, which is an aspect of an object in a review.
(4). (e, a) is an entity with attribute; e is an entity, and a is its attribute. Usually, it is a word pair.

(5). $MI(e_i, a_j)$ is the mutual information of entities and their attributes.
(6). V_n are nodes to denote the entities with attributes in the directed weighted model. $\vec{V}_n$ is the sorted set in the order in which the elements appear in the reviews.
(7). $N(V_i, V_j)$ is the frequency of V_i and V_i coexisting in the same documents. $N(V_i)$ and $N(V_j)$ are the frequencies of V_i and V_i independently existing, respectively.
(8). L_{ij} is the length of the directed weighed path from one node to another, and W_{ij} is the weight of the link between two nodes that are directed linked.
(9). C is the set of Sentiment Multiclassification, and C_i is the *i-th* subset of it.

B. DEFINITIONS

Sentiment analysis is based on the directed weighted model. In the structure of the model, the sentiment, the sentiment multiclassification, the object, the entity, the directed weighted link, the directed weighted path, and other related conceptions are involved. The relevant definitions of these conceptions are as following.

Definition 1 (Sentiment): Sentiment is a three-dimensional function. The function is expressed as $S(O, H, T)$, in which O is an object that is the target of reviews, such as a good, service, transaction, or an entire process of shopping with various attributes in E-commerce. Sometimes a topic is used instead of an object, and we simply consider both as objects. H is the sentiment holder and usually is a consumer. T is the time that the consumer expressed the sentiment. $S(O, H, T)$ represents the sentiment of the holder H for an object at a certain time. Usually, we use $S(O)$ as its simple form.

Definition 2 (Entity): An entity is represented as e. The aspects of objects are looked at as entities. The entities are the things related to the object, including color, price, weight, quality, etc. In practical application, an object is often described as being structured with multiple entities. These entities are characteristics of the object as in equation (1).

$$O = \sum_{i=1}^{n} e_i \tag{1}$$

Definition 3 (Entity With Attribute): An entity with an attribute is represented as (e, a), and it is a component of sentiment for an object. An entity with attribute refers to an entity described in a review and the consumer's attitude towards it. An entity with attribute consists of two parts: one part is the contents of the entity, and the other part is its attributes. The attributes are the reviewer's attitudes to these entities, and they include the characteristics, performances, and tendencies, etc. An entity can have multiple attributes, and an attribute is used against multiple entities. All entities with attributes of an object in a review comprise the global sentiment meaning of the object. In the form of an entity with attribute (e, a), the entity is e, and its attribute is a. Thus, there may be a series of such binary pairs to describe the sentiment of an object, such as the form (e_i, a_j).

In example of section II, the review is a customer's evaluation of the cell phone on Amazon.com. A certain type of cell phone is the object, e.g., "the cell phone." It has a set of entities, e.g., "picture," "work," "shape," and "battery," and a set of attributes, e.g., "stickers," "well," "bumping," "popping," "long," and "heavy." There is a set of entities with attributes, which are "picture"-"stickers," "work"-"well," "shape"-"bumping," "shape"-"popping," "battery"-"long," and "battery"-"heavy.".

Some reviews may not focus exclusively on a specific object. They also focus on certain topics. A topic can be an object too, e.g., "tax increase," with its parts "tax increase for the poor," "tax increase for the middle class" and "tax increase for the rich." To simplify the process, the approach to the topic is the same as the approach to the object in our work.

In our directed weighted model, each entity with attributes is viewed as a node because it can express a sentiment independently. We use V to represent the node of entities with attributes, and it is shown in equation (2).

$$V = (e, a) \tag{2}$$

Definition 4 (Directed Weighted Link): A directed weighted link is a directed edge from one node to another. From the start node, the order of the nodes is the same as the appear order of entities. For each node V_i and V_j, if the node V_j presents next after the node V_i, there may be a directed link from V_i to V_j, represented as $V_i \rightarrow V_j$. The existence of directed links depends on the sentiment similarity of the two nodes. A directed weighted link is represented as in equation (3)

$$V_i \rightarrow V_j | S(V_i) = S(V_j) \tag{3}$$

where $S(V_i)$ and $S(V_j)$ are the sentiments of entities with attributes. We use the link weight to represent the degree of the link of sentiment similarity. The weight is represented as W, and its value range is (0, 1). The greater the weight is, the greater the similarity between the two nodes is. In contrast, the smaller the weight is, the greater the difference between the two nodes is.

The nodes of entities with attributes may have different entities and their attributes, which represent sentiments. Some of them with the same sentiment similarity may have a relation which is established by sentiment similarity computing between the two nodes. If there is a similarity relation between them, it shows that the two nodes belong to the same sentiment classification and are represented as a directed weighted link from the first node to the second.

Definition 5 (Directed Weighted Path): A directed weighted path is a set of associated nodes that constitute a connected sequence of directed connections. A directed weighted path is represented as C. In the directed graph, the *i-th* directed path from start node V_u to end node V_v can be represented as in equation (4).

$$C_i = (V_u, V_{i1}, V_{i2}, \ldots V_{im}, V_v | S(V_u) = S(V_{i1}) = \ldots = S(V_v)) \tag{4}$$

The length of a directed path is the length of all the connecting links that pass between the beginning and the ending of the path. It expresses a sentiment similarity of all nodes on the path. We use L to represent the path length. The longer the path length is, the smaller the similarity of the nodes on the path will be. By comparison, the longer the path length is, the greater the similarity of the nodes on the path will be.

Definition 6 (Sentiment Multiclassification): Sentiment multiclassification refers to the description of consumers' different attitudes or tendencies towards objects based on the nodes of entities with attributes. We use C to represent the set of sentiment multiclassifications. The *i-th* element of C is C_i, and it is a type of classification of sentiment that has the same sentiment or opinion tendency. C_i and C are defined by equation (5) and (6), respectively.

$$C_i = (V_u \ldots V_v, \min(L_{u,v}) | \min(L_{u,v}) \leq \beta) \tag{5}$$

and

$$C = \sum_{i=1}^{n} C_i \tag{6}$$

Thus, the goal of sentiment multiclassification is to find a set of nodes in which the length of the shortest path connected with the nodes in the set is less than a certain default threshold. In a such path, the nodes are connected by a directed path which is used for constituting a set of nodes of entities with attributes that having the sentiment similarity. A path with the same or similar sentiments is considered as a classification of sentiments. Multiple this kind of paths are found by setting different default thresholds, and in this way multiclassification of sentiments can be achieved.

IV. FRAMEWORK OF SENTIMENT MULTICLASSIFICATION ANALYSIS

The sentiment multiclassification analysis is based on the directed weighted model. Our study aims to analyze E-commerce reviews. We can find the object and the main topic from an online E-commerce website by the ID of the product and the description of its style. These information constitutes the metadata about a product, and the metadata can be considered as objects and topics. Typically, in an E-commerce environment, customer reviews tend to target specific objects and topics. These objects and topics are often aimed at specific goods or services, and most consumers' reviews can be clearly divided into objects or topics, that is, the objects and topics of these reviews can be determined.

Usually, objects and topics are composed of multiple aspects, and sometimes users do not focus on the whole of the object in a reviews but instead, on a certain aspect of the object. These aspects are the core content of expressing the user's sentiments. It is necessary to extract the aspects that describe the object. The aspects of the objects are often present as various nouns. These nouns are the entities that make up the objects and the topics. People's emotional vocabulary, attitudes, and preferences in reviews are attributes of these entities. Next, we should extract all entities and related attributes That are related to the objects and the topics.

We consider the entities with attributes in reviews as the nodes and the sentiment similarity relations of these nodes as weighted links. The direction of the directed links is determined in the order in which the entities appear in reviews. To find the sentiment classification, we propose a general three-step sentiment multiclassification computing framework, as shown in Figure 1.

First, the extraction of sentiment nodes is performed through two steps, which are the extraction of entity words and the mining of entity-sentiment word pairs [43]. The entity-sentiment word pairs in the paper are called entities with attributes, and they are the keywords of sentiments to express various sentiments of consumer reviews on objects. We use a mutual information formula and classic sentiment lexicon to mine entity-sentiment word pairs. Each entity-sentiment word pair corresponds to an entity with attribute. In addition, an entity with attribute is a node in the directed weighted model. Nodes represent the resulting vocabulary, which is the set of sentiment nodes.

Second, we take the cosine formula to calculate the similarity between two nodes. A link between two nodes is based on whether there is similarity between them, and the weight is by cosine formula. In general, we set a fixed threshold value as the valid weight of the link between nodes. When the weight value of the link is larger than the threshold, it means that there is a close similarity between nodes, and a directed link is constructed between the two nodes. When the link weight is less than the threshold, it means that the sentiment similarity between the nodes is very small, and that there is no directed link between them.

Third, calculation of the shortest path is used to find the node with the closest sentiment relation. The entities with attributes as nodes are arranged based on the order in which they appear in the consumer reviews. In addition, we set the node without input links as the start node and the nodes without output links as end nodes. The shortest path calculation is to find the sets of nodes and links that have the shortest path between all the start nodes and the end nodes. We introduced an improved shortest path algorithm to calculate the shortest paths. Meanwhile, the link threshold and path length threshold are used to improve the accuracy and reduce the complexity of the algorithm. Each shortest path represents that all the nodes on a path have similar sentiment relations. We can think of this type of sentiment as a sentiment classification of the reviews. When there are more than two shortest paths in a review, it indicates that the review expresses more than one sentiment and can be divided into several classifications. However, not all these shortest paths represent a valid classification. Some shortest paths with overly small total weights that is lower than a pre-set experience threshold may express very weak sentiment similarity which can be ignored. Only those with weights that is higher than the threshold can be considered as a valid classification, while those that weights do not reach the threshold cannot be considered as a valid classification. During the training process, only the sentiments that are repeatedly emphasized

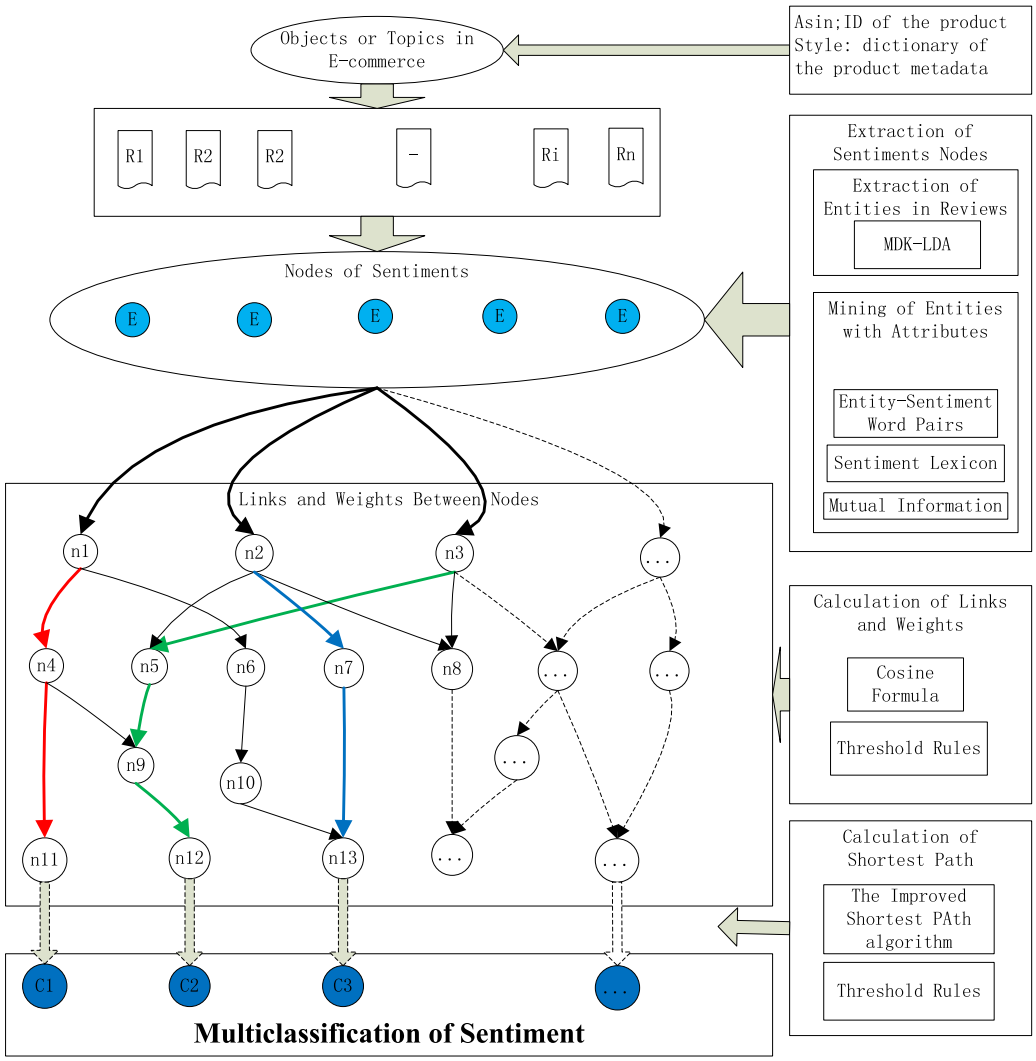


FIGURE 1. Sentiment multiclassification computing framework.

in the reviews can be considered as a valid expression of a sentiment, and the sentiments that are lightly mentioned are not a valid sentiment.

V. CALCULATION METHOD OF SENTIMENT MULTICLASSIFICATION

A. EXTRACTION OF ENTITIES FROM REVIEWS

Entity extraction is a principled approach for discovering aspects of objects from a large corpus of review documents [2]. The most common outputs of an entity extraction are a set of word clusters and an entity distribution for each document. Each word cluster is an entity. In E-commerce, consumer reviews are usually the evaluation of commodities and services for a certain online transaction process. Several aspects of commodities and services are the entities of objects, such as price, color, size and performance. In addition, because there are often comparisons between the same aspect of similar commodities and services in the consumer reviews, it is inevitable that there are also comments on the same aspects of other objects. Therefore, it is necessary to assemble the entities of objects more accurately for consumer reviews discussed based on combining relevant commodities and services of E-commerce reviews.

Entity extraction is a popular method in sentiment analysis. However, unsupervised entity extraction models often generate incoherent aspects. To address the issue, several knowledge-based models have been proposed to incorporate prior knowledge provided by the user to guide modeling [44], [45]. Based on the LDA model, Chen *et al.* [46] introduced a latent variable z and proposed an MDK-LDA method, which denotes the *s-set* assignment to each word. Assuming that there are *s-sets* in total. We use this method to extract entities.

Under MDK-LDA, the document of reviews from E-commerce is D, and the probability of word z given entity e, i.e., $p_e(z)$, is given by equation (7) [46].

$$p_e(z) = \sum_{m=1}^{M} \varphi_e(M) \cdot \eta_{e,m}(z) \tag{7}$$

m is an *M-set*, z is a word, φ is an *Entity-M-set* distribution, φ_e is an *M-set* distribution of entity e, η is an *Entity-M-set-Word*

distribution, and $\eta_{e,m}$ is the word distribution of the entity e, *M-set m*.

where $\varphi_e(m)$ denotes the probability of *M-set m* occurring under entity e and $\eta_{e,m}(z)$ is the probability of a word appearing in *M-set z*and m under entity e.

B. MINING OF ENTITIES WITH ATTRIBUTES

The mining of entities with attributes is to extract the relevant attributes that describe the entity, and constitute the entity with attribute, that is, the entity-attribute pair. If an entity has multiple attributes, multiple entity-attribute pairs will all be extracted. Each extracted entity-attribute pair represents a node in the directed weighted graph.

We consider the existing lexicon of sentiment words, WordNet [47], [48], as the lexicon. Within a certain window distance of an entity word, if the word belongs to the lexicon, we extract the word as the attribute corresponding to the entity and compose an entity-attribute word pair. We look at the pair as a node of the directed weighted graph. If there is more than one attribute word within a certain window distance, the entity and these attribute words will form entity-attribute word pairs separately.

We use a mutual information method to calculate the relation between an entity and its attribute. Mutual information of an entity word e_i and an attribute word a_j can be calculated by formula (8) [43].

$$MI(e_i, a_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(e_i, a_j) \log \frac{p(e_i, a_j)}{p(e_i) \cdot p(a_j)} \tag{8}$$

The mutual information between an entity and an attribute indicates the strength of the relation between the entity and the attribute. Entity-attribute word pairs with strong relations can represent a specific sentiment and can be regarded as nodes of a directed weighted graph model. Extraction of all such nodes, which can express specific sentiments, is the first step in the further analysis of different sentiments.

C. CONSTRUCTION OF THE DIRECTED WEIGHTED MODEL

The construction of the directed weighted model includes two sections. One is the directed links, and the other is the links weights. Whether there is a link between two nodes depends on the sentiment similarity of the two nodes. If there is a type of strong sentiment similarity from one node to another in documents, there is a directed link between the two nodes based on the order of appearance of the entities with attributes. The direction of the link is from the first to the second. The weight of the link represents the degree of sentiment similarity of nodes, and it is computed by the frequencies of the nodes.

In the node space of the model, V_i is the start node, and V_j is the other following connecting node. The first task is sorting of V_n. For the node set V_n, it is sorted based on the order in which each node appears in the document. The node that appeared first is denoted by V_i, and the following node is denoted by V_j. After identifying the order of all nodes and sorting them, the sorted set of nodes is $\vec{V}_n$.

For all nodes $\vec{V}_n$, one node V_i as a link-start node that can correspond to multiple link-end nodes V_j and one node V_j as a link-end node can correspond to multiple link-start nodes V_i. For the former, it means that an entity with attribute can be further divided into several different aspects, and each aspect is a type of sentiment. For the latter, it means that several entities with attributes have similar sentiments to the link-end node. **Definition 7** (Link Weight): We use $N(V_i, V_j)$ to denote the frequency of V_i and V_j coexisting in the same documents. $N(V_i)$ and $N(V_j)$ are the frequencies of V_i and V_i independently existing, respectively. The link weight of $V_i \rightarrow V_j$ is defined as $W_{i,j}$. It is a cosine similarity and computed by formula (9) [49].

$$W_{i,j} = \frac{N(V_i, V_j)}{\sqrt{N(V_i)^2 \times N(V_j)^2}} \tag{9}$$

Due to the complexity of human language and the randomness of review content, many possible sentiment similarity relations between nodes are weak links. That is, the sentiment similarity expressed by these links is low. If all these links were included in the model, the dimensions of the model would be out of control, and the model could be in a very complex state. To ensure the effectiveness and complexity, we set a manual empirical parameter to limit the minimum weight of a valid link. When the weight of the link is greater than or equal to the pre-set value, the link exists and has weight $W_{i,j}$. The link will not exist when its weight is less than the pre-set value.

By the link construction and weight calculation, we can construct a directed weighted graph model in which the nodes are entities with sentiments. Next, we can use the directed weighted graph model as the basis to set up sentiment classification.

D. MULTICLASSIFICATION OF SENTIMENTS BASED ON SHORTEST PATHS

In the directed weighted graph model, the initial start node is a set of consumer reviews, and the other nodes are extracted from sentiment entities with attributes. These nodes are processed in ordered. The first node extracted from each review that does not appear in all previous reviews can be viewed as a start node in the model, and there may be multiple start nodes depends on the number of nodes extracted from the reviews. Those nodes that do not have any subsequent nodes are considered as the end nodes. There is a directed weighted path from the start node to each end node. All nodes on the path have a certain sentiment similarity and express a similar emotion. The directed weighted path denotes a type of sentiment classification. There are many such paths in one reviews document, and different paths represent different sentiment classifications. Thus, we can implement the approach of the sentiment multiclassification of reviews.

The main method of sentiment classification is determined by calculating the length of each path which is from the start node to one of the end nodes. In the actual calculation, the path length is calculated by the link weight, and the link

Algorithm 1 Sentiment Multiclassification

Begin
Initialize
$k = 1$; $C_k = \varphi$; *//C_k is the kth classification.*
$\vec{V}_n = \{V_1, V_2, \ldots V_n\}$;
$V_s = V_1$ is the $1th$ node of $\vec{V}_n$; $i = 1$;
$L_{s,u} = \infty$; $L_{s,i} = \infty$; $Father(V_s) = \varphi$; *// The link start node of V_s.*
$Father(V_i) = \varphi$ for all $i \in n, i \neq s$;
$C_k = \{V_s\}$; $Q = \{V_s\}$. *//C_i is the ith sentiment classification and Q is the candidate nodes set.*
While ($\vec{V}_n \neq \Phi$) **do //Search** *Q to find the node which being with minimum path length to node V_s*;
$V_u = \{V_i | L_{u,i} \leq L_{u,j}, i, j \in Q, i \neq j\}$;
$Q = Q - \{V_u\}$;
$C_k = C_k \cup \{V_u\}$; *//the V_u is the node with shortest path from V_s to V_u and put it into C_k.*
For each $(V_u \rightarrow V_v) \in Out(V_u)$ **such that** $L_{s,u} + L_{u,v} < L_{s,v}$ and $L_{u,v} \leq \alpha$ **do //Update the path length label of** *V_v under the threshold α, and $Out(V_u)$ is an outgoing links set. α is the empirical value of the direct link between two nodes.*
$L_{s,v} = L_{s,u} + L_{u,v}$; $Father(V_v) = V_u$;
If $V_v \notin Q$ **then** $Q = Q \cup \{V_v\}$;
If $Out(V_v) = \varphi$ and $L_{s,v} \leq \beta$ **then** $k = k + 1$; *//If the node has no successor and the path length of the shortest path is not greater than the threshold β less than, a classification is established and go on next. β is the empirical value of the path length.*
Enddo
If $V_u \in \vec{V}_n$ **then** $\vec{V}_n = \vec{V}_n - \{V_u\}$; $V_s = V_{s+1}$;
Endwhile
end

weight is represented by the similarity of two nodes. The two nodes are those linked directed by a link. Therefore, we will take the reciprocal of the link weight as the path length. The larger the value of link weight between nodes is, the shorter the path length is. Conversely, the smaller the link weight between nodes is, the longer the path length is.

Our goal is to calculate the paths length from the start node to one of the end nodes, and to find the shortest path among them. Each set of nodes on the shortest path represent a different type of sentiment or opinion. We can consider it as a classification. It indicates that those nodes on the shortest path express a type of sentiment which can be classified as the same classification. In addition, since there could be more than one end nodes corresponding to one start node, it is necessary to find the all shortest paths from the start nodes to all the end nodes,

A threshold is set as the effective path length to verify the classifications. The path length of any shortest path that is less than the threshold can be considered an effective sentiment classification. A shortest path which path length exceeds the threshold is a path that is not a valid classification. The threshold is an empirical value, so a different number of shortest paths can get be obtained by setting different thresholds. Thus, the types of sentiment classification are also different, and the sentiment multiclassification is implemented.

Definition 8 (Path Length): We define the reciprocal of the weights as the path length between two directly connected nodes, and the length is denoted by $L_{i,j}$ in equation (10).

$$L_{i,j} = {}^{1}\!/_{W_{i,j}}, \quad W_{i,j} > 0. \tag{10}$$

According to definition (7), we know that the weight represents the degree of sentiment similarity between two nodes. The greater the weight is, the higher the sentiment similarity of the two nodes is. As well as, the smaller the weight is, the lower their sentiment similarity is. As for the definition (8), the path length is the reciprocal of the weight, and its meaning is the opposite of the weight. The shorter the path length is, the higher the sentiment similarity of the two nodes is. Besides, the longer the path length is, the lower their sentiment similarity is.

The algorithm of sentiment multiclassification based on the shortest path is presented as algorithm 1.

We can obtain the entire sentiment multiclassification based on the equation of $\sum_{k=1}^{K} C_k$. Each C_k represents a classification that differs from others. The review opinions and the current hot issues of public concern could be understanding in this way.

VI. EXPERIMENT EVALUATION

A. DATASET COLLECTION

Our experimental dataset is collected from the Amazon Review Data (2018). The data includes reviews in the period of May 1996 - Oct 2018. We use the 5-core subset of the data in which all consumers and items have at least 5 reviews. A small subset concerning cell phones and accessories is used for experimentation [50]–[52]. The review data file used in the experiment consists of the IDs of the reviewers, IDs of the products, texts of the reviews, summaries of the reviews, etc. The detailed structure of the file is shown in Table 1.

TABLE 1. Review data structure.

Field	Description
reviewerID	ID of the reviewer
asin	ID of the product
reviewerName	name of the reviewer
vote	helpful votes of the review
style	a dictionary of the product metadata
reviewText	text of the review
overall	rating of the review
summary	summary of the review
unixReviewTime	time of the review
reviewTime	time of the review
image	images that consumers post after they have received the product

TABLE 2. Label the review text by sentiment keywords.

ReviewText	Label of Classification
"These make using the home button easy. My daughter and I both like them. I would purchase them again. Well worth the price."	"button easy" "price well"
"Fit and finish very nice. Battery strength very poor. Takes forever to recharge phone. I would not suggest this to anyone."	"fit nice" "battery poor"
"it came in ok but there was a crack on the left side. also it slides apart really easily but its color is fine."	"side crack" "slides easily" "color fine"
……	……

10,000 reviews related to 2,000 cell phone and accessories products from the review data were selected as our test data and were labeled manually. The labels are the keywords of entities with attributes. Each review may have several labels depends on its content. The types of the sentiment multiclassification are constituted by all these labels. Samples of labels for reviews are shown in Table 2.

B. MEASURES

According to the data description in Tables 1 and 2, we used three fields to test the directed weighted model. These are "reviewerID," "asin," and "reviewText." When we analyze a review, if we find that the node set C_k of its classification contains all of the labels listed by the manual classification, it indicates that the classification is correct.

Definition 9 (Correct Reviews Number): The number of correct reviews in the classification is defined as equation (11). $R_i \in C_k \wedge Label_i \in C_k$ is the *ith* review, $Label_i$is the feature manually labeled to $R_i \in C_k \wedge Label_i \in C_k$. The meaning of $R_i \in C_k \wedge Label_i \in C_k$ is that when a review and a label belong to the same classification, then the classification of the reviews is correct.

$$CorNum = \sum_{i=1}^{n} R_i, R_i \in C_k \wedge Label_i \in C_k. \tag{11}$$

Definition 10 (Incorrect Reviews Number): If a review belongs to a classification and its manual label is not in the same one, the classification of the review**is**incorrect. The meaning of $R_i \in C_k \wedge Label_i \notin C_k$ is that when a review belong to a classification C_k, but its label is not involve in C_k. The number of incorrect reviews in the classification is defined as equation (12).

$$IncorNum = \sum_{i=1}^{n} R_i, R_i \in C_k \wedge Label_i \notin C_k \tag{12}$$

Definition 11 (Missing Reviews Number): When some of the manual labels to a review are included in the classification and others are missing, the review in the classification is defined as missing review. The meaning of $R_i \in C_k \wedge (not\ all\ Label_i \in C_k)$ is that when a review belong to a classification C_k, but not all its manual labels are involved in the C_k. The number of missing reviews in classification is defined as equation (13).

$$MissNum = \sum_{i=1}^{n} R_i R_i \in C_k \wedge (not\ all\ Label_i \in C_k). \tag{13}$$

Considering the correct, incorrect and missing cases, *Precision* and *Recall* can be calculated by equations (14) and (15) [53], [54].

$$Precision = \frac{CorNum}{CorNum + IncorNum} \tag{14}$$

$$Recall = \frac{CorNum}{CorNum + MissNum} \tag{15}$$

The *F-value* is usually typically used to represent the common impact of *Precision* and *Recall*, as shown in equation (16) [53], [54].

$$F - value = \frac{2Pecision \times Recall}{Pecision + Recall} \tag{16}$$

Moreover, the CPU time is used to perform the efficiency of the algorithm. The CPU time of the algorithm is the sum of CPU time of every part involved in the algorithm.

C. EXPERIMENTAL RESULTS AND ANALYSIS

We divided the records of the dataset into 10 sections on average. Each section contains 1,000 reviews. First, one section of the dataset is used as a testing set, and the 9 sections remaining are used as a training dataset to calculate the accuracy and the efficiency. The accuracy includes *Precision, Recall* and *F-value* in the experiment. The efficiency is the CPU time of the algorithm spent. Then, in the next step, another section is selected as the test set, and the 9 sections remaining in the dataset are used as the train set, including the previous section which was the test set. The *Precision, Recall, F-value, CPU time* are calculated again until all 10 sections are used as the test set once.

1) ACCURACY OF THE DIRECT WEIGHTED MODEL

BERT is a famous model for NPL since 2018 [55]. They use pre-training and fine-tuning to create state-of-the-art models for a wide range of tasks. One of the important functions of BERT is to accept a sentence and output a representation of a word. The BERT model can be used to analyze the relation between reviews and keywords of a classification. The vocabulary obtained by the BERT model is consistent with the labeled keywords. It shows that the result

TABLE 3. Accuracy comparison of the two models.

Section	BERT			Direct Weighted Model		
	Precision	Recall	F value	Precision	Recall	F value
1	0.801	0.736	0.767	0.780	0.829	0.804
2	0.701	0.589	0.640	0.688	0.752	0.719
3	0.703	0.825	0.759	0.758	0.780	0.769
4	0.768	0.653	0.706	0.685	0.749	0.715
5	0.799	0.784	0.792	0.751	0.854	0.799
6	0.715	0.907	0.800	0.801	0.865	0.832
7	0.799	0.667	0.727	0.746	0.832	0.786
8	0.767	0.623	0.688	0.788	0.700	0.741
9	0.753	0.766	0.759	0.745	0.844	0.792
10	0.745	0.736	0.740	0.735	0.804	0.768
Ave	**0.801**	0.736	0.767	0.747	**0.800**	**0.772**

is correct. We compared the work in this paper with BERT, and the relevant comparison results are shown in Table 3 ($\alpha = 1000$, and $\beta = 3000$). Among them, the BERT model uses unsupervised learning methods, while this paper uses semi-supervised learning methods.

We can see from Table 3 that, in 8 out of the 10 test datasets, the *Precision* values of the BERT model are slightly better than those of our directed weighted model, and the mean value is also greater than that of the directed weighted model. However, the *Recall* results of 8 out of the 10 test datasets by BERT were worse than those of the model in this paper. The average recall of the model in this paper is also greater r than that of the BERT model. This is mainly because the BERT model can find keywords more accurately, but for the case of multiple sentiments, it is relatively weak and with more missing words. Although the accuracy of the model proposed in this paper is slightly smaller than that of the BERT model, it is able to find the keywords that express weak sentiments by ours model. Therefore, ours model can give more sentiment classification, and the absence of keywords is relatively small. The main factor which may cause this result is the training mode of the model. Our model uses a semi-supervised method, which is controlled by manual labeling and threshold methods. While the BERT model adopts unsupervised learning, which is not supported by a large amount of common-sense background knowledge of human beings. What the BERT model learns is the features and representations of the sample space, which can be regarded as a large text matching model. A large amount of background knowledge is implicit and fuzzy, which is difficult to reflect in the pre-training data. The directed weighted model is based on entities with attributes, which are the most objective expression of sentiments. Most non-emotional entities and words are effectively filtered out through entity-with-attributes extraction. As a result, fewer keywords will be lost than with BERT.

α and β are the two threshold parameters of the shortest path. Different threshold values lead to different shortest paths and then generate different classifications. The results based on different thresholds are shown in Table 4.

We can see from Table 3 that the α and β of the thresholds are very important for the accuracy of the directed weighted model. The values of α and β must be at a certain matching value so that the accuracy of the model can reach the maximum. When one of the two parameters is adjusted individually, the accuracy also changes simultaneously. The smaller the parameter is, the lower the accuracy is. The reason for this is that some nodes that are not of strong sentiment are not added to the classification node set. When the parameter is large, the accuracy of the model will decrease after reaching a maximum value, which is due to many nodes with weaker sentiment being added to the classification set. Therefore, there is a maximum point in model training that is the best match of α and β.

TABLE 4. Accuracy of directed weighted model with different α and β.

α	β	Average		
		Precision	Recall	F value
500	2000	0.364	0.334	0.348
500	3000	0.476	0.313	0.378
500	5000	0.489	0.472	0.480
1000	2000	0.699	0.656	0.677
1000	**3000**	**0.747**	**0.800**	**0.772**
1000	5000	0.714	0.760	0.737
1500	2000	0.590	0.720	0.649
1500	3000	0.641	0.611	0.625
1500	5000	0.725	0.703	0.714
2000	3000	0.423	0.611	0.500
2000	5000	0.531	0.593	0.560
2000	7000	0.422	0.609	0.498

2) EFFICIENCY OF THE ALGORITHM

The efficiency of the algorithm is closely related to the complexity of the directed graph. The thresholds of α and β are two different parameters that control the complexity of the directed graph. Under the two parameters, we conduct the following three experiments. All experiments are performed on a PC workstation with dual 2.3GHz Xeon CPU, 64GB RAM, NVIDIA Quadro P4000 and Tensorflow for GPU 2.0.

In the experiment 1, we select the best case in accuracy experiments, that is on $\alpha = 1000$ and $\beta = 3000$. The datasets containing different number samples are used for testing. The numbers of samples contained in these datasets are 100, 200, and 300, …, 1000 respectively. The method of the sample selection is: randomly selecting 100 samples from the whole dataset as the first test samples, then randomly selecting another 100 sample from the remaining samples. Next, adding these samples to the previous dataset as the second test samples. Continue the same pipeline until all samples are selected. The efficiency on CPU time of the algorithm and the BERT [55] is shown in figure 2.

In the experiment 2, we set α ($\alpha = 1000$) as a fixed value, and β to 2000, 3000 and 5000 respectively. In the experiment 3, we set β ($\beta = 3000$) as a fixed value, and α to 500, 1000, 1500 and 2000 respectively. We use the dataset division method in the previous accuracy experiment for the two efficiency tests. We divide the whole samples into 10 datasets on average. Each dataset contains 100 samples. We select 9 datasets and remove 1 dataset to constitute a new training sample dataset. The new sample dataset includes 900 samples. The removed dataset is different at each time. Thus there are 10 training datasets totally. The efficiency on

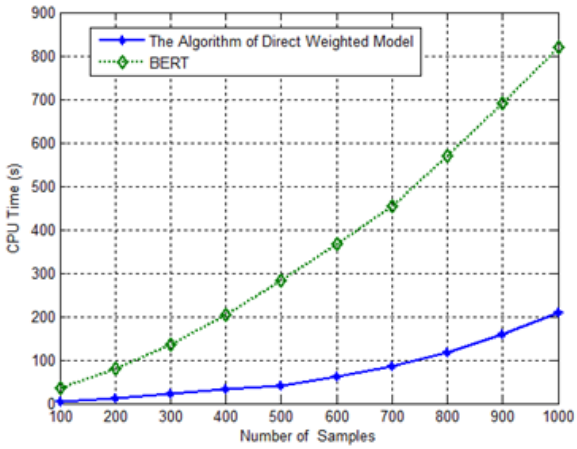


FIGURE 2. CPU time of the algorithm of direct weighted model and the BERT.

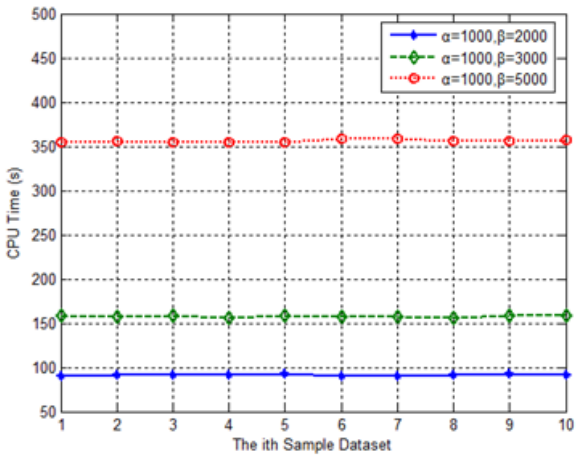


FIGURE 3. CPU time of the algorithm of direct weighted model on $\alpha = 1000$ and $\beta = 2000, 3000, 5000$.

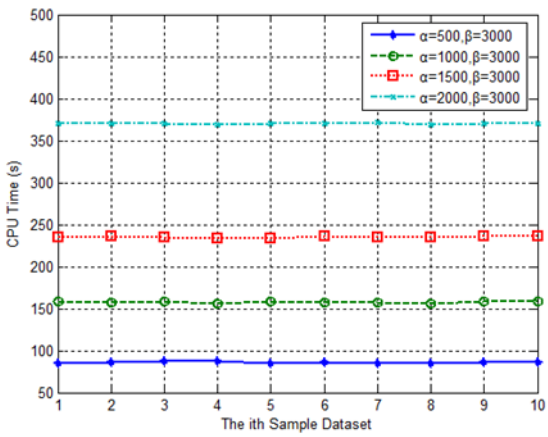


FIGURE 4. CPU time of the algorithm of direct weighted model on $\alpha = 500, 1000, 1500, 2000$ and $\beta = 3000$.

CPU time under these thresholds with 10 different training datasets is shown in figure 3 and 4.

We can see from figure 2 that in the case of the same sample size, the efficiency of CPU time of the algorithm proposed in this paper is significantly better than the method using BERT. The main reason is that the thresholds in the paper can quickly end the loops of algorithm. We can see from figure 3 and 4 that, the thresholds have an important impact on efficiency. When α takes a fixed value, the larger the value of β, the more CPU time for the algorithm. This is because when the similarity between two nodes is limited to a certain range, the longer the path length, the more the algorithm loops. When β takes a fixed value, α has a great influence on CPU time. The smaller the α, the less the algorithm CPU time; conversely, the larger the α, the more the algorithm CPU time. This is mainly because a larger α leads to more nodes with sentiment similarity into the consider space. Then the complexity of the model increases, and the algorithm CPU time also increases. Under the demand of a good accuracy, an appropriate value of α and β should be taken.

VII. CONCLUSION AND FUTURE WORK

Here, we addressed the problem of multiclassification of sentiment for E-commerce reviews. We introduced a sentiment classification method based on a directed weighted model. We turn the problem of sentiment similarity into the problem of shortest path computation through the extraction of entity words with attributes, the analysis of sentiment similarity relations, and the calculation of shortest paths between nodes. The model adopts the basic theory and method of nodes, links, and paths in a directed weighted graph. It represents the sentiment entity vocabulary as the sentiment nodes and the relations between nodes as the directed weighted links. The sentiment entity vocabulary is a type of entity with attributes, which is extracted through entity sentiment word pair analysis. Directed weighted links represent the sentiment similarity between two nodes of entities with attributes and are determined by the direct correlation calculation between them. The paths are all connected directed links from one node to another, which are composed of several nodes and links with high sentiment similarity.

Due to the complexity of human natural language, the content of user reviews exhibits great complexity and uncertainty. It is very difficult to distinguish the multilevel sentiments expressed by users in reviews. Future research will focus on how to level the core objectives of users' attentions. This type of division should consider the relation with the entity itself in detail, besides, it should clarify the hierarchical relation between entities. The purpose of hierarchy is to facilitate more detailed division of users' sentiments. Therefore, how to divide levels and how to distinguish the same level of content and different levels of content will be the focus of the future research.

REFERENCES

[1] S. Bhatia, P. Chaudhary, and N. Dey, *Opinion Mining in Information Retrieval*, 1st ed. Berlin, Germany: Springer-Verlag, 2020, pp. 1–22. [Online]. Available: https://link.springer.com/book/10.1007/978-981-15-5043-0

[2] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 1st ed. New York, NY, USA: Cambridge Univ. Press, 2015.

[3] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Berlin, Germany: Springer-Verlag, 2008.

[4] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, N. Indurkhya and F. J. Damerau, Eds., 2nd ed. Boca Raton, FL, USA: CRC Press, 2010.

[5] E. Cambria, M. Grassi, A. Hussain, and C. Havasi, "Sentic computing for social media marketing," *Multimedia Tools Appl.*, vol. 59, no. 2, pp. 557–577, Jul. 2012.

[6] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

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

[8] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar. 2016, doi: 10.1109/MIS.2016.31.

[9] J. R. Hobbs and E. Riloff, "Information extraction," in *Handbook of Natural Language Processing*, N. Indurkhya and F. J. Damerau, Eds. Boca Raton, FL, USA: CRC Press, 2010, pp. 1–31.

[10] R. C. Bunescu and R. J. Mooney, "Subsequence kernels for relation extraction," in *Proc. 18th Int. Conf. on Neural Inf. Process. Syst.* Cambridge, MA, USA: MIT Press, Dec. 2005, pp. 71–178.

[11] S. Sarawagi, "Information extraction," *Found. Trends Databases*, vol. 1, no. 3, pp. 261–377, 2008.

[12] A. Go, R. Bhayani, and L. Huang. "Twitter sentiment classification using distant supervision," Stanford, CA, USA, Tech. Rep. CS224N, 2009. [Online]. Available: https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf

[13] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[14] A. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," *Comput. Intell.*, vol. 22, no. 2, pp. 110–125, May 2006.

[15] K. Sarkar, "Using character N-gram features and multinomial Naïve Bayes for sentiment polarity detection in Bengali tweets," in *Proc. 5th Int. Conf. Emerg. Appl. Inf. Technol. (EAIT)*, Kolkata, India, Jan. 2018, pp. 12–13, doi: 10.1109/EAIT.2018.8470415.

[16] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. New York, NY, USA: ACM, 2004, pp. 168–177.

[17] S. Naz, A. Sharan, and N. Malik, "Sentiment classification on Twitter data using support vector machine," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Santiago, Chile, Dec. 2018, pp. 3–6, doi: 10.1109/WI.2018.00-13.

[18] J. Ding, H. Sun, X. Wang, and X. Liu, "Entity-level sentiment analysis of issue comments," in *Proc. 3rd Int. Workshop Emotion Awareness Softw. Eng. (SEmotion)*, Gothenburg, Sweden, 2018, pp. 7–13.

[19] Y. Xia, E. Cambria, A. Hussain, and H. Zhao, "Word polarity disambiguation using Bayesian model and opinion-level features," *Cognit. Comput.*, vol. 7, no. 3, pp. 369–380, Jun. 2015, doi: 10.1007/s12559-014-9298-4.

[20] L. Zhang, Y. Zhou, R. Chen, and X. Duan, "Sentiment-polarized word embedding for multi-label sentiment classification," in *Proc. IEEE 4th Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Dec. 2018, pp. 7–10, doi: 10.1109/CompComm.2018.8780754.

[21] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool, May 2012. [Online]. Available: https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

[22] P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2001, pp. 417–424.

[23] M. Venugopalan and D. Gupta, "Exploring sentiment analysis on Twitter data," in *Proc. 8th Int. Conf. Contemp. Comput. (IC3)*, Noida, India, Aug. 2015, pp. 241–247.

[24] G. Li and F. Liu, "A clustering-based approach on sentiment analysis," in *Proc. IEEE Int. Conf. Intell. Syst. Knowl. Eng.*, Hangzhou, China, Nov. 2010, pp. 331–337.

[25] R. Liu, R. Xiong, and L. Song, "A sentiment classification method for Chinese document," in *Proc. 5th Int. Conf. Comput. Sci. Edu.*, Hefei, China, Aug. 2010, pp. 918–922.

[26] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proc. 7th Conf. Int. Lang. Resour. Eval. (LREC)*, Valletta, Malta, May 2010, pp. 1320–1326.

[27] H.-Y. Kao and Z.-Y. Lin, "A categorized sentiment analysis of Chinese reviews by mining dependency in product features and opinions from blogs," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, Toronto, ON, Canada, vol. 1, Aug. 2010, pp. 456–459.

[28] D. Zhang, S. Li, Q. Zhu, and G. Zhou, "Multi-modal sentiment classification with independent and interactive knowledge via semi-supervised learning," *IEEE Access*, vol. 8, pp. 22945–22954, 2020, doi: 10.1109/ACCESS.2020.2969205.

[29] K. Ghag and K. Shah, "Comparative analysis of the techniques for sentiment analysis," in *Proc. Int. Conf. Adv. Technol. Eng. (ICATE)*, Mumbai, India, Jan. 2013, pp. 1–7, doi: 10.1109/ICAdTE.2013.6524752.

[30] S. M. Liu and J.-H. Chen, "A multi-label classification based approach for sentiment classification," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1083–1093, Feb. 2015.

[31] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis on Twitter: Classification performance and challenges," *Big Data Mining Anal.*, vol. 2, no. 3, pp. 181–194, Sep. 2019, doi: 10.26599/BDMA.2019.9020002.

[32] K. H.-Y. Lin, C. Yang, and H.-H. Chen, "Emotion classification of online news articles from the Reader's perspective," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, Sydney, NSW, Australia, Dec. 2008, pp. 220–226, doi: 10.1109/WIIAT.2008.197.

[33] M. Bouazizi and T. Ohtsuki, "A pattern-based approach for multi-class sentiment analysis in Twitter," *IEEE Access*, vol. 5, pp. 20617–20639, 2017, doi: 10.1109/ACCESS.2017.2740982.

[34] M. Bouazizi and T. Ohtsuki, "Sentiment analysis: From binary to multi-class classification: A pattern-based approach for multi-class sentiment analysis in Twitter," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[35] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis in Twitter: What if classification is not the answer," *IEEE Access*, vol. 6, pp. 64486–64502, 2018, doi: 10.1109/ACCESS.2018.2876674.

[36] J. Tao and X. Fang, "Toward multi-label sentiment analysis: A transfer learning based approach," *J. Big Data*, vol. 7, no. 1, pp. 1–26, Dec. 2020, doi: 10.1186/s40537-019-0278-0.

[37] D. Li and J. Qian, "Text sentiment analysis based on long short-term memory," in *Proc. 1st IEEE Int. Conf. Comput. Commun. Internet (ICCCI)*, Wuhan, China, Oct. 2016, pp. 471–475.

[38] J. Kikuchi and V. Klyuev, "Gathering user reviews for an opinion dictionary," in *Proc. 18th Int. Conf. Adv. Commun. Technol. (ICACT)*, Pyeongchang, South Korea, Jan. 2016, pp. 566–569.

[39] Z. Hameed, B. Garcia-Zapirain, and I. O. Ruiz, "A computationally efficient BiLSTM based approach for the binary sentiment classification," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Ajman, United Arab Emirates, Dec. 2019, pp. 1–4, doi: 10.1109/ISSPIT47144.2019.9001781.

[40] Z. Hameed and B. Garcia-Zapirain, "Sentiment classification using a single-layered BiLSTM model," *IEEE Access*, vol. 8, pp. 73992–74001, 2020, doi: 10.1109/ACCESS.2020.2988550.

[41] J. Akaichi, "Social networks 'Facebook' statutes updates mining for sentiment classification," in *Proc. Int. Conf. Social Comput.*, Alexandria, VA, USA, Sep. 2013, pp. 886–891.

[42] F. Colace, M. D. Santo, and L. Greco, "Sentiment mining through mixed graph of terms," in *Proc. 17th Int. Conf. Netw.-Based Inf. Syst.*, Salerno, Italy, Sep. 2014, pp. 324–330.

[43] S. Zhang and H. Zhong, "Mining users trust from E-Commerce reviews based on sentiment similarity analysis," *IEEE Access*, vol. 7, pp. 13523–13535, 2019.

[44] R. Ghani, K. Probst, Y. Liu, M. Krema, and A. Fano, "Text mining for product attribute extraction," *ACM SIGKDD Explor. Newslett.*, vol. 8, no. 1, pp. 41–48, Jun. 2006, doi: 10.1145/1147234.1147241.

[45] P. Pantel, E. Crestan, A. Borkovsky, A.-M. Popescu, and V. Vyas, "Web-scale distributional similarity and entity set expansion," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Singapore, 2009, pp. 938–947.

[46] Z. Chen, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Leveraging multi-domain prior knowledge in topic models," in *Proc. Int. Joint Conf. Artif. Intell.*, Beijing, China, 2013, pp. 2071–2077.

[47] Princeton University. *About WordNet*. [Online]. Available: https://wordnet.princeton.edu/

[48] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[49] D. Soyusiawaty and Y. Zakaria, "Book data content similarity detector with cosine similarity (Case study on digilib.Uad.AC.id)," in *Proc. 12th Int. Conf. Telecommun. Syst., Services, Appl. (TSSA)*, Yogyakarta, Indonesia, Oct. 2018, pp. 1–6, doi: 10.1109/TSSA.2018.8708758.

[50] J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image-based recommendations on styles and substitutes," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, Santiago, Chile, 2015, pp. 43–52.

[51] J. Ni, J. Li, and J. McAuley, "Justifying recommendations using distantly-labeled reviews and fine-grained aspects," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, Hong Kong, 2019, pp. 188–197.

[52] J. Ni. (2018). *Amazon Review Data*. [Online]. Available: https://nijianmo.github.io/amazon/index.html

[53] T. Zhang, K. Zhu, and D. Niyato, "A generative adversarial learning-based approach for cell outage detection in self-organizing cellular networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 171–174, Feb. 2020, doi: 10.1109/LWC.2019.2947041.

[54] L. Dong, W. Wang, and X. She, "An improved algorithm of Chinese comments opinion mining based on adverbs," in *Proc. 7th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Beijing, China, Aug. 2016, pp. 211–214, doi: 10.1109/ICSESS.2016.7883051.

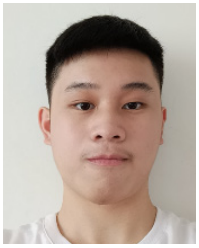
[55] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, Minneapolis, Minnesota, Jun. 2019, pp. 4171–4186, doi: 10.18653/v1/n19-1423.

SHAOZHONG ZHANG was born in Jinzhou, Liaoning, China, in 1969. He received the B.S. degree in computer science from Shenyang University, Jianzhu, in 1991, the M.S. degree in world economy from Nankai University, Tianjing, China, in 1999, and the Ph.D. degree in computer science from the Dalian University of Technology, Dalian, Liaoning, China, in 2004.

From 2006 to 2008, he was a Postdoctoral Researcher with Beihang University, Beijing, China. He was a Senior Visiting Scholar with Dalarna University from 2015 to 2016. Since 2008, he has been a Professor with the College of Electronic and Computer Science, Zhejiang Wanli University, China. He is the author of two books and more than 80 articles. His research interests include data mining and knowledge discovery, sentiment analysis, big data mining, and e-commerce and business intelligence.

DINGKAI ZHANG was born in Jinzhou, Liaoning, China, in 1996. She received the B.S. degree in computer science from Zhejiang Sci-Tech University, Zhejiang, China, in 2016, and the M.S. degree in computer information technology from Purdue University, Hammond, IN, USA, in 2020. Her main current research interests include machine learning, sentiment analysis, and big data mining.

HAIDONG ZHONG was born in Wuhan, Hubei, China, in 1982. He received the B.S. degree in computer science and technology, in 2004, and the Ph.D. degree in cartography and geography information system from East China Normal University, in 2011. He is currently an Associate Professor with the College of Logistics and E-Commerce, Zhejiang Wanli University. His main current research interests include mobile e-commerce, personalized recommendation, and business intelligence.



GUORONG WANG was born in Wenzhou, Zhejiang, China, in 1999. He is currently pursuing the bachelor's degree with the College of Information and Intelligence Engineering, Zhejiang Wanli University. His main current research interests include big data mining and e-commerce.