Devising a Cross-Domain Model to Detect Fake Review Comments

Chen-Shan Wei[1], Ping-Yu Hsu[1], Chen-Wan Huang[1], Ming-Shien Cheng[2](✉), and Grandys Frieska Prassida[1]

[1] Department of Business Administration, National Central University, No. 300, Jhongda Road, Jhongli City, Taoyuan County 32001, Taiwan (R.O.C.)
984401019@cc.ncu.edu.tw

[2] Department of Industrial Engineering and Management, Ming Chi University of Technology, No. 84, Gongzhuan Road, Taishan District, New Taipei City 24301, Taiwan (R.O.C.)
mscheng@mail.mcut.edu.tw

Abstract. The online reviews not only have huge impact on consumer shopping behavior but also online stores' marketing strategy. Positive reviews will have positive influence for consumer's buying decision. Therefore, some sellers want to boost their sales volume. They will hire spammers to write undeserving positive reviews to promote their products. Currently, some of the researches related to detection of fake reviews based on the text feature, the model will reach to high accuracy. However, the same model test on the other dataset the accuracy decrease sharply. Relevant researches had gradually explored the identification of fake reviews across different domains, whether the model built using comprehensive methods such as text features or neural networks, encountering the decreasing of accuracy. On the other hand, the method didn't explain why the model can be applied to cross-domain predictions. In our research, we using the fake reviews and truthful reviews from Ott et al. (2011) and Li, Ott, Cardie, and Hovy (2014) in the three domain (hotel, restaurant, doctor). The cross-domain detect model based on Stimuli Organism Response (S-O-R) combine LIWC (Linguistic Inquiry and Word Count), add word2vec quantization feature, overcoming the decreasing accuracy situation. According to the research result, in the method one SOR calculation of feature weight of reviews, the DNN classification algorithm accuracy is 63.6%. In the method two, calculation of frequent features of word vectors, the random forest classification algorithm accuracy is 73.75%.

Keywords: Fake reviews · Stimuli-Organism-Response (S-O-R) framework · word2vec

1 Introduction

A study by Klaus and Changchit (2017) proposed that when the Internet review comment system is reliable, consumers' purchase decisions reflect this system. Thus, electronic commerce companies focus heavily on comment quality management to

M. Hernes et al. (Eds.): ICCCI 2020, CCIS 1287, pp. 714–725, 2020.
https://doi.org/10.1007/978-3-030-63119-2_58

prevent malicious or fake comments from reducing the quality of the comment system and influencing consumers' purchase decisions.

Studies on identifying true and fake comments have used language types to extract their characteristics, which results in over-reliance on terminology from the same batch of data. Thus, once the data are changed to those of another batch, the accuracy of original identification methods performing well decreases substantially. Universal identification rules cannot be formed, and future applications are limited.

For scholars currently conducting interdisciplinary research, Li, Ott, Cardie, and Hovy (2014) developed general rules for data collected by Ott et al. (2011). Word calculation was implemented using language characteristic models from unigram, linguistic inquiry and word count (LIWC), and part of speech (POS), and evaluation rules were established combining support vector machines (SVMs) and the synthetic, augmentative, generative, experiential (SAGE) model. Ren and Ji (2017) used an artificial neural network (ANN) to identify fake comments. Finally, increases in interdisciplinary accuracy were higher than those in the experiment results of Li et al. (2014). However, no clear explanation has emerged for why words can be applied to interdisciplinary predictions.

Aslam et al. (2019) proposed that although fake comments in the filters of Yelp have similar language with that of true comments on the platform, the fake comments in the Yelp filters still leave psychological tracks when written. Therefore, this study applied psychological theories and the stimuli-organism-response (S-O-R) framework and established a classification model applicable across disciplines. Thus, the problem of substantially reduced accuracy during interdisciplinary identification from former studies can be overcome.

This paper organized as follow: (1) Introduction: research background, motivation and purpose. (2) Related work: review of scholars researches on the interdisciplinary identification of fake comments and stimulus-organism-response (S-O-R) framework. (3) Research methodology: content of research process in this study. (4) Research result: two method results and discussion. (5) Conclusion and future research: contribution of the study, and possible future research direction is discussed.

2 Related Work

2.1 Studies on Interdisciplinary Identification of Fake Comments

Previously, during interdisciplinary identification of true and fake comments, Li et al. (2014) attempted to capture general differences between fake and true comments from language usage. These differences were used to help clients make purchase decisions and maintain the quality of platform comment systems. The three types of language characteristic models comprised LIWC, POS, and unigram, and the two analysis models comprised SVM and SAGE. The study applied psychological theories but yielded unfavorable results. The maximum accuracy in the interdisciplinary section reached 78.5%, which was much lower than predictions within the same domain. In addition, Ren and Ji (2017) used an ANN, which presented relatively favorable identification of subtle language characteristics for identifying fake comments. The general regression neural network model had the most favorable performance and reached 83.5% accuracy in the restaurant domain and 57% in the doctor domain, which were both higher than the 78.5% and 55% by Li et al. (2014). However, models were established through language characteristic methods without explained reasons for interdisciplinary identification. Therefore, this study used psychological S-O-R theory to establish an interdisciplinary model.

2.2 Stimulus-Organism-Response (S-O-R) Framework

S-O-R theory was initially proposed by psychologist Woodworth in 1929. It posits that environmental stimuli change people's minds or individual statuses, and behavioral intentions or avoidance responses are generated. With the rise of online electronic commerce, studies have increasingly focused on stimulation from the Internet environment on consumers' purchases. Eroglu, Machleit, and Davis (2001) studied high and low correlation clues influencing online consumers' behavior. In addition, atmospheric stimulation was defined as cumulative hints seen and heard on the website, such as words, links, colors, animations, and sounds. Studies have indicated that Internet interfaces influence the emotions and cognitive status of Internet consumers and thus influence purchase behavior.

Apart from discussing stimulation's influence on consumers, some scholars have studied stimulation's internal influence on individuals. Adelaar, Chang, Lancendorfer, Lee, and Morimoto (2003) explored the relationship between the media's influence on personal emotions and impulsive purchases. The PAD (pleasure, arousal, and dominance) emotional state model and ANCOVA tests were used to reveal that personal emotions are positively correlated with impulse purchases. Menon and Kahn (2002) discussed the Internet atmosphere's influence on consumers' emotions and subsequent behavior. Experiencing pleasurable feelings when online shopping positively influences consumers' behavior.

For Internet environments, S-O-R theory is primarily the basic structure for discussing how stimulation factors on websites influence consumers' purchase behavior. With changes in consumers' purchase behavior, Internet comments have become a factor changing consumers' behavior and an indispensable link of the consumer purchase process. With Internet comments confirmed as a factor stimulating consumers' behavior, this paper discusses an identification model for fake comments from the S-O-R theory framework.

3 Research Methodology

According to the study structure, to establish an interdisciplinary identification model for comments, the following steps were conducted. Figure 1 presents the study structure and process. First, the comment databases established by Ott et al. (2011) and Li et al. (2014) were collected as original data of true and fake comments. The S-O-R psychological theory was used as the background, and writers' psychological states when writing comments were simulated. Related weighting values were calculated, and two methods were derived. Method 1 combines the 2015 LIWC dictionary with S-O-R theory and selects 14 categories of conforming dictionaries for the first comment study. Each comment's weighting values corresponding to the 14 categories were calculated. To study the relationship between selected words in the comments and the 14 categories from dictionaries, this study extended to Method 2. A second interdisciplinary identification method was conducted through frequency characteristic extraction of word2vec word vectors to achieve the expected method results.

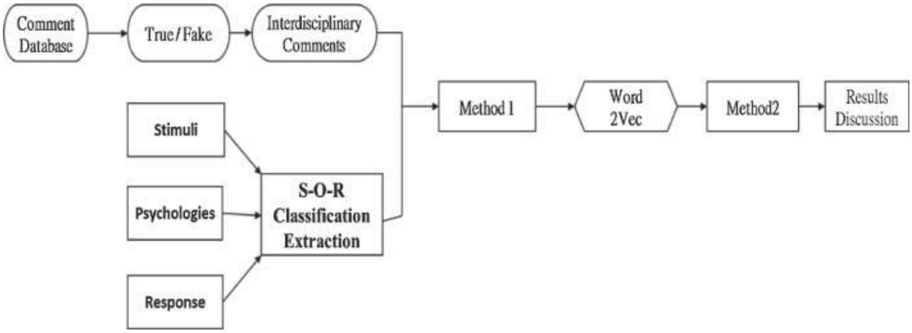

Fig. 1. Research process

3.1 S-O-R Category Filter Method

The LIWC dictionary was chosen to select categories conforming to definitions from the S-O-R framework. Gatautis and Vaiciukynaite (2013) stated that Internet elements in a virtual environment comprise website design, Internet communication elements, website content, and website instructions. Consumers' purchase results are easily influenced by website content. Comments on the Internet are also included in this. The five sensory processes are the receiving media for users to absorb information. Consequently, the perceptual process dictionary and its subcategories see, hear, and feel in LIWC were used.

In the S-O-R framework, organism is an internal process between external stimulation and final personal choice. we noted that activities include cognition, physiology, feeling, and thought. Emotions created after an individual internalizes information influence consumption behavior. Wolfe's psychological theory divides an individual's mind into two sections: desire and cognition. From this, the creation of online shopping experience is related to the external stimulation of individual cognition. Thus, this study used the cognitive process dictionary and the subcategories insight, causation, discrepancy, tentative, certainty, and differentiation.

For desire, and Kavanagh, Andrade, and May (2005) proposed that the generation of desire often results in pleasure or unhappy emotions. Boujbel and d'Astous (2015) defined desire as when consumers begin wanting to possess an object or behavior and begin imagining the sense of pleasure it brings after this desire is fulfilled and the sense of loss from when this desire is not satisfied. Because LIWC currently does not provide emotional dictionaries related to desire, this study implemented the WordNet expansion of desire-related dictionaries to identify hyponyms with desire as a vocabulary concept.

According to the word meaning of desire by Boujbel and d'Astous (2015), which is to strongly feel or have a desire for something or feel unsatisfied, the first and second desire dictionaries were constructed separately.

The response in the S-O-R framework represents the final outcomes and decisions of consumers, who can be approached or avoided. Chang, Eckman, and Yan (2011) defined approach behavior as positive actions or performance of an action such as staying, browsing, and making purchases. LIWC did not have dictionaries for reactions; therefore, the action dictionaries provided by Jeff Gardner were used to represent dictionaries of the reaction category.

3.2 Method 1: S-O-R and the Characteristic Weights of Comments

Before calculation, the comments underwent world pretreatment by the Natural Language Toolkit in Python. To pretreat English word exploration data, mostly stemming, lemmatization, normalization, and noise cancellation were implemented. In this study, only noise cancellation was adopted. Symbols or numbers that may have interfered with the analysis were deleted, all capital letters were converted into small letters, and words were segmented. The former three treatments were not adopted because the authenticity, writing styles, and words used in manual writing were hoped to be reserved in this study.

The characteristic weight values of every article were those calculated for the 14 categories selected from LIWC by using S-O-R theory. Consequently, 14 weight values resulted for each comment. This study used the R language to calculate values as fractions. The numerator was the number of words from a single category dictionary appearing in one comment, and the denominator was the total words used in that comment.

3.3 Method 2: Frequency Characteristics of Word Vectors

In Method 2, the relationship between words used in each comment and the 14 category dictionaries were discussed. Accordingly, the frequency characteristics of word vectors were developed. Word2vec was used to generate word vectors. After neural network training to learn a large number of articles, word2vec can simplify words into vectors in space. Thus, distances between vectors can be combined with the quality "words with close semantic meanings are also close in space" to represent similarity levels between word semantics.

The word2vec training database used in this study is an open English database downloaded from Wikipedia. Wikipedia data was downloaded in the xml format and converted into the text format. In addition, Python was adopted to train 300-dimension word vectors, and 1,500,000 words were ultimately trained. After the training was complete, every word could be expressed as a vector composed of 300 dimensions.

The average word vectors of a single comment and category were calculated using the word2vec word vector method. First, trained word vectors were used to calculate Euclidean distances between all words in one single comment and all words in a category. Then, the average word vector of one category could be calculated by measuring the average of all values. In addition, an extra cell, characteristic word extraction, was added. It can be a fake weight keyword or a truth weight keyword, and thus one weight word can be given according to the comment content. When conducting the classification algorithm, the S-O-R and comment characteristic weights from the original Method 1 were added with the word vector frequency characteristic values from Method 2.

4 Research Results

4.1 Data Collection

In the literatures of Ott et al. (2011) and Li et al. (2014), the review data of the three domains of hotel, restaurant and doctor were collected. Real reviews come from online review sites, fake reviews come from Amazon's mass intelligence platform (AMT) and experts and scholars.

Among the fake reviews, hotel domain: Ott et al. (2011) collected a total of 400 reviews all from AMT and Li et al. (2014) collected a total of 540 reviews of which 140 were written by experts and 400 articles were collected from the research of Ott et al. (2011); restaurant domain, a total of 320 reviews were collected, among them, 200 articles came from Amazon's AMT, 120 articles came from restaurant employees; doctors domain, a total of 232 reviews were collected, of which 200 articles came from Amazon's AMT and 32 articles came from hospital doctors.

In real reviews, Ott et al. (2011) collected a total of 400 reviews in the hotel sector, from TripAdvisor. Li et al. (2014) scholars collected 200 reviews each in the restaurant and doctor domains, all from TripAdvisor review sites. This study uses reviews from these two researches, the data collection can be seen in Table 1.

Table 1. Data collections

	True reviews	Fake review (AMT/expert)	Source
Hotel	400	400/140	Ott et al. (2011) Li et al. (2014)
Restaurant	200	200/120	Li et al. (2014)
Doctor	200	200/32	Li et al. (2014)

4.2 S-O-R Category Data

Some examples of the 14 categories and the number of words in each category are listed in Table 2.

Table 2. Volume of words of 14 categories

Categories	Example	No. of words
S (Stimulate)		
Perceptual	ache, aching, acid, acrid, appear…	433
see	appear, appearance, appeared, appearing…	126
hear	audibl, audio, boom, cellphone, choir…	92
feel	ache, aching, brush, burn, burned…	128
O (Organism)		
cogproc	abnormal, absolute, absolutely, accept…	797
insight	accept, accepts, accepted, accepting…	256
causation	activate, affect, affected, affecting, affects	133
Discrepancy	abnormal, besides, could, couldn't…	82
tentant	allot, almost, alot, ambigu, any…	179
certainty	absolute, absolutely, accura, all…	113
differentation	actually, adjust, against, aint, ain't…	77
desire1	abjure, abnegate, absence, ache…	351
desire2	ache, actuation, aim, allurement…	136
R (response)		
action	accomplish, achieve, attain, benefit…	32

4.3 Comments and S-O-R Word Characteristic Weights

The characteristic weights were calculated for every article in three domains. The weight value was between 0 and 1, and this showed the word usage weight ratio of comment in the category. Tables 3, 4 and 5 show the result.

4.4 Method 1: S-O-R and Characteristic Weights of Comments

The identification accuracy of Method 1 results are presented in Table 6. Eight classification algorithms were used: k-nn, naïve bayes (NB), decision tree, random forest (RF), gradient boosting machine, SVM, XGBoost (XGB), and deep neural networks (DNNs). DNNs had the highest performance in the hotel and restaurant domains with an accuracy of 62.08% and 63.6%, respectively.

4.5 Method 2: Word Vector Frequency Characteristics

Tables 7, 8 and 9 present an average word vector summary of the three domains, v1–v14 were the following 14 categories in order: action, differ, discrepancy, hear, certainty, see, feel, causation, desire2, tentant, insight, desire1, perceptual, and cogproc, D1–D400 were 400 fake comments, and T1–T400 were 400 true comments, D and T Average were the average distances of each category between fake and true comments.

Table 3. Average characteristic weights in hotel field

S (Stimulate)		
Category	Fake	True
insight	**0.0070**	0.0059
tentant	**0.0277**	0.0257
causation	**0.0061**	0.0049
Discrepancy	**0.0142**	0.0117
certainty	**0.0201**	0.0158
differ	0.0159	**0.0228**
desire1	**0.0228**	0.0218
desire2	**0.0110**	0.0073
cogproc	**0.0850**	0.0798
O (Organism)		
Category	Fake	True
Perceptual	**0.0177**	0.0146
see	**0.0097**	0.0076
hear	0.0015	**0.0031**
faal	**0.0054**	0.0036
R (response)		
Category	Fake	True
action	**0.0088**	0.0070

Table 4. Average characteristic weights in restaurant field

S (Stimulate)		
Category	Fake	True
insight	**0.0077**	0.0071
tentant	**0.0328**	0.0324
causation	**0.0066**	0.0056
Discrepancy	0.0109	**0.0115**
certainty	**0.0230**	0.0150
differ	0.0221	**0.0298**
desire1	**0.0228**	0.0182
desire2	**0.0099**	0.0067
cogproc	**0.0965**	0.0930
O (Organism)		
Category	Fake	True
Perceptual	0.0175	**0.0188**
see	0.0060	**0.0076**
hear	0.0023	**0.0038**
faal	**0.0035**	0.0030
R (response)		
Category	Fake	True
action	0.0088	**0.0089**

Table 5. Average characteristic weights in doctor field

S (Stimulate)		
Category	Fake	True
insight	0.0146	**0.0187**
tentant	0.022	**0.0330**
causation	0.0108	**0.0110**
Discrepancy	0.0144	**0.0154**
certainty	0.0178	**0.0214**
differ	0.0180	**0.0235**
desire1	0.0145	**0.0209**
desire2	**0.0162**	0.0121
cogproc	**0.0408**	0.1138
O (Organism)		
Category	Fake	True
Perceptual	0.0180	**0.0215**
see	**0.0099**	0.0086
hear	0.0066	**0.0049**
faal	**0.011**	0.0076
R (response)		
Category	Fake	True
action	**0.0115**	0.0158

Table 6. The Method 1 result

Classification algorithm	Hotel	Restaurant	Doctor	Average
KNN	54.16%	57.02%	55.35%	55.51%
NB	59.15%	61.98%	60.11%	60.41%
DT	57.08%	57.02%	60.71%	58.27%
RF	59.16%	59.50%	61.90%	60.19%
GBM	57.08%	57.85%	61.30%	58.74%
SVM	47.91%	47.93%	63.09%	52.98%
XGB	58.33%	60.33%	63.09%	60.58%
DNN	62.08%	63.60%	61.90%	62.53%

The average word vector tables of the hotel and restaurant domains revealed that the values of D Average were smaller than those of T Average, which signified that words used for writing fake comments were closer to the categories selected with S-O-R theory. In the original assumptions, fake comment writers may have used the S-O-R

Table 7. Average word vector summary in hotel domain

Hotel	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14
D1	3.692	3.686	3.612	3.856	3.572	3.802	4.017	3.804	3.697	3.748	3.697	3.854	3.940	3.703
D2	3.692	3.693	3.635	3.857	3.587	3.799	4.014	3.810	3.704	3.752	3.700	3.855	3.942	3.710
D3	3.604	3.599	3.504	3.733	3.472	3.686	3.908	3.721	3.580	3.638	3.592	3.743	3.819	3.602
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
D398	3.840	3.827	3.749	3.970	3.719	3.930	4.098	3.936	3.842	3.885	3.862	3.980	4.031	3.849
D399	3.644	3.616	3.572	3.805	3.508	3.719	3.946	3.735	3.638	3.689	3.644	3.790	3.869	3.644
D400	3.755	3.742	3.683	3.892	3.610	3.789	4.002	3.853	3.727	3.808	3.771	3.875	3.936	3.762
T1	3.703	3.669	3.641	3.836	3.567	3.755	3.969	3.800	3.687	3.744	3.713	3.831	3.900	3.705
T2	3.712	3.679	3.642	3.835	3.576	3.756	3.979	3.797	3.689	3.752	3.714	3.841	3.901	3.710
T3	3.765	3.738	3.718	3.893	3.654	3.819	4.026	3.856	3.751	3.813	3.782	3.895	3.958	3.776
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
T398	3.779	3.752	3.710	3.888	3.654	3.830	4.044	3.864	3.760	3.814	3.779	3.905	3.966	3.777
T399	3.675	3.654	3.600	3.801	3.547	3.747	3.955	3.768	3.662	3.715	3.675	3.812	3.880	3.675
T400	3.762	3.736	3.705	3.885	3.653	3.821	4.039	3.856	3.748	3.809	3.773	3.893	3.961	3.771
D Average	**3.692**	**3.678**	**3.619**	**3.828**	***3.568***	**3.765**	**3.982**	**3.792**	**3.678**	**3.740**	**3.695**	**3.832**	**3.905**	**3.697**
T Average	3.729	3.704	3.662	3.860	3.605	3.793	4.008	3.822	3.717	3.773	3.735	3.865	3.933	3.733
|DA - DT|	0.037	0.026	0.043	0.032	0.037	0.028	0.026	0.03	0.039	0.033	0.04	0.033	0.028	0.036

Table 8. Average word vector summary in restaurant domain

Restaurant	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14
D1	3.654	3.640	3.585	3.792	3.532	3.738	3.937	3.751	3.635	3.704	3.655	3.790	3.858	3.659
D2	3.757	3.724	3.690	3.887	3.621	3.820	4.026	3.846	3.730	3.791	3.756	3.883	3.945	3.753
D3	3.699	3.680	3.614	3.816	3.547	3.768	3.970	3.800	3.662	3.725	3.690	3.823	3.887	3.691
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
D200	3.584	3.594	3.483	3.700	3.451	3.679	3.881	3.707	3.561	3.622	3.576	3.738	3.795	3.586
D201	3.596	3.603	3.501	3.710	3.479	3.693	3.885	3.732	3.582	3.634	3.599	3.756	3.800	3.606
D202	3.673	3.634	3.588	3.797	3.541	3.736	3.932	3.762	3.646	3.700	3.666	3.802	3.852	3.665
T1	3.726	3.713	3.641	3.832	3.604	3.800	4.020	3.822	3.694	3.762	3.714	3.853	3.920	3.722
T2	3.731	3.715	3.673	3.866	3.634	3.814	4.001	3.848	3.740	3.775	3.747	3.883	3.927	3.747
T3	3.830	3.841	3.782	3.948	3.746	3.900	4.053	3.947	3.815	3.889	3.867	3.966	3.974	3.860
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
T198	3.982	3.973	3.927	4.066	3.881	4.036	4.219	4.083	3.950	4.011	3.993	4.102	4.134	3.990
T199	3.930	3.890	3.866	4.044	3.809	3.977	4.130	4.013	3.902	3.956	3.945	4.044	4.057	3.930
T200	3.875	3.858	3.810	3.990	3.778	3.940	4.126	3.967	3.863	3.918	3.889	4.009	4.044	3.886
D Average	**3.736**	**3.719**	**3.655**	**3.858**	**3.609**	**3.809**	**4.002**	**3.835**	**3.712**	**3.770**	**3.736**	**3.869**	**3.919**	**3.736**
T Average	3.776	3.748	3.703	3.897	3.655	3.842	4.032	3.869	3.758	3.811	3.781	3.907	3.953	3.777
|DA - DT|	0.04	0.029	0.048	0.039	0.046	0.033	0.03	0.034	0.046	0.041	0.045	0.038	0.034	0.041

Table 9. Average word vector summary in doctor domain

Doctor	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14
D1	3.526	3.587	3.295	3.433	3.345	3.526	3.717	3.717	3.371	3.476	3.509	3.617	3.605	3.500
D2	3.857	3.816	3.764	3.965	3.694	3.923	4.133	3.912	3.818	3.879	3.805	3.962	4.055	3.823
D3	3.596	3.591	3.483	3.719	3.458	3.699	3.921	3.695	3.580	3.627	3.549	3.744	3.835	3.579
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
D355	3.597	3.608	3.494	3.747	3.470	3.722	3.948	3.701	3.591	3.655	3.561	3.756	3.861	3.593
D356	3.620	3.602	3.541	3.808	3.490	3.751	3.970	3.712	3.637	3.685	3.597	3.784	3.898	3.620
D357	3.727	3.705	3.628	3.831	3.586	3.780	3.984	3.823	3.693	3.747	3.698	3.851	3.913	3.711
T1	3.675	3.682	3.610	3.882	3.554	3.819	4.014	3.773	3.696	3.762	3.669	3.847	3.954	3.690
T2	3.536	3.551	3.451	3.676	3.405	3.648	3.866	3.658	3.529	3.581	3.510	3.705	3.785	3.536
T3	3.646	3.649	3.552	3.777	3.520	3.752	3.971	3.755	3.640	3.692	3.619	3.807	3.889	3.644
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
T199	3.640	3.628	3.535	3.766	3.503	3.733	3.947	3.746	3.609	3.673	3.602	3.768	3.868	3.626
T200	3.541	3.583	3.425	3.644	3.392	3.654	3.857	3.678	3.509	3.568	3.495	3.702	3.770	3.531
T201	3.523	3.561	3.422	3.662	3.403	3.672	3.870	3.672	3.526	3.572	3.501	3.708	3.790	3.532
D Average	3.625	3.608	3.535	3.782	3.484	3.736	3.945	3.714	3.616	3.674	3.597	3.772	3.873	3.616
T Average	**3.613**	**3.600**	**3.520**	**3.760**	**3.470**	**3.721**	**3.931**	**3.710**	**3.600**	**3.656**	**3.584**	**3.760**	**3.856**	**3.604**
|DA - DT|	0.012	0.008	0.015	0.022	0.014	0.015	0.014	0.004	0.016	0.018	0.013	0.012	0.017	0.012

Table 10. The Method 2 result

	Hotel	*Restaurant*	*Doctor*
KNN	66.66%	65.28%	69.64%
NB	68.75%	65.29%	66.66%
DT	63.75%	64.46%	65.47%
RF	73.75%	73.55%	71.42%

structure to promote products in the store. Thus, consumers accepted related stimulation and purchased products after reading the comments.

However, different outcomes were observed in the medical domain, and the D Average values were all larger than the T Average values in every category. Therefore, the words used in the true comments were related to selected dictionaries with S-O-R theory. The differences between D Average and T Average in the hotel, restaurant, and medical domains were 0.033, 0.039, and 0.014, respectively. This finding can be used as reference when discussing word usage differences when writing comments on the three domains.

Table 11. Result comparison of Method 1 and 2

Doctor	Method 1 (SOR characteristic weight)	Method 2 (SOR characteristic weight + Word vector)
KNN	55.35%	64.28%
NB	60.11%	60.07%
DT	60.71%	63.09%
RF	61.90%	66.07%
GBM	61.30%	67.26%
SVM	63.09%	67.26%
XGB	63.09%	69.06%
DNN	61.90%	63.25%
Average	60.93%	65.04%

Method 2 indicated the classification results of adding 15 words vector frequency characteristics with S-O-R characteristic values from Method 1. The method results are listed in Table 10. The classification algorithm RF are the greatest performance.

Table 11 presents the accuracy of true and fake comment identification in the medical domain after training the data from two other domains and excluding data from the medical domain. In the method results of Method 2, the XGB classification algorithm performed the most satisfactorily (69.06%), followed by GMB and SVM (67.26%). These results all were higher than the 55.7% produced by Ren and Ji (2017) by using ANNs and the 64.7% by Li et al. (2014) by using SAGE.

All classification algorithms were compared with Method 1 and listed. All algorithms of Method 2 except for NB exhibited higher performance than did those of Method 1. The average value of all classification algorithms was 60.93%, and this improved by 4.11% when Method 2 was used. From this experiment, the addition of word vector frequency characteristics could enhance the accuracy of the extraction of the word characteristic frequency used in true and fake comments. Thus, establishing an interdisciplinary identification model for fake comments can be accomplished.

5 Conclusion and Future Research

Online shopping has a new consumption model. Online reviews are of great significance to consumers, e-commerce platforms, and stores. Due to the habit of consumers to refer to reviews before buying, shop owners will write fake reviews to promote products as a boost. In the face of real and fake comments on the Internet, e-commerce platforms have an obligation to protect consumers and maintain the quality on the platform. Therefore, the ability to develop a system that can quickly identify the authenticity of online reviews is an important issue for online platforms.

This research is aimed at the reviews in the three domains of hotel, restaurant, and doctor, and attempts to develop a model that can discriminate true and fake reviews across domains. After using SOR theory to select the corresponding category fonts from the LIWC font to calculate the feature values, word2vec is used to increase the frequent features of the word vector, and it becomes a cross-domain model that can effectively discriminate between true and fake reviews.

Many deficiencies were discovered after the end of the research. We have provided 2 points below in terms of limitations and recommendations:

1. In this research, the labeled reviews are used to focus on the three areas of hotel, restaurant, and doctor. In the future, if data from other domains can be used as training materials for modeling, we can explore whether data in different domains can also be applied rule.
2. In future research, because the idioms used in each generation are different. Therefore, it is suggested to expand or update the related category fonts based on the SOR theory to maintain or improve the discrimination accuracy, and at the same time explore whether the category fonts will change over time.

References

Adelaar, T., Chang, S., Lancendorfer, K.M., Lee, B., Morimoto, M.: Effects of media formats on emotions and impulse buying intent. J. Inf. Technol. **18**(4), 247–266 (2003). https://doi.org/10.1080/0268396032000150799

Aslam, U., Jayabalan, M., Ilyas, H., Suhail, A.: A survey on opinion spam detection methods. Int. J. Sci. Technol. Res. **8**(9) (2019)

Boujbel, L., d'Astous, A.: Exploring the feelings and thoughts that accompany the experience of consumption desires. Psychol. Mark. **32**(2), 219–231 (2015). https://doi.org/10.1002/mar.20774

Chang, H.-J., Eckman, M., Yan, R.-N.: Application of the Stimulus-Organism-Response model to the retail environment: the role of hedonic motivation in impulse buying behavior. Int. Rev. Retail Distrib. Consum. Res. **21**(3), 233–249 (2011)

Eroglu, S.A., Machleit, K.A., Davis, L.M.: Atmospheric qualities of online retailing: a conceptual model and implications. J. Bus. Res. **54**(2), 177–184 (2001)

Gatautis, R., Vaiciukynaite, E.: Website atmosphere: towards revisited taxonomy of website elements. Econ. Manag. **18**(3) (2013)

Kavanagh, D.J., Andrade, J., May, J.: Imaginary relish and exquisite torture: the elaborated intrusion theory of desire. Psychol. Rev. **112**(2), 446 (2005)

Klaus, T., Changchit, C.: Toward an understanding of consumer attitudes on online review usage. J. Comput. Inf. Syst. **59**(3), 277–286 (2017). https://doi.org/10.1080/08874417.2017.1348916

Li, J., Ott, M., Cardie, C., Hovy, E.: Towards a general rule for identifying deceptive opinion spam. Paper Presented at the Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (2014)

Liu, W., Jing, W., Li, Y.: Incorporating feature representation into BiLSTM for deceptive review detection. Computing **102**(3), 701–715 (2019). https://doi.org/10.1007/s00607-019-00763-y

Menon, S., Kahn, B.: Cross-category effects of induced arousal and pleasure on the Internet shopping experience. J. Retail. **78**(1), 31–40 (2002)

Ott, M., Choi, Y., Cardie, C., Hancock, J.T.: Finding deceptive opinion spam by any stretch of the imagination. Paper Presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (2011)

Ren, Y., Ji, D.: Neural networks for deceptive opinion spam detection: an empirical study. Inf. Sci. **385–386**, 213–224 (2017). https://doi.org/10.1016/j.ins.2017.01.015