Review Spam Detection using Multi-View Deep Learning Combining Content and Behavioral Features

Giuseppina Andresini[1,2,*], Andrea Iovine[1], Roberto Gasbarro[1], Marco Lomolino[1], Marco de Gemmis[1] and Annalisa Appice[1,2]

[1]Department of Computer Science, University of Bari Aldo Moro, Bary,Italy
[2]Consorzio Interuniversitario Nazionale per l'Informatica - CINI, Bari, Italy

Abstract

Nowadays, online reviews are the main source to customer opinions. They are especially important in the realm of e-commerce, where reviews regarding products and services influence the purchase decisions of customers, as well as the reputation of the commerce websites. Unfortunately, not all the online reviews are truthful and trustworthy. Therefore, it is crucial to develop machine learning techniques to detect review spam. This study describes EUPHORIA – a novel classification approach to distinguish spam from truthful reviews. This approach couples multi-view learning to deep learning, in order to gain accuracy by accounting for the variety of information possibly associated with both the reviews' content and the reviewers' behavior. Experiments carried out on two real review datasets from Yelp.com – Hotel and Restaurant – show that the use of multi-view learning can improve the performance of a deep learning classifier trained for review spam detection.

Keywords

review spam detection, deep learning, multi-view learning, word embedding

1. Introduction

In the last two decades, the widespread diffusion of customer reviews has raised the risk of review spam attacks towards several commerce websites (e.g., Amazon.com, Yelp.com). Customer reviews are user-contributed consumer opinions posted to commerce websites and originated from the users' experiences regarding specific products or services. They represent the most valuable source of information that can be used to determine the public opinion on the reviewed products or services. In fact, customer reviews are one of the primary factors in a customer's decision to purchase a product or service [1]. Furthermore, there are increasing efforts to incorporate the rich information embedded in reviews into the process of user modeling and recommendation generation or justification [2, 3]. On the other hand, as anyone can easily produce opinions and post fake reviews, i.e., spam reviews, to social media with no constraints, certain product vendors or service providers may abuse this situation to promote their products and services, or to criticize their competitors unfairly. Due to the real risk of review spam

ITADATA2022: The 1st Italian Conference on Big Data and Data Science, September 20–21, 2022, Milan, Italy
*Corresponding author.
giuseppina.andresini@uniba.it (G. Andresini); andrea.iovine@uniba.it (A. Iovine); r.gasbarro1@studenti.uniba.it (R. Gasbarro); m.lomolino@studenti.uniba.it (M. Lomolino); marco.degemmis@uniba.it (M. d. Gemmis); annalisa.appice@uniba.it (A. Appice)

attacks, developing effective review spam detection approaches is a crucial task to secure the reliability of online opinions. Today, distinguishing spam reviews from truthful (non-spam) reviews is a challenging task that attracts growing attention in the machine learning community, since it is difficult, if not impossible, to recognize fake reviews by manually reading their content [4].

Although content features are widely investigated in the review spam detection literature, several studies assess that a purely content-based approach is not sufficient to train a review spam classifier with adequate accuracy performance [5, 4]. The emerging research trend is to improve the accuracy of review spam classifiers by taking additional features into account. These features include the review's post date/time, the behavior associated with the review's writer, and the deviation from other reviews concerning the same product or service [6]. Coupling review content features with reviewer behavioral features has proven to be more effective than using each type of feature alone [5, 4].

In this paper, we investigate the effects of jointly performing supervised learning on both content and behavioral features for the task of review spam detection. We define a novel review classification approach, named EUPHORIA (nEural mUlti-view aPproacH fOr RevIew spAm) that couples deep learning with multi-view learning, in order to elaborate knowledge pertaining to both the review content and the reviewer behavior. With regard to the feature extraction technique, we employ word embedding models used to derive a feature vector representation of the review content. This step is performed to capture complex global semantic information that is hidden in reviews and difficult to express using traditional bag-of-words features. In particular, we explore the performance of two word embedding models of the review content, namely Word2Vec [7] and BERT [8], and a temporal-aware representation of the reviewer behavior. In addition, we identify a set of behavioral reviewer features, based on the related literature, that can aid in enhancing the performance of the review content-based classification. Another distinctive characteristic of the proposed solution is that, with regard to the classification technique, we adopt a combination of deep learning and multi-view learning. In fact, we handle both content and behavioral features as multiple views of the same review corpus and propose a deep neural network architecture to learn an accurate classification model to distinguish spam reviews from non-spam reviews. Specifically, we train a multi-input neural network that is able to share knowledge among review content-based and reviewer behavior-based views. This architecture allows us to gain in classification performance. The original contributions of this work is reported in [9].

The paper is organized as follows. Section 2 illustrates the related work. Section 3 presents the proposed EUPHORIA approach. Section 4 describes the data collections processed in the experiments, the experimental setting and the relevant results. Finally, Section 5 draws conclusions and outlines the future directions of this work.

2. Related work

Several supervised approaches are formulated in the machine learning literature to distinguish spam reviews from non-spam reviews [4, 10]. The seminal studies in [11, 12] started the investigation of the review spam detection task in the context of product reviews. Starting

from these studies, many researchers have investigated the performance of several feature representation models adopted to describe structural properties of review content or delineate behavioral characteristics of reviewers. In particular, state-of-the-art approaches for review spam detection can be categorized into two main groups: methods based on features created from the review content and methods based on features created on the reviewer's behavior [13]. Content-based approaches extract features based on the review text. The majority of content-based approaches adopt the traditional bag-of-words model, which represents text as sets of words, ignoring their order. However, recent studies have started to explore word embedding models that allow us to overcome the limitations of the bag-of-words model by capturing complex global information contained in the text. For example, TopicSpam [14] uses the Latent Dirichlet Allocation (LDA) algorithm to identify slight differences between the distribution of the keywords in the spam and non-spam reviews. More recently, large pre-trained models have been employed for this task. An example of such model is BERT [8], a powerful Transformer-based encoder model trained to generate a bidirectional representation of text. BERT has already been proven to be successful in a variety of NLP tasks. The application of BERT on the review spam detection task has been studied in [15], which aims to capture the semantic relevance in the review's sentences. The experimental results reported in this study show that BERT can generate a text representation with richer content information compared to traditional text representation approaches based on bag-of-words/n-gram features.

Behavior-based approaches focus on studying the behavior of the reviewers, rather than analyzing their reviews as individual units [13]. Starting from the seminal study of [6], various behavioral features (e.g., maximum number of reviews per day, percentage of positive reviews, total number of reviews for a reviewer, content similarity) have been studied to improve the classification performance achieved with content features. A recently emerging research trend has combined the two approaches, improving the performance of classifiers trained to detect spam reviews [11, 13, 16, 17]. For example, in [16], both content features (e.g., uni-gram, bi-gram, similarity scores with other reviews) and behavioral features (e.g., authority score of a reviewer, rating deviation score) are analyzed for the review classification. Classification with both behavioral features and content features has been recently investigated also in [13] and [17].

Finally, few recent studies have shown the superiority of deep learning approaches compared to traditional classification methods in several problems of review spam detection [18, 19, 20]. In [19], a Recurrent Neural Network with Attention mechanism (GRNN) is trained to capture the non-local information over sentence vectors. In [18], a deep learning-based approach is used to learn a document-level representation to identify spam reviews. In particular, the approach combines Word2Vec and Convolutional Neural Networks in order to learn the representation of the reviews, and extract higher-level n-gram features of the review content. A Bi-directional LSTM is finally used for the review classification. The work in [20] uses word embeddings trained on a large Amazon review dataset using the Continuous Bag-of-Words (CBOW) algorithm, and trains a model that combines Convolutional Neural Networks and Gated Recurrent Neural Networks.

Even though recent literature in review spam detection strongly suggests that the analysis of the reviewer behavior covers a crucial role in improving review classification performance, prior studies exploring deep learning methods for review spam detection are mostly limited to

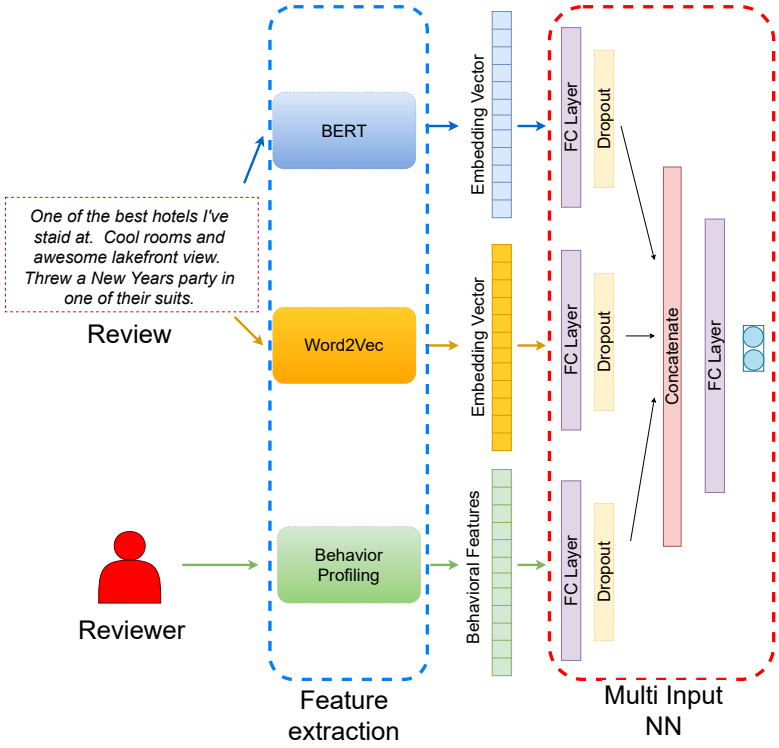

Figure 1: Schema of EUPHORIA. Abbreviations: FC = fully-connected.

processing the review content.

3. Proposed method

EUPHORIA uses the multi-view learning approach in order to detect fake reviews more effectively, by exploiting multiple (overlapping) views of the data. In our approach, each data view is represented as a distinct feature vector, which is then processed by the multi-input neural network. Currently, the model supports three different data views, two related to textual features (Word2Vec and BERT), and one related to behavioral features. Both Word2Vec and BERT are used to extract text-centric features from text. The two techniques represent to different *views* of the textual data, which focus on different aspects of the review text

Word2Vec is one of the first attempts to learn dense and continuous textual representations in the form of a numerical vector, which describes the meaning of a particular word. The meaning of a specific word is derived from the *context* that the word appears in, i.e., the words that commonly co-occur with it. To represent a sentence or a paragraph, word embeddings are combined through an *aggregation function*, where the representation of a review is obtained by averaging the word embeddings of all tokens contained in that review. For the purposes of this study, we decided to train a new set of word embeddings using Continuous Bag-of-Words (CBOW) algorithm. In the CBOW algorithm, a feed-forward single-layer neural network is

trained to predict a specific word, given a set of history words (which precede the word to be predicted) and a set of future words (which follow the word to be predicted). The size of the *context window*, i.e., the number of history and future words given as input is considered as a hyperparameter With these embeddings, we are able to obtain a more *domain-focused* view of the text data, which we can use to identify patterns that are specific to the review domain.

It is important to note that while the meaning of each word depends on its context (co-occurring words), the embedding of a specific word computed with Word2Vec does not change depending on the sentence the word appears in. This can cause problems with *polysemous* words, which have multiple meanings. *Contextualized word embeddings* may overcome this issue by adding the ability to produce different vector representations for the same word, based on the context of the entire sentence. BERT [8], that stands for *Bidirectional Encoder Representation from Transformers*, is a state-of-the-art pre-trained model for generating contextual representations of text. In particular, BERT is based on the bidirectional Transformer architecture [21], which features an encoder and a decoder with several multi-head attention layers. Due to its bidirectional and contextual nature, BERT can encode the meaning of an entire sentence/paragraph in a single vector. The vector representation of a sentence was constructed by using the final hidden state of the special *classification* (CLS) token, which is the common approach for representing sequences of text for classification purposes with BERT [8]. Finally, the vector representation of the entire review was obtained by averaging the vectors of each sentence.

Finally, we calculated six behavioral features, which represent the *profile* of the reviewer over the course of time. These features are *reviewer-centric*, as they consider all the reviews written by the reviewer in the specified time range.

The reviewer profile includes the following behavioral dimensions:

- Maximum Number of Reviews per Day (MRD), that measures the maximum number of reviews that the reviewer u has written in a single day up to the current time t. A reviewer who writes a large amount of reviews in a single day can thus be a potential threat.
- Positive Ratio (PR), that is the percentage of positive reviews written by a reviewer up to the time t. Specifically, we consider a review as *positive* if its rating is higher than 3 out of 5.
- Average Review Length (ARL), that is the average length of the reviews (number of words) written by u up to the time t. It has been noted that fake reviews tend to be generally shorter than legitimate ones.
- Reviewer Deviation (RD) that is defined as the absolute difference between the average rating obtained by the item, and the rating that u has assigned to the same product. In fact, it has been observed that the ratings of fake reviews tend to deviate much more frequently from the average rating.
- Average Review Similarity (ARS), that measures the similarity of all reviews written by u up to the time t. The idea is to detect reviewers who use templates to write large amounts of reviews, which will be similar in both syntax and meaning.
- Maximum Review Similarity (MRS), that measures the maximum similarity between the current review r and the reviews previously written by the same reviewer u. The similarity is again calculated as the cosine between the Word2Vec feature vectors of the

Table 1
Statistics on reviews and reviewers collected in Hotel and Restaurant datasets

review/reviewer	Hotel	Restaurant
Spam reviews (target service)	779	8301
Non-spam reviews (target service)	5078	58716
Spam reviews (every business)	267453	326979
Non-spam reviews (every business)	420785	461490
Reviewers	5123	16941

two reviews. Similarly to ARS, MRS can be used to detect reviewers that use standard templates to generate fake reviews.

4. Empirical evaluation and discussion

We evaluated the accuracy performance of the proposed approach by performing several experiments on two benchmark review datasets. These experiments aimed to investigate the achievements of the multi-view learning approach, as well as the accuracy performance of the deep learning architecture. The datasets are presented in Section 4.1. The implementation details of the trained neural network architecture are reported in Section 4.2. The experimental setting is described in Section 4.3, while the results are discussed in Section 4.4.

4.1. Data

Obtaining gold standard datasets for detecting spam review is a challenging problem. Due to the large amount of reviews online, manual labeling for ground truth reviews is complex and costly. In our study, we consider two datasets from Yelp.com, which are first used in [6]. Yelp.com is a well-known large-scale online review site that provides labeled datasets for the evaluation of review spam detection approaches. The two datasets, namely Hotel and Restaurant, contain reviews across 85 hotels and 130 restaurants, respectively, in the Chicago area. T Both datasets collect reviews for the target services of this experimental study, i.e., hotels and restaurants, as well as reviews obtained from the reviewers' profile pages for any product or service that they wrote a review for. In this study, we classify reviews of hotels and restaurants separately, as any domain has specific characteristics to take into account for review spam analysis.

However, we also investigate the performance that we can achieve by integrating the user profile with reviews coming from other domains. Dataset statistics are reported in Table 1. The class distribution is imbalanced in both datasets.

4.2. Implementation details

EUPHORIA was implemented in Python 3. In particular, the multi-input neural network architecture was realized using the high-level neural network API Keras 2.4 with TensorFlow as

Table 2
Hyper-parameter search space for the multi-input neural network

Hyper-parameter	Values
Mini-batch size	$\{2^4, 2^5, 2^6, 2^7, 2^8, 2^9\}$
Learning rate	[0.0001, 0.01]
Dropout	[0,1]
# of neurons per hidden layer	$\{2^6, 2^7, 2^8, 2^9\}$

back-end. For each dataset, we performed automatic hyper-parameter optimization using the tree-structured Parzen estimator algorithm, as implemented in the Hyperopt library.[1] Hyper-parameter optimization was performed by using 20% of the entire training as a validation set. We selected the configuration of the hyper-parameters that achieved the lowest validation set loss. The hyper-parameter search space is reported in Table 2. The standard Rectified Linear Unit (ReLU) was selected as the activation function for each hidden layer, while for the last layer the softmax activation function has been used. The neural network was trained with mini-batches by back-propagation, while the gradient-based optimization was performed using the Adam update rule

To preprocess of the text of each review, we adopted the Natural Language Toolkit (NLTK), a suite of libraries for natural language processing for English language [2]. The Word2Vec algorithm adopted in the proposed method is included in the Gensim library [3]. Finally, we used the existing implementation of BERT offered by the Transformers library[4], which was pre-trained on a large corpus derived from the Toronto Book Corpus and Wikipedia.

4.3. Experimental setting

For each dataset, reviews were sorted according to their post date/time. The first ordered 80% of reviews were used as training set, while the remaining 20% of the reviews were used as testing set. The classification models were learned on each training set, and their accuracy was evaluated on the corresponding testing set.

Notice that the experimental setting adopted in this study is different from the one commonly adopted in the review spam detection literature, where training-testing splits are generated randomly, therefore neglecting the post date/time of reviews. We believe that ignoring the temporal ordering of the reviews might reduce the fairness of the evaluation phase, because at test time the model may make use of information that was produced after the review had been written, which would not normally be possible in a realistic scenario. Therefore, in the validation and test sets, we update the values of the behavioral features for each review, based on the order in which they were written. Really, at testing time the model can only access information that was collected up to the post date/time of the current review. This setup also correctly captures the behavior of reviewers, which naturally changes over the time as they

[1] https://github.com/hyperopt/hyperopt
[2] https://www.nltk.org/
[3] https://radimrehurek.com/gensim/models/word2vec.html
[4] https://huggingface.co/docs/transformers/model_doc/bert

post new reviews. We measured F-score, G-mean and AUC-ROC to evaluate the accuracy performance of the compared approaches.These metrics were measured by considering the class "spam" as the positive class of the classification.

4.4. Results and discussion

The empirical validation was done to answer the following questions:

- To what extent each individual view influences the accuracy of the classification model? (Section 4.4.1)
- Is a multi-input network more powerful than a single-input network? (Section 4.4.2)
- Is the analysis of reviews written on any domain helpful in improving the accuracy of the spam classifier learned in the domain of a specific product or service? (Section 4.4.3)
- How does the multi-input neural network compare to recent review spam detection methods that use traditional classification algorithms? (Section 4.4.4)

4.4.1. Data view analysis (Q1)

We analyzed how the knowledge enclosed in both the content and behavioral views can influence the performance of EUPHORIA . To this aim, we performed an ablation study, where we measured the accuracy of the following baselines of EUPHORIA :

- Word2Vec that elaborated content features extracted through Word2Vec.
- BERT that elaborated content features extracted through BERT.
- Behav that elaborated Behavioral features.
- Word2Vec + Behav that elaborated both Word2Vec features and Behavioral features.
- Word2Vec + BERT that elaborated both Word2Vec and BERT features.
- EUPHORIA (Word2Vec + BERT + Behav) that elaborated both Word2Vec and BERT features, as well as Behavioral features.

The results collected in Table 3 show that the behavioral features convey the most relevant information to detect fake reviews: Behav outperforms Word2Vec, BERT in both datasets. On the other hand, content features extracted with Word2Vec and BERT disclose non-redundant knowledge on the review text. In fact, the trained neural networks gained accuracy when the content features extracted through both Word2Vec and BERT were processed jointly instead than separately (Word2Vec + BERT outperforms both Word2Vec and BERT). Finally, the main outcome of this experiment is that the richer the processed information, the higher the learning ability of the trained the accuracy of the learned multi-input deep neural network is actually improved when both content and behavioral knowledge are taken into account. In fact, EUPHORIA achieved the highest accuracy in both datasets.

4.4.2. Single-input versus multi-input analysis (Q2)

We continue the study of the effectiveness of the multi-view learning schema of EUPHORIA by investigating the improvement achieved by learning a "multi-input" neural network. To

Table 3
F-score, *G-Mean* and *AUC-ROC* of EUPHORIA and its baselines. The best results are in bold.

Dataset	View	F-score	G-mean	AUC-ROC
Hotel	Word2Vec	0.030	0.126	0.504
	BERT	0.076	0.205	0.514
	Behav	0.580	0.809	0.809
	Word2Vec+Behav	0.574	0.794	0.795
	BERT+Behav	0.581	0.806	0.806
	Word2Vec + BERT	0.175	0.340	0.535
	EUPHORIA (Word2Vec + BERT + Behav)	**0.592**	**0.813**	**0.813**
Restaurant	Word2Vec	0.312	0.610	0.624
	BERT	0.286	0.564	0.598
	Behav	0.328	0.656	0.656
	Word2Vec+Behav	0.334	0.665	0.665
	BERT+Behav	0.368	0.691	0.691
	Word2Vec + BERT	0.328	0.618	0.635
	EUPHORIA (Word2Vec + BERT + Behav)	**0.372**	**0.706**	**0.708**

Table 4
F-score, *G-Mean* and *AUC-ROC* of Single NN and EUPHORIA . The best results are in bold.

Dataset	Architecture	F-score	G-mean	AUC-ROC
Hotel	Single NN	0.508	0.613	0.685
	EUPHORIA	**0.592**	**0.813**	**0.813**
Restaurant	Single NN	0.318	0.594	0.622
	EUPHORIA	**0.372**	**0.706**	**0.708**

this purpose, we considered a single-input baseline (denoted as Single NN) of EUPHORIA . For Single NN, we first computed the feature vectors on the three distinct views (i.e., Word2Vec, BERT and Behav). Then we concatenated the three feature vectors in a single input vector. Finally, we processed the concatenated data as input of a single-input neural network trained to learn the classification model.

Table 4 reports the *F-score*, *G-mean* and *AUC-ROC* of both Single NN and EUPHORIA. These results show that the multi-input neural network of EUPHORIA can actually take advantage of the richness of multi-view data. In fact, they confirmed that processing data of the individual views though a multi-input architecture is more powerful than pre-concatenating these multiple inputs and training a single-input neural network, whose performances are worsened by the effects of the curse of dimensionality.

4.4.3. Multi-domain review analysis (Q3)

In the proposed approach, the classifier is trained to detect fake reviews written for a specific domain of products or services (e.g., hotels or restaurants). However, the behavioral features can be computed over time by accounting for reviews written by the same reviewer on a multitude

Table 5
F-score, *G-Mean* and *AUC-ROC* of EUPHORIA(B) and EUPHORIA . The best results are in bold.

Dataset	Architecture	F-score	G-mean	AUC-ROC
Hotel	EUPHORIA(B)	0.364	0.614	0.686
	EUPHORIA	**0.592**	**0.813**	**0.813**
Restaurant	EUPHORIA(B)	0.340	0.678	0.684
	EUPHORIA	**0.372**	**0.706**	**0.708**

Table 6
F-score, *G-Mean* and *AUC-ROC* of SVM, Ens-SVM and EUPHORIA . The best results are in bold.

Dataset	Architecture	F-score	G-mean	AUC-ROC
Hotel	SVM	0.530	0.779	0.779
	Ens-SVM	0.445	0.688	0.705
	EUPHORIA	**0.592**	**0.813**	**0.813**
Restaurant	SVM	0.351	0.704	0.692
	Ens-SVM	0.349	0.685	0.687
	EUPHORIA	**0.372**	**0.706**	**0.708**

of products. To evaluate the effectiveness of this choice, we formulated the counterpart (denoted as EUPHORIA(B)) of EUPHORIA, which computed behavioral features by only using reviews written in the target domain (i.e., only restaurant reviews in the Restaurant dataset, and only hotel reviews in the Hotel dataset).

The results in Table 5 confirmed that the more the knowledge take into account to represent the behavior of a reviewer, the higher the accuracy gained leveraging this knowledge to detect spamming phenomena.

4.4.4. Deep learning analysis

Finally, we evaluate the effectiveness of the adopted neural network architecture for the considered classification task, by comparing the performance of EUPHORIA to that of competitors with SVM as classification algorithm. In particular, we measured the accuracy of two SVM-based competitors:

- SVM that concatenated the three feature vectors produced by Word2Vec, BERT and Behav in a single input vector and processed the concatenated data as the input of a SVM classifier;
- Ens-SVM that trained an ensemble of three separate SVMs trained from the three feature vectors produced by Word2Vec, BERT and Behav, respectively, and used the ensemble majority rule for the final classification.

The results in Table 6 show that the highest accuracy was achieved with EUPHORIA by training a multi-input neural network. Notably, the main finding of this experiment is that

learning separate classifiers for each view performs worse than learning a single classifier by concatenating all views together.

5. Conclusions

In this paper, we have illustrated a novel predictive approach for review spam detection, which is able to take advantage of both content and behavioral characteristics possibly hidden in online product reviews. In particular, we have proposed to process these multi-view data in their raw format, leaving the task of sharing information (and consequently relationships) across multiple views to the deep learning architecture. We have coupled a multi-view learning approach with a deep learning architecture, in order to gain predictive accuracy from the diversity of data in each view without suffering from the curse of dimensionality. The experiments performed on two benchmark datasets confirm the effectiveness of the proposed approach. One limitation of our methodology is the absence of an online learning phase able to periodically increment the trained classification model as new reviews are recorded over time. In fact, the field of *online learning* with deep neural networks is still mostly unexplored in the review spam detection literature. An interesting avenue for future work is to explore further fine-tuning strategies, which were recently explored in other tasks such as network intrusion detection [22].

Finally, we plan to investigate the use of *graph embedding* techniques, in order to dynamically represent relationships between reviewers and products.

References

[1] A. Heydari, M. ali Tavakoli, N. Salim, Z. Heydari, Detection of review spam: A survey, Expert Systems with Applications 42 (2015) 3634–3642.

[2] C. Musto, M. de Gemmis, P. Lops, G. Semeraro, Generating post hoc review-based natural language justifications for recommender systems, User Model. User Adapt. Interact. 31 (2021) 629–673.

[3] L. Chen, G. Chen, F. Wang, Recommender systems based on user reviews: the state of the art, User Modeling and User-Adapted Interaction 25 (2015).

[4] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, H. A. Najada, Survey of review spam detection using machine learning techniques, J. Big Data 2 (2015) 23.

[5] E. Ferrara, The history of digital spam, Commun. ACM 62 (2019) 82–91.

[6] A. Mukherjee, V. Venkataraman, B. Liu, N. S. Glance, What yelp fake review filter might be doing?, in: E. Kiciman, N. B. Ellison, B. Hogan, P. Resnick, I. Soboroff (Eds.), Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM 2013, Cambridge, Massachusetts, USA, July 8-11, 2013, The AAAI Press, 2013.

[7] T. Mikolov, Q. V. Le, I. Sutskever, Exploiting similarities among languages for machine translation, ArXiv abs/1309.4168 (2013).

[8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, ArXiv abs/1810.04805 (2019).

[9] G. Andresini, A. Andrea Iovine, R. Gasbarro, M. Lomolino, M. de Gemmis, A. Appice, Euphoria: A neural multi-view approach to combine content and behavioral features in

review spam detection, Journal of Computational Mathematics and Data Science 3 (2022) 100036.
[10] N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain, M. Kaleem, Spam review detection techniques: A systematic literature review, Applied Sciences 9 (2019).
[11] N. Jindal, B. Liu, Opinion spam and analysis, in: Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM 2008, Association for Computing Machinery, New York, NY, USA, 2008, p. 219–230.
[12] N. Jindal, B. Liu, Analyzing and detecting review spam, in: Seventh IEEE International Conference on Data Mining (ICDM 2007), 2007, pp. 547–552.
[13] G. Satia Budhi, R. Chiong, Z. Wang, S. Dhakal, Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews, Electronic Commerce Research and Applications 47 (2021) 101048.
[14] J. Li, C. Cardie, S. Li, Topicspam: a topic-model based approach for spam detection, in: ACL, 2013.
[15] Y. Shang, M. Liu, T. Zhao, J. Zhou, T-bert: A spam review detection model combining group intelligence and personalized sentiment information, in: I. Farkaš, P. Masulli, S. Otte, S. Wermter (Eds.), Artificial Neural Networks and Machine Learning – ICANN 2021, Springer International Publishing, Cham, 2021, pp. 409–421.
[16] F. Li, M. Huang, Y. Yang, X. Zhu, Learning to identify review spam., 2011, pp. 2488–2493.
[17] N. Hussain, H. Mirza, I. Hussain, F. Iqbal, I. Memon, Spam review detection using the linguistic and spammer behavioral methods, IEEE Access PP (2020) 1–1.
[18] P. Bhuvaneshwari, A. Rao, H. Robinson, Spam review detection using self attention based cnn and bi-directional lstm, Multimedia Tools and Applications 80 (2021) 1–18.
[19] Y. Ren, Y. Zhang, Deceptive opinion spam detection using neural network, in: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, The COLING 2016 Organizing Committee, Osaka, Japan, 2016, pp. 140–150.
[20] Y. Ren, D. Ji, Neural networks for deceptive opinion spam detection: An empirical study, Information Sciences 385-386 (2017) 213–224.
[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, \. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.
[22] G. Andresini, F. Pendlebury, F. Pierazzi, C. Loglisci, A. Appice, L. Cavallaro, INSOMNIA: towards concept-drift robustness in network intrusion detection, in: N. Carlini, A. Demontis, Y. Chen (Eds.), AISec@CCS 2021: Proceedings of the 14th ACM Workshop on Artificial Intelligence and Security, Virtual Event, ACM, 2021, pp. 111–122.