An ML Model for Predicting Information Check-Worthiness using a Variety of Features

Md Zia Ullah

IRIT, UMR5505 CNRS
mdzia.ullah@irit.fr
118 Route de Narbonne, 31062 Toulouse CEDEX 9, France

Abstract. In this communication, we introduce the important problem of information check-worthiness. We present the method we developed to automatically answer this problem. This method makes use of an elaborated information representation that combines the "information nutritional label" features along with word-embedding features. The information check-worthy claim is then predicted by training a machine learning model based on these features. Our model outperforms the official participants' runs of CheckThat! 2018 challenge.

Keywords: Information check-worthiness; Information nutritional label; Machine learning based model

1 Introduction

The main problems associated to automatic fact-checking consist of (1) deciding whether a piece of information is worth being reviewed or not and (2) finding evidence that helps in detecting if the fact is correct or if it is a fake. Information check-worthiness refers to the first challenge and is specifically critical in political debates [8,2] where facts can be manipulated, denied, or hidden.

2 Method

The approach we developed to tackle this problem relies both on word embedding using Word2Vec model [14] and on the Information Nutritional Label for online documents [5]. The former is now a common model to represent texts for various tasks [18,15]. On the other hand, the information nutritional label which was initially introduced to "help readers making more informed judgments about the items they read" provides scores for various criteria to qualify the content of a text and have shown to be helpful for deciding whether a piece of information should be prioritized for checking or not [13,1].

2.1 Information representation

The information representation combines (a) the information nutritional label features and (b) word embedding features.

Information nutritional label. The information nutritional label for online documents [5] corresponds to a description of the textual information unit according to nine criteria as follows:

1. *Factuality*: the number of facts it mentions,
2. *Readability*: the ease with which a reader can understand it,
3. *Virality*: the speed at which it is propagated,
4. *Emotion*: its emotional impact, both positive and negative emotion.
5. *Opinion*: the number of opinionated sentences it contains,
6. *Controversy*: the number of controversial issues it addresses,
7. *Authority/Trust/Credibility*: its credibility and the authority and trust of the source it belongs to,
8. *Technicality*: the number of technical issues it addresses and technical terms used,
9. *Topicality*: its current interest which is time-dependent.

From the initial label our method makes use of the ones that are underlined (factuality, emotion, controversy, and technicality) in our model. Lespagnol et al. [13] discusses this point in more details.

Word embedding : Word embedding refers to the representation of a word in a semantic space as a vector of numerical values. Words that are semantically and syntactically similar tend to be close in this embedding space. To represent a sentence, we use the pre-trained "Word vectors" which was trained on GoogleNews corpus using Word2Vec model [14]. We average the word vectors of every word in a sentence. When we could not find a word in the model, we represent it with a zero vector. Although zero vector affects the mean [20], this is indeed essential when we could not find any word of the sentence in the model.

2.2 Machine learning

We have considered a machine learning model based on stochastic gradient descent classifier with "log loss" function (AKA, Logistic regression). We keep the default values of other hyper-parameters of the ML algorithm from Scikit-learn (version 3.2.4) [17].

3 Results

We used the CLEF18 CheckThat! 2018 collection (CT-CWC-18) [16] for evaluation. It corresponds to the transcriptions of political debates or speeches from the 2016 US Presidential campaign. For each line of the transcription the training data set includes a label indicating whether this statement is check-worthy (1) or not (0).

The CT-CWC-18 consists of 3 sub-datasets with a total of $4,064$ sentences from which 90 are check-worthiness. The test set consists of 7 sub-datasets for a total of $4,882$ sentences from which 192 are check-worthiness. The data set is strongly unbalanced in favor to sentences that are not worth checking. While oversampling the minority class is common practice in machine learning[3,11], it does not guarantee the best results [21,19]. In our experiments, we studied both cases and report here the best only, which is achieved without oversampling, keeping the initial data as it is.

In Table 1, the results are presented in terms of mean average precision (MAP) which is the official measure for the CLEF track [16]; we used the scripts from the CheckThat! Lab organizers.

While in [13] we evaluated various other features and other feature combinations, the best results were obtained when combining word embeddings and information nutritional label based features. Moreover, also in [13] we consider various machine learning models. The best results have been obtained when using SGD_Logloss (Stochastic gradient descent classifier training using "log" loss function) [12].

Table 1. MAP of the SGD_Logloss ML algorithm, considering features based on Nutritional label (N), Word-embedding (W), or the combination of both (NW) - without oversampling. First row is the best MAP achieved at Checkthat! 2018 challenge.

	Method	**MAP**
Our model	SGD_Logloss – N	.079
	SGD_Logloss – W	.210
	SGD_Logloss – NW	**.230**
CheckThat!	Prise de Fer [23]	.133
	Copenhagen [9]	.115
	UPV-INAOE [7]	.113
	IRIT [1]	.063

We also compared our method to the teams that participated in CLEF track, including Prise de Fer [23], Copenhagen [9], UPV-INAOE-Autoritas [7], and IRIT [1]. Among the participants, the best performing system is Prise de Fer [23] that obtained a MAP score of 0.133. Prise de Fer [23] represented the sentence using word-embedding combined with POS-tags, syntactic dependencies, and some features including named entities, sentiment, and verbal forms. They trained a multi-layer perceptron (MLP) model with two hidden layers (100 units and 8 units, respectively) and the hyperbolic tangent (tanh) as an activation function. The Copenhagen team [9] represented each sentence using word-embedding combined with POS tags and syntactic dependencies. They trained an attention based RNN with GRU memory units and obtained a MAP score of 0.115. The UPV-INAOE team [7] obtained a MAP score of .113 where they used character n-grams as features and k-nearest neighbors as the model. The IRIT team [1] used the features based on information nutritional label, and trained an SVM model which obtained a MAP score of 0.063.

In Table 1, we describe three variants of our method namely SGD_Logloss based on information nutritional label based features (SGD_Logloss-N), word-embedding based features (SGD_Logloss-W), and the combination of information nutritional label and word embedding (SGD_Logloss-NW). We can see the SGD_Logloss-NW produces the

http://alt.qcri.org/clef2018-factcheck

best performance compared to the other two variants. Our method also outperforms all the participating teams' approaches in the CLEF2018 CheckThat! track.

4 Related work

Identifying check-worthy statements has been recently investigated in different studies. In ClaimBuster [10], the authors used the transcripts of all of the US presidential debates that were manually annotated. The authors proposed a SVM-based model with sentence-level features such as sentiment, length, TF-IDF, POS-tags, and Entity Types. Gencheva et al. integrated several context-aware and sentence-level features to train both SVM and Feed-forward Neural Networks [6]. This approach outperforms the ClaimBuster system in terms of MAP and precision.

The best performing system in CheckThat! Lab at CLEF 2018 related shared task is Prise de Fer [23] with MAP of 0.133. The sentence level features they used are word-embedding combined with POS-tags, syntactic dependencies, named entities, sentiment, and verbal forms. They trained a multi-layer perceptron (MLP) consisting of two hidden layers and the hyperbolic tangent as the activation function.

The second best performing system is Copenhagen team's [9] that obtained a MAP of 0.115. The authors represented the sentence using word embedding combined with POS tags and syntactic dependency based features. This representation was used as input to an RNN with GRU memory units, where the output from each word was aggregated using attention, followed by a fully connected layer, from which the output was predicted using a sigmoid function [9].

The other participants used different representations such as character n-grams [7] or topics [22]; different machine learning algorithms such as SVM [1], Random Forest [1], k-nearest neighbors [7], or Gradient boosting [22].

5 Conclusion

In this communication, we present a method for predicting information check-worthiness that was developed in [13].

Experimental results on the CheckThat! 2018 collection shows that combing information nutritional label and word-embedding using SGD_Logloss model produces the best performance and outperforms the known related methods. Oversampling the training set have not improved the results although the training examples are unbalanced. In future work, we would like to improve the model by integrating additional components from the information nutritional label such as readability and other language model such as BERT [4].

Ethical issue. While Check That challenge has its proper ethical policies, detecting information check-worthiness raises ethical issues that are beyond the scope of the paper.

Acknowledgement. This work has been partially funded by the European Union's Horizon 2020 H2020-SU-SEC-2018 under the Grant Agreement n°833115 (PREVISION project `https://cordis.europa.eu/project/id/833115`). The paper reflects only the authors' view and the Commission is not responsible for any use that may be made of the information it contains

References

1. Agez, R., Bosc, C., Lespagnol, C., Petitcol, N., Mothe, J.: IRIT at checkthat! 2018. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France (2018)
2. Bond, G.D., Schewe, S.M., Snyder, A., Speller, L.F.: Reality monitoring in politics. In: The Palgrave Handbook of Deceptive Communication, pp. 953–968. Springer (2019)
3. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority oversampling technique. Journal of artificial intelligence research **16**, 321–357 (2002)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Fuhr, N., Giachanou, A., Grefenstette, G., Gurevych, I., Hanselowski, A., Jarvelin, K., Jones, R., Liu, Y., Mothe, J., Nejdl, W., et al.: An information nutritional label for online documents. In: ACM SIGIR Forum. vol. 51, pp. 46–66. ACM (2018)
6. Gencheva, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., Koychev, I.: A context-aware approach for detecting worth-checking claims in political debates. In: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017. pp. 267–276 (2017)
7. Ghanem, B., Montes-y-Gómez, M., Pardo, F.M.R., Rosso, P.: UPV-INAOE - check that: Preliminary approach for checking worthiness of claims. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France (2018)
8. Graves, L.: Deciding what's true: The rise of political fact-checking in American journalism. Columbia University Press (2016)
9. Hansen, C., Hansen, C., Simonsen, J.G., Lioma, C.: The copenhagen team participation in the check-worthiness task of the competition of automatic identification and verification of claims in political debates of the CLEF-2018 checkthat! lab. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France (2018)
10. Hassan, N., Adair, B., Hamilton, J.T., Li, C., Tremayne, M., Yang, J., Yu, C.: The quest to automate fact-checking. world (2015)
11. Khan, S.H., Hayat, M., Bennamoun, M., Sohel, F.A., Togneri, R.: Cost-sensitive learning of deep feature representations from imbalanced data. IEEE transactions on neural networks and learning systems **29**(8), 3573–3587 (2017)
12. Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M.: Logistic regression. Springer (2002)
13. Lespagnol, C., Mothe, J., Ullah, M.Z.: Information nutritional label and word embedding to estimate information check-worthiness. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 941–944 (2019)
14. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems 26, pp. 3111–3119. Curran Associates, Inc. (2013)
15. Mothe, J.: "recherche d'information textuelle, apprentissage et plongement de mots". In: Document numérique. Hermès (2020)
16. Nakov, P., Barrón-Cedeno, A., Elsayed, T., Suwaileh, R., Màrquez, L., Zaghouani, W., Atanasova, P., Kyuchukov, S., Da San Martino, G.: Overview of the CLEF-2018 CheckThat! Lab on automatic identification and verification of political claims. In: Proceedings of

the Ninth International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction (CLEF'18). pp. 372–387. Springer (2018)
17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
18. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. arXiv preprint arXiv:1802.05365 (2018)
19. Reshma, I.A., Gaspard, M., Franchet, C., Brousset, P., Faure, E., Mejbri, S., Mothe, J.: Training set class distribution analysis for deep learning model – application to cancer detection (2019)
20. Ullah, M.Z., Shajalal, M., Chy, A.N., Aono, M.: Query subtopic mining exploiting word embedding for search result diversification. In: Asia Information Retrieval Symposium. pp. 308–314. Springer (2016)
21. Weiss, G.M., Provost, F.: Learning when training data are costly: The effect of class distribution on tree induction. Journal of Artificial Intelligence Research **19**, 315–354 (2003)
22. Yasser, K., Kutlu, M., Elsayed, T.: bigir at CLEF 2018: Detection and verification of check-worthy political claims. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France (2018)
23. Zuo, C., Karakas, A., Banerjee, R.: A hybrid recognition system for check-worthy claims using heuristics and supervised learning. In: Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France (2018)