Predicting News Values from Headline Text and Emotions

Maria Pia di Buono[1]   **Jan Šnajder**[1]   **Bojana Dalbelo Bašić**[1]
Goran Glavaš[2]   **Martin Tutek**[1]   **Natasa Milic-Frayling**[3]

[1] TakeLab, Faculty of Electrical Engineering and Computing, University of Zagreb
`first.namelastname@fer.hr`

[2] Data and Web Science Group, University of Mannheim, Germany
`goran@informatik.uni-mannheim.de`

[3] School of Computer Science, University of Nottingham, UK
`natasa.milic-frayling@nottingham.ac.uk`

Abstract

We present a preliminary study on predicting *news values* from headline text and emotions. We perform a multivariate analysis on a dataset manually annotated with news values and emotions, discovering interesting correlations among them. We then train two competitive machine learning models – an SVM and a CNN – to predict news values from headline text and emotions as features. We find that, while both models yield a satisfactory performance, some news values are more difficult to detect than others, while some profit more from including emotion information.

1   Introduction

News values may be considered as a system of criteria applied to decide about the inclusion or exclusion of material (Palmer, 2000) and about the aspects of the selected material that should be emphasized by means of headlines. In fact, the informative value of headlines lays its foundations in their capability of optimizing the relevance of their stories for their users (Dor, 2003). To the intent of being optimizers of the news relevance, headlines carry out a set of different functions while meeting two needs: attracting users' attention and summarizing contents (Ifantidou, 2009). In order to attract users' attention, headlines should provide the triggers for the emotional impact of the news, accounting emotional aspects related to the participants of the event or to the actions performed (Ungerer, 1997). As far as the summarization of contents is concerned, headlines may be distinguished on the basis of two main goals: headlines that represent the abstract of the main event and headlines that promote one of the details in the news story (Bell, 1991; Nir, 1993). Furthermore, Iarovici and Amel

(1989) recognize two simultaneous functions: "a semantic function, regarding the referential text, and a pragmatic function, regarding the reader (the receiver) to whom the text is addressed."

In this work we present a preliminary study on predicting news values from headline text and emotions. The study is driven by two research questions: (1) what are the relations among news values conveyed by headlines and the human emotions triggered by them, and (2) to what extent can a machine learning classifier successfully identify the news values conveyed by headlines, using merely text or text and triggered emotions as input? To this end, we manually annotated an existing dataset of headlines and emotions with news values. To answer the first question, we carried out a multivariate analysis, and discovered interesting correlations among news values and emotions. To answer our second research question, we trained two competitive machine learning models – a support vector machine (SVM) and a convolutional neural network (CNN) – to predict news values from headline text and emotions. Results indicate that, while both models yield a satisfactory performance, some news values are more difficult to detect, some profit from including emotion information, and CNN performs better than SVM on this task.

2   Related work

Despite the fact that news values has been widely investigated in Social Science and journalism studies, not much attention has been paid to its automatic classification by the NLP community. In fact, even if news value classification may be applied in several user-oriented applications, e.g., news recommendation systems, and web search engines, few scholars (De Nies et al., 2012; Piotrkowicz et al., 2017) have been focused on this particular topic. Related to our work is the work on predicting

emotions in news articles and headlines, which has been investigated from different perspectives and by means of different techniques. Strapparava and Mihalcea (2008) describe an experiment devoted to analyze emotion in news headlines, focusing on six basic emotions and proposing knowledge-based and corpus-based approaches. Kozareva et al. (2007) extract part of speech (POS) from headlines in order to create different bag of words pairs with six emotions and compute for each pair the Mutual Information Score. Balahur et al. (2013) test the relative suitability of various sentiment dictionaries in order to separate positive or negative opinion from good or bad news. Ye et al. (2012) deal with the prediction of emotions in news from readers' perspective, based on a multi-label classification. Another strand of research more generally related to our work is short text classification. Short text classification is technically challanging due to the sparsity of features. Most work in this area has focused on classification of microblog messages (Sriram et al., 2010; Dilrukshi et al., 2013; Go et al., 2009; Chen et al., 2011).

3 Dataset

As a starting point, we adopt the dataset proposed for the SemEval-2007 Task 14 (Strapparava and Mihalcea, 2007). The dataset consists of 1250 headlines extracted from major newspapers such as New York Times, CNN, BBC News, and Google News. Each headline has been manually annotated for valence and six emotions (Anger, Disgust, Fear, Joy, Sadness, and Surprise) on a scale from 0 to 100. In this work, we use only the emotion labels, and not the valence labels.

News values. On top of the emotion annotations, we added an additional layer of news value labels. Our starting point for the annotation was the news values classification scheme proposed by Harcup and O'Neill (2016). This study proposes a set of fifteen values, corresponding to a set of requirements that news stories have to satisfy to be selected for publishing. For the annotation, we decided to omit two news values whose annotation necessitates contextual information: "Audio-visuals", which signals the presence of infographics accompanying the news text, and "News organization's agenda", which refers to stories related to the news organization's own agenda. This resulted in a set of 13 news value labels.

News value	IAA		IAA (adj)		Support
	κ	F1	κ	F1	
Bad news	**0.47**	0.526	**0.72**	0.744	85
Celebrity	**0.51**	0.545	**0.74**	0.761	82
Conflict	0.19	0.245	**0.52**	0.564	86
Drama	0.25	0.383	**0.58**	0.663	178
Entertainment	**0.53**	0.684	**0.76**	0.843	351
Follow-up	0.10	0.129	**0.43**	0.451	29
Good news	0.23	0.268	**0.54**	0.563	65
Magnitude	0.08	0.121	0.34	0.371	45
Shareability	0.05	0.101	0.29	0.335	130
Surprise	0.06	0.102	0.38	0.409	43
Power elite	0.36	0.472	**0.66**	0.718	166

Table 1: Original and adjudicated interannotator agreement (Cohen's κ and F1-macro scores) and counts for each news value (agreement scores averaged over three annotator pairs and four annotator groups; moderate/substantial κ agreement shown in bold).

Annotation task. We asked four annotators to independently label the dataset. The annotators were provided short guidelines and a description of the news values. We first ran a calibration round on a set of 120 headlines. After calculating the inter-annotator agreement (IAA), we decided to run a second round of calibration, providing further information about some labels conceived as more ambiguous by the annotators (e.g., "Bad news" vs. "Drama" vs. "Conflict" and "Celebrity" vs. "Power elite"). For the final annotation round, we arranged the annotators into four distinct groups of three, so that each headline would be annotated by three annotators. The annotation was done on 798 headlines using 13 labels. Annotation analysis revealed that two of these labels "Exclusivity" and "Relevance", have been used in a marginal number of cases so we decide to omit these labels from the final dataset.

Table 1 show the Cohen's κ and F1-macro IAA agreement scores for the 11 news value labels. We observe a moderate agreement of $\kappa \geq 0.4$ (Landis and Koch, 1977) only for the "Bad news", "Celebrity", and "Entertainment" news values, suggesting that recognizing news values from headlines is a difficult task even for humans. To obtain the final dataset, we adjudicated the annotations of the three annotators my a majority vote. The adjudicated IAA is moderate/substantial, except for "Magnitude", "Shareability", and "Surprise".

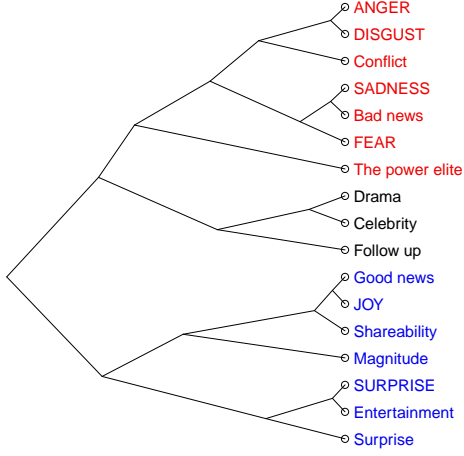**Factor analysis.** As a preliminary investigation of the relations among news values and emotions in

Figure 1: A dendrogram of the correlations among factor loadings for news values and emotions. (Emotions are shown in caps.)

headlines, we carry out a multivariate data analysis using factor analysis (FA) (Hair et al., 1998). The main goal of FA is to measure the presence of underlying constructs, i.e., factors, which in our case represent the correlation among emotions and news values, and their factor loading magnitudes. The use of FA is justified here because (1) we deal with cardinal (news values) and ordinal (emotions) variables and (2) the data exhibits a substantial degree of multicollinearity. We applied varimax, an orthogonal factor rotation used to obtain a simplified factor structure that maximizes the variance. We then inspected the eigenvalue scree plot and chose to use seven factors whose values were larger than 1 as to reduce the number of variables without loosing relevant information. To visualize the factor structure and relations among news values and emotions, we performed a hierarchical cluster analysis, using complete linkage with one minus Pearson's correlation coefficient as the distance measure.

Fig. 1 shows the resulting dendrogram. We can identify three groups of news values and emotions. The first group contains the negative emotions related to "Conflict" and "Bad news", and the rather distant "Power elite". The second group contains only news values, namely "Drama", "Celebrity", and "Follow up". The last group is formed by two positive emotions, joy and surprise, which are the kernels of two sub-groups: joy is related to "Good news", "Shareability" and, to a lesser extent, to "Magnitude", while surprise emotions relates to "Entertainment" and "Surprise" news values.

4 Models

We consider two classification algorithms in this study: a support vector machine (SVM) and the CNN. The two algorithms are known for their efficiency in text classification tasks (Joachims, 1998; Kim, 2014; Severyn and Moschitti, 2015). We frame the problem of news values classification as a multilabel task, and train one binary classifier for each news value, using headlines labeled with that news value as positive instances and all other as negative instances.

Features. We use the same feature sets for both SVM and CNN. As textual features, we use the pretrained Google News word embeddings, obtained by training the skip-gram model with negative sampling (Mikolov et al., 2013). For emotion features, we used the six ground-truth emotion labels from the SemEval-2007 dataset, standardized to zero mean and unit variance.

SVM. An SVM (Cortes and Vapnik, 1995) is a powerful discriminate model trained to maximize the separation margin between instances of two classes in feature space. We follow the common practice of assuming additive compositionality of the word embeddings and represent each headline as one 300-dimensional vector by averaging the individual word embeddings of its constituent words, whereby we discard the words not present in the dictionary. Note that this representation is not sensitive to word order. We use the SVM implementation from scikit-learn (Pedregosa et al., 2011), which in turn is based on LIBSVM (Chang and Lin, 2011). To maximize the efficiency of the model, we use the RBF kernel and rely on nested 5×5-cross-validation for hyperparameter optimization, with $C \in \{1, 10, 100\}$ and $\gamma \in \{0.01, 0.1\}$.

CNN. A CNN (LeCun and Bengio, 1998) is a feed-forward neural network consisting of one or more convolutional layers, each consisting of a number of filters (parameter matrices). Convolutions between filters and slices of the input embedding matrix aim to capture informative local sequences (e.g., word 3-grams). Each convolutional layer is followed by a pooling layer, which retains only the largest convolutional scores from each filter. A CNN thus offers one important advantage over SVM, in that it can detect indicative word sequences – a capacity that might be crucial when classifying short texts such as news headlines.

News value	SVM		CNN	
	T	T+E	T	T+E
Bad news	0.652	0.763*	0.778 †	**0.848***†
Celebrity	**0.553**	0.534	0.496	0.526
Conflict	0.526	0.487	0.654 †	**0.659** †
Drama	0.636	0.637	0.668	**0.681**
Entertainment	0.832	0.783*	0.803	**0.841***
Good news	0.414	0.513	0.509	**0.578**
Magnitude	0.299	**0.515***	0.438	0.507
Power elite	0.596	0.570	0.695 †	**0.700** †
Shareability	0.309	0.318	**0.427** †	0.425 †

Table 2: F1-scores of SVM and CNN news values classifiers using text ("T") or text and emotions ("T+E") as features. Best result for each news value are shown in bold. "*" denotes a statistically significant difference between feature sets "T" and "T+E" for the same classifier, and "†" a statistically significant difference between SVM and CNN classifiers with the same features (p<0.05, two-tailed permutation test).

In our experiments, we trained CNNs with a single convolutional and pooling layer. We used 64 filters, optimized filter size ($\{3,4,5\}$) using nested cross-validation, and performed top-k pooling with $k = 2$. For training, we used the RMSProp algorithm (Tieleman and Hinton, 2012).

In addition to the vanilla CNN model that uses only the textual representation of a headline, we experimented with a model that additionally uses emotion labels as features. For each headline, the emotion labels are concatenated to the latent CNN features (i.e., output of the top-k pooling layer) and fed to the output layer of the network. Let $\mathbf{x}_T^{(i)}$ be the latent CNN vector of the i-th headline text, and $\mathbf{x}_E^{(i)}$ the corresponding vector of emotion labels. The output vector $\mathbf{y}^{(i)}$, a probability distribution over labels, is then computed as:

$$\mathbf{y}^{(i)} = softmax\big(\mathbf{W} \cdot [\mathbf{x}_T^{(i)}; \mathbf{x}_E^{(i)}] + \mathbf{b}\big)$$

where $\mathbf{W}$ and $\mathbf{b}$ are the weight matrix and the bias vector of the output layer.

5 Evaluation

Table 2 shows the F1-scores of the SVM and CNN news values classifiers, trained with textual features ("T") or both textual and emotion features ("T+E"). We report the results for nine out of 11 news values from Table 1; the two omitted labels are "Follow-up" and "Surprise", for which the number of instances was too low to successfully train

the models. Models for the remaining nine news values were trained successfully and outperform a random baseline (the differences are significant at p<0.001; two-sided permutation test (Yeh, 2000)).

We can make three main observations. First, there is a considerable variance in performance across the news values: "Bad news" and "Entertainment" seems to be the easiest to predict, whereas "Shareability", "Magnitude", and "Celebrity" are more difficult. Secondly, by comparing "T" and "T+E" variants of the models, we observe that adding emotions as features improves leads to further improvements for the "Bad news" and "Entertainment" news values (differences are significant at p<0.05) for CNN, and for SVM also for "Magnitude", but for other news values adding emotions did not improve the performance. This finding is aligned with the analysis from Fig. 1, where "Bad news" and "Entertainment" are the two news values that correlate the most with one of the emotions. Finally, by comparing between the two models, we note that CNN generally outperforms SVM: the difference is statistically significant for "Bad news", "Conflict", "Power elite", "Shareability", regardless of what features were used. This suggest that these news values might be identified by the presence of specific local word sequences.

6 Conclusions and Future Work

We described a preliminary study for predicting news values using headline text and emotions. A multivariate analysis revealed a three-way grouping of news values and emotions. Experiments with predicting news values revealed that both a support vector machine (SVM) and a convolutional neural network (CNN) can outperform a random baseline. The results further indicate that some news values are more easily detectable than others, that adding emotions as features helps for news values that are highly correlated with emotions, and that CNNs ability to detect local word sequences helps in this task, probably because of the brevity of headlines.

This works opens up a number of interesting research directions. One is to study the relation between the linguistic properties of headlines and news values. Another research direction is the comparison between headlines and full-text stories as features for news value prediction. It would also be interesting to analyze how news values correlate with properties of events described in text. We intend to pursue some of this work in the near future.

Acknowledgments

References

Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik Van Der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2013. Sentiment analysis in the news. *arXiv preprint arXiv:1309.6202*.

Allan Bell. 1991. *The language of news media*. Blackwell Oxford.

Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27.

Mengen Chen, Xiaoming Jin, and Dou Shen. 2011. Short text classification improved by learning multi-granularity topics. In *Twenty-Second International Joint Conference on Artificial Intelligence*.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Tom De Nies, Evelien Dheer, Sam Coppens, Davy Van Deursen, Erik Mannens, and Rik Van de Walle. 2012. Bringing newsworthiness into the 21st century. *Web of Linked Entities (WoLE) at ISWC*, 2012:106–117.

Inoshika Dilrukshi, Kasun De Zoysa, and Amitha Caldera. 2013. Twitter news classification using svm. In *Computer Science & Education (ICCSE), 2013 8th International Conference on*, pages 287–291. IEEE.

Daniel Dor. 2003. On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5):695–721.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12).

Joseph F Hair, William C Black, Barry J Babin, Rolph E Anderson, Ronald L Tatham, et al. 1998. *Multivariate data analysis*, volume 5. Prentice hall Upper Saddle River, NJ.

Tony Harcup and Deirdre O'Neill. 2016. What is news? News values revisited (again). *Journalism Studies*, pages 1–19.

Edith Iarovici and Rodica Amel. 1989. The strategy of the headline. *Semiotica*, 77(4):441–460.

Elly Ifantidou. 2009. Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts. *Journal of Pragmatics*, 41(4):699–720.

Thorsten Joachims. 1998. Text categorization with support vector machines: Learning with many relevant features. *Machine learning: ECML-98*, pages 137–142.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

Zornitsa Kozareva, Borja Navarro, Sonia Vázquez, and Andrés Montoyo. 2007. Ua-zbsa: a headline emotion classification through web information. In *Proceedings of the 4th international workshop on semantic evaluations*, pages 334–337. Association for Computational Linguistics.

J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Yann LeCun and Yoshua Bengio. 1998. The handbook of brain theory and neural networks. chapter Convolutional Networks for Images, Speech, and Time Series, pages 255–258. MIT Press, Cambridge, MA, USA.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Raphael Nir. 1993. A discourse analysis of news headlines. *Hebrew Linguistics*, 37:23–31.

Jerry Palmer. 2000. *Spinning into control: News values and source strategies*. A&C Black.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Alicja Piotrkowicz, Vania Dimitrova, and Katja Markert. 2017. Automatic extraction of news values from headline text. In *Proceedings of EACL*. Association for Computational Linguistics.

Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 959–962, New York, NY, USA. ACM.

Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Fer-hatosmanoglu, and Murat Demirbas. 2010. Short text classification in twitter to improve information filtering. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 841–842. ACM.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics.

Carlo Strapparava and Rada Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560. ACM.

Tijmen Tieleman and Geoffrey Hinton. 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. Technical Report 2.

Friedrich Ungerer. 1997. Emotions and emotional language in english and german news stories. *The language of emotions*, pages 307–328.

Lu Ye, Rui-Feng Xu, and Jun Xu. 2012. Emotion prediction of news articles from reader's perspective based on multi-label classification. In *Machine Learning and Cybernetics (ICMLC), 2012 International Conference on*, volume 5, pages 2019–2024. IEEE.

Alexander Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 18th conference on Computational linguistics-Volume 2*, pages 947–953. Association for Computational Linguistics.