The Thirty-Second International Florida
Artificial Intelligence Research Society Conference (FLAIRS-32)

Using Correlation for Labelset Selection in Multi-Label Classification of Users Reactions

Zacarias Curi,[1] **Alceu de Souza Britto Jr.,**[1, 2] **Emerson Cabrera Paraiso,**[1]
[1]Pontifícia Universidade Católica do Paraná (PUCPR) - Curitiba, PR, Brazil
[2]Universidade Estadual de Ponta Grossa (UEPG) - Ponta Grossa, PR, Brazil
{zacarias, alceu, paraiso}@ppgia.pucpr.br

Abstract

The increasing use of social networks has made opinion mining an important field in the area of Natural Language Processing. The analysis of texts from the reader perspective tends to generate multi-label data since one can interpret the text using different contexts. In this paper, a new method for multi-label classification is proposed to identify reactions or emotions in texts. The new method uses data correlation to improve the class ensemble process used to create the classifiers. In addition to the new method, a new corpus of news written in Brazilian Portuguese labeled with user reactions is presented. Experiments performed with the new corpus and with two existing corpora have demonstrated that the proposed method generates statistically superior or equivalent results, requiring fewer classifiers or classes than traditional problem transformation methods.

Introduction

The increasing number of Internet users and the growth of social networks generated a large amount of textual data. This data can be used for several applications but its quantity prevent this analysis from being performed manually.

In addiction to analyzing the opinions of users, analyzing the reactions of them after reading a text may also be helpful. The term reaction is used in this work as the attitude acquired by a person after receiving an external stimulus. According to Desmet (2003), reactions form emotions and they can be divided into behavioral reactions, expressive reactions, and physiological reactions. One way to classify expressive reactions can be performed by using emojis. After reading a news or post, the user can select the image that best matches their reaction. Figure 1 shows the emojis used on Facebook for this purpose.

Figure 1: Emojis used on Facebook

The classification of user reactions can be used to assist the task of recommending new texts to users, as well as to help choose texts that deserves greater prominence and tend to become popular. Another application of this field is into the advertising recommendation system. User reactions can be used to analyze if the content of the text can negatively influence the content of the advertisement. An example of a situation where this problem occurs is the presentation of an advertisement from an airline company in a news story about an airplane accident. In this situation, people may unconsciously associate the advertising company with accidents.

Reactions are expressed from the perspective of the reader leading to a multi-label case by nature. Each user has he/she individuality, which can generate different reactions for the same text.

As presented in Zhang and Zhou (2014), there are two strategies to perform the classification of multi-label corpora: algorithm adaptation and problem transformation. In this work it is considered a third category, the class ensemble. The algorithm adaptation methods consists of classification algorithms that are capable of directly addressing the multi-label problem. The problem transformation methods consist of making transformations into the problem, transforming it into one or more problems of classification or ranking. The methods of class ensemble consist of dividing the problem into smaller problems, performing an ensemble with several groups of classes and performing the classification with some method of algorithm adaptation or problem transformation. Several class ensemble approaches are presented in the literature, such as the Ensembles of Classifier Chains (ECC), Ensembles of Pruned Sets (EPS), and the Random k-labelsets (RAkEL).

One of the most popular class ensemble methods in the literature is RAkEL, initially presented in Tsoumakas and Vlahavas (2007). This method starts with a random selection of m labelsets with k problem classes, after this selection each group is classified using the problem transformation method called Label Powerset (LP). After the classification of all groups, the average decision of the predicted classes is calculated, creating the set that represents the final prediction.

In this work, a method of class ensemble that uses the correlation of the classes for the selection of labelsets is presented. The method uses characteristics of the reactions in texts to aid in the classification process. The objective of the

proposed method is to remove the random selection of labelsets, allowing the use of sets that enable the extraction of better results from the problem transformation method used. Besides the proposal of the new method, a new corpus of news written in Brazilian Portuguese labeled with the user's reactions is presented. In this paper, we used annotated corpora with reactions and emotions.

The next section presents the main works related to the classification of reactions and emotions in texts and the main methods of class ensemble in the literature.

Related Work

The multi-label classification methods can be divided into three categories: algorithm adaptation, problem transformation, and class ensemble. A description of the main classification methods existing in the literature can be found in (Tsoumakas, Katakis, and Vlahavas 2009; Sorower 2010; Zhang and Zhou 2014).

In Madjarov et al. (2012), the authors presented a comparison between different classification approaches for several domains. Almeida et al. (2018) also presented a comparison between different classification approaches, but they used the identification of emotions in news texts from the reader's perspective.

Curi and colleagues were interested in identifying emotions and reactions in news texts (Curi, Britto Jr, and Paraiso 2018). In their work, the authors compared an approach using the Binary Relevance method and the LSTM algorithm with several traditional classification methods. In the work of Liu and Chen (2015), the authors presented a hybrid approach composed of three components: text segmentation, feature extraction, and multi-label classification. The authors used texts extracted from microblogs and labeled it with 10 different emotions.

In Zhang, Li, and Lu (2017) the authors also used a corpus composed of texts of microblogs, but labeled with the six categories of emotions proposed by Ekman. They proposed a framework with the use of emotion label correlation and social correlation. Ye and colleagues presented a comparison with several multi-label classification algorithms, problem transformation methods, and several feature selection methods (Ye, Xu, and Xu 2012). The results obtained by Ye, Xu, and Xu showed that the best performance was obtained by the method RAkEL.

In addition to the work focused on the emotions mining, there are reports of the efficiency of the method RAkEL in different applications. This method was initially presented in Tsoumakas and Vlahavas (2007) and some changes have been proposed in the literature.

In 2014, Lo, Lin, and Wang presented an expansion of the RAkEL method in order to minimize the global error between the prediction and the ground truth. The method, called Generalized k-labelsets Ensemble, has a novel objective function to learn the expansion coefficients of the base classifiers and found a solution to learn the coefficients efficiently. Wu and Lin (2017) presents a method called progressive random k-labelsets (PRAkEL). The PRAkEL method is able to transfer the information to the subproblems in the training stage to optimize the classification process.

According to Gharroudi, Elghazel, and Aussem (2015), the RAkEL method is usually badly calibrated due to the problems raised by the imbalanced label representation and proposed three solutions to this problem. The first one is to increase the diversity of the classifiers in the ensemble. The second is to smooth the label powerset probability that is estimated during the ensemble aggregation process, and the third solution is about to calibrate the label decision thresholds. Nasierding, Kouzani, and Tsoumakas (2015) presents a triple-random ensemble learning method for multi-label classification. The method integrates the concepts of random subspace, bagging and RAkEL methods to form an approach to classify multi-label data.

The method presented in this paper differs from those presented in the literature review because it has a class selection method based on the data correlation. The strategy presented in this paper uses the structure of the data labeled with emotions and reactions to avoid the variation brought by the random selection of labels, as it occurs in the RAkEL method. The method is evaluated using a new multi-label news corpus labeled with the reader's point of view. The next section presents the proposed method.

Proposed Method

The method presented in this paper works similarly to the RAkEL method. The main difference between them is in the class selection step. As RAkEL, the new method can also be classified as a class ensemble method. Different from the traditional ensemble of classifiers, the class ensemble methods perform the creation of labelsets and use the same classification process for all existing sets. Similar to most class ensemble methods, the proposed method has three steps: group creation, classification, and selection of the final labels. The Figure 2 presents the steps of the proposed method.

In the group creation stage, the classes correlations of the training database are used for the creation of the labelsets. The use of correlation is possible due to the characteristics of the reactions in texts. Although readers may have different reactions to the same text, there are some combinations of classes that tend to occur frequently. For example, when reading a news that reports some tragedy, it is common for some people to be sad and others to be afraid. The use of this type of combination allows the accomplishment of an efficient division, providing simpler data for classification stage. The main objective of this step is to prevent classes that can cause confusion (with a strong correlation between them) to be kept in the same labelset.

The group selection process begins with the choice of the number of groups m and the number of labels in each group k. The calculations of these values is performed as demonstrated in (Tsoumakas, Katakis, and Vlahavas 2011). The value of k is calculated by $L/2$, where L is the number of classes in the problem. The main advantage of this division is that each class appears in only one labelset, avoiding the creation of many classes or many classifiers by the classification method of the next step. This simplification allows the reduction of the processing necessary for classification.

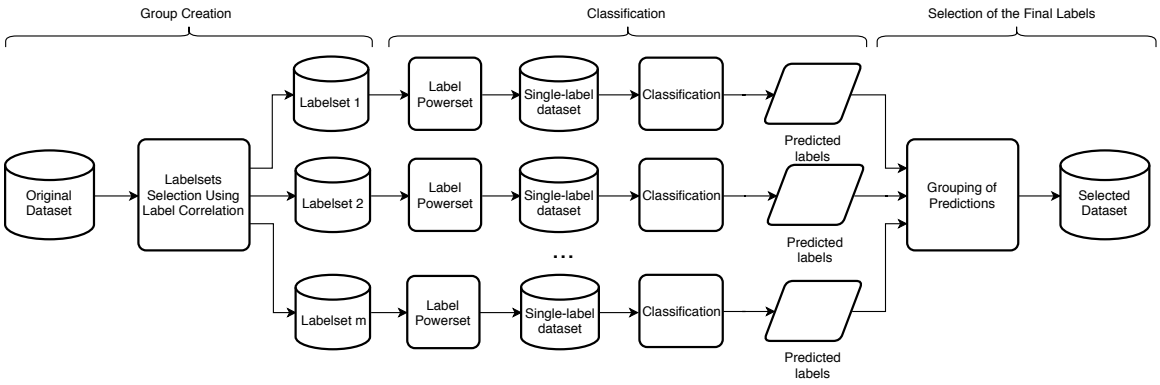

Figure 2: Proposed method

After defining the values of k and m the correlation is calculated through the Pearson correlation coefficient. This coefficient is used to measure the linear correlation between two variables X and Y. As can be seen in the equation 1, the coefficient of each combination of the problem (represented by ρ) is calculated by dividing the covariance of the variables X and Y by multiplying the standard deviation of these variables. After calculating the correlation between all possible combinations, the lowest correlations are grouped in m groups with a maximum of k elements. With the creation of the groups, the classification stage begins.

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

In the classification step, the problem transformation method called Label Powerset (LP) is used for each of the m groups created in the previous step. The basic strategy of LP is to create a new label for each label combination in the problem, turning the multi-label dataset into a single label dataset. With the creation of a single label dataset, a traditional multi-class classifier can be applied. The main advantage of the strategy used by the LP is the need for only one single label classifier, but in many cases it may presents the creation of many new classes. The use of LP in a class ensemble reduces this problem because each classifier needs to handle only a portion of the classes of the problem. After the prediction of all the groups created, the results are grouped, creating the set of labels that define the final result. As in the group creation step we use $k = L/2$ each class is presented in only one labelset, thus avoiding the need for a selection threshold.

The proposed method presents some advantages over other methods in the literature. The new method allows to remove the randomness of the RAkEL, using the correlation coefficient of the labels for this. The method also allows the reduction of the complexity of the LP, as well as avoiding the creation of excessive new classifiers, as in Binary Relevance (BR) and Classifier Chains (CC) methods. The next section presents the experimental protocol used.

Experiments

This section presents the experiments performed, the corpora used, and the evaluation metrics. The proposed method was implemented in the software meka[1]. In all experiments, the corpus was divided in three uses of 3-folds cross-validation. The use of three folds allows the creation of larger test divisions, generating more accurate tests for unbalanced databases. The methods and algorithms used in experiments are presented in the next section.

Methods and Algorithms

To evaluate the proposed method the results of the experiments were compared to the results generated in all possible combinations of RAkEL for $k = L/2$. This was done to represent all possible combinations that can be generated by the RAkEL method, as well as to demonstrate the variation between the combinations. In addition to the RAkEL method, was used the Binary Relevance (BR), Classifier Chains (CC) and Label Powerset (LP) methods. In all cases, Naive Bayes (NB) and Support Vector Machine (SVM) induction algorithms were used.

The problem transformation method BR consists of transforming the multi-label problem into a series of binary problems, where each class of the original problem is transformed into a classifier. One of the main limitations presented by the BR method is that each class is treated independently, ignoring the correlations between classes. One way to solve this limitation is presented by the CC method. The CC creates a chain, using the output of one classifier as input to the next. This strategy allows adding the existing link between classes to optimize the classification process.

Another strategy used for the comparison is LP. This method of problem transformation consists of creating a new class for each existing class combination in the training database. The LP method is used in the classification step of the method presented in this work.

Datasets

In this paper, the evaluation of the proposed method is performed using two corpora already available in the literature

[1] http://waikato.github.io/meka/

and a new corpus of news labeled with user reactions. A summary of the characteristics of the corpora is presented in Table 1.

Table 1: Used corpora

Corpus	# Intances	# Labels	Cardinality	Density
BF	7,879	8	3.9277	0.4910
G1	2,000	7	1.9635	0.2805
GP	668	6	2.1572	0.3595

The G1 corpus was initially presented in (Dosciatti, Ferreira, and Paraiso 2015). This corpus is composed by 2,000 news written in Brazilian Portuguese collected from the portal G1[2]. The news were labeled by specialists using the six basic emotions proposed by Ekman and the neutral for cases where none of the emotions was present in the text. The classes used in the annotation process were: anger, disgust, fear, happiness, sadness, surprise and neutral.

The BF corpus was initially presented in (Curi, Britto Jr, and Paraiso 2018). This corpus is composed of entertainment news taken from the BuzzFeed Brazil[3] and labeled with user reactions. In this paper, we used a variation of the original dataset, where we considered only the news with more than three votes. The corpus is composed of 7,879 news articles labeled with the reactions: cute, fail, funny, hate, love, shock, skeptic and win.

We also developed a new corpus called GP. This corpus is composed by 668 news written in Brazilian Portuguese collected from the news website called Gazeta do Povo[4]. The annotation of this corpus was made based on the choices of users. At the end of each news, the website provides a series of emoticons that allows the user to express their reaction. The emoticons provided by the website are: anger, funny, like, love, sad and surprise. As in the BF corpus, only news with more than three votes were considered. A change was also made on the labels, where a threshold was applied to remove classes with less than 3% of the total votes. The corpora used in this paper can be obtained on the website[5].

On these three corpora, used in this paper, were removed the special characters and stopwords. All links, emails, numbers, and currency symbols were replaced by tokens. After these operations, a stemming was applied and the TF-IDF (term frequency-inverse document frequency) method was used to create the input vectors of the classifiers.

Evaluation Measures

To compare the results of each methods, three evaluation metrics were used allowing a more detailed analysis.

The Micro F1 for multi-label data works similarly to F1 for single-label data, establishing the harmonic mean between precision and recall. The main difference between the single-label version and the multi-label is the input of the data in the calculation, as presented in the equation 2. The Micro F1 allows the visualization of the impact of the imbalance in the classification method. This metric is represented by the equation 3, where the values used are calculated by equation 2.

$$B_{micro}(h) = B\left(\sum_{j=1}^{|L|} VP_j, \sum_{j=1}^{|L|} FP_j, \sum_{j=1}^{|L|} VN_j, \sum_{j=1}^{|L|} FN_j\right) \quad (2)$$

$$F^{\beta}() = \frac{(1+\beta^2)\,.\,VP_j}{(1+\beta^2)\,.\,VP_j + \beta^2\,.\,FN_j + FP_j} \quad (3)$$

The Jaccard Index metric is used to compare the similarity and diversity of the data. This metric examines the proportion of positive labels correctly predicted. The Jaccard Index is represented by the equation 4.

$$ACC_{ml}(h) = \frac{1}{|X|}\sum_{i=1}^{|X|}\frac{|h(x_i) \cap Y_i|}{|h(x_i) \cup Y_i|} \quad (4)$$

The Hamming Loss metric considers the classifier errors but ignores the unbalance of the data. Unlike the other metrics presented, the lower the result obtained by this evaluation metric better is the result obtained by the classifier. The Hamming Loss is defined by the equation 5, where $\triangle$ represents the symmetric difference between two sets.

$$HammingLoss(h) = \frac{1}{|X|}\frac{1}{|L|}\sum_{i=1}^{|X|}|h(x_i) \triangle Y_i| \quad (5)$$

Results and Discussion

As presented in the previous section, the proposed method was compared to all possible combinations of the RAkEL method for $k = L/2$ and with the BR, CC and LP problem transformation methods. All methods were used with the SVM and NB induction algorithms. The obtained results are shown in Table 2. The values presented in RAkEL represent the mean value obtained by the combinations tested in the RAkEL method. It is important to note that none of the RAkEL combinations provided the best result for all the evaluation metrics with the two induction algorithms used. The method proposed in this paper (presented as New in the table) achieved, in all cases, better results than the average of all combinations of the RAkEL method for all the metrics used.

The method presented in this paper, as well as the other methods of class ensemble, allows the classification without the need of L classifiers as in the binary methods. Besides, it uses a smaller number of new classes if compared to the LP method. In the traditional LP method, up to 2^L new classes can be created, with an ensemble, the number of new classes is maximum 2^k. The basic idea of the proposed method is to keep the existing benefits of class ensemble methods and to use existing features in user reactions to improve the classification process. This is done through the characteristics of the news, which tend to generate strongly correlated reactions in

[2]https://g1.globo.com/

[3]https://www.buzzfeed.com/?country=br

[4]https://www.gazetadopovo.com.br/

[5]https://www.ppgia.pucpr.br/~paraiso/mineracaodeemocoes

Table 2: Obtained results

		Micro F1 ↑		Jaccard index ↑		Hamming Loss ↓	
		NB	**SVM**	**NB**	**SVM**	**NB**	**SVM**
GP	**New**	**0.6084±0.0185**	**0.6004±0.0113**	**0.4939±0.0178**	**0.4852±0.0146**	**0.2552±0.0122**	**0.2613±0.0100**
	BR	0.5925±0.0299	0.5752±0.0315	0.4757±0.0317	0.4406±0.0234	0.2927±0.0205	0.2977±0.0054
	CC	0.5901±0.0345	0.5770±0.0301	0.4647±0.0337	0.4440±0.0214	0.2864±0.0114	0.2987±0.0043
	LP	0.6073±0.0266	0.3002±0.0266	0.4931±0.0222	0.1825±0.0205	0.2822±0.0183	0.7370±0.0292
	RAkEL*	0.6047±0.0247	0.5978±0.0290	0.4903±0.0312	0.4795±0.0358	0.2607±0.0176	0.2668±0.0164
BF	**New**	**0.6427±0.0026**	**0.6411±0.0045**	**0.4884±0.0034**	**0.4817±0.0048**	**0.3537±0.0029**	**0.3571±0.0042**
	BR	0.6182±0.0155	0.6234±0.0170	0.4628±0.0176	0.4621±0.0180	0.3746±0.0096	0.3636±0.0033
	CC	0.6211±0.0200	0.6236±0.0158	0.4667±0.0210	0.4628±0.0162	0.3770±0.0083	0.3647±0.0037
	LP	0.6223±0.0154	0.4319±0.0723	0.4687±0.0184	0.2709±0.0538	0.3707±0.0082	0.6311±0.0287
	RAkEL*	0.6395±0.0128	0.6371±0.0091	0.4827±0.0147	0.4775±0.0089	0.3550±0.0121	0.3585±0.0081
G1	**New**	0.5362±0.0091	**0.5148±0.0137**	0.4272±0.0104	**0.4000±0.0158**	0.2669±0.0064	**0.2578±0.0060**
	BR	0.5273±0.0143	0.4923±0.0054	0.4013±0.0139	0.3659±0.0015	**0.2651±0.0080**	0.2845±0.0007
	CC	**0.5478±0.0075**	0.4934±0.0022	**0.4274±0.0087**	0.3746±0.0016	0.2762±0.0042	0.2846±0.0029
	LP	0.4429±0.0067	0.2817±0.0114	0.3322±0.0057	0.1713±0.0076	0.3128±0.0033	0.7427±0.0070
	RAkEL*	0.5356±0.0249	0.5108±0.0219	0.4220±0.0252	0.3974±0.0240	0.2725±0.0154	0.2584±0.0126

The value presented for the RAkEL method is the average of all possible combinations for $k = L/2$

some cases. The presented method avoids that classes with a strong correlation are treated by the same classifier, thus avoiding some possible ambiguity in the classification.

The results from the new method was compared with all existing combinations of the RAkEL method using the Friedman test. The tests with a significance of $\alpha = 0.05$ showed that only the comparisons made in the GP corpus with the SVM algorithm for the micro F1 metric and with the SVM and the NB for the Hamming Loss metric showed no statistical difference. In all other cases, one or more combinations showed statistically lower results. Using the Nemenyi test we also observed that our method generated statistically higher results than at least one combination in all cases where there was a statistical difference. This result showed that for most of the cases there is a significant difference for the results obtained by the RAkEL method. This exposes how the random selection of labelsets can influence the final prediction. These results also allowed the verification that in all cases the proposed method generated a result statistically superior or equal to the best result obtained by the RAKEL method, demonstrating that the correlation can be used to select the labelsets.

In addition to comparisons with the RAkEL method, the new method was compared with the BR, CC, and LP problem transformation methods. The results obtained by these methods with the classification algorithms SVM and NB can be observed in Table 2. To compare the results, the Friedman test with a significance of $\alpha = 0.05$ was used. The test demonstrated that there is a statistical difference between the methods for the three metrics used. The Figure 3 presents the result of the Nemenyi test for the comparison among the used algorithms. As can be observed, the presented method generated the best results for the three metrics used.

The Nemenyi test also allows to observe that for the metric F1, the proposed method presented results statistically superior to the three methods of problem transformation with the SVM algorithm and to the LP method with the NB

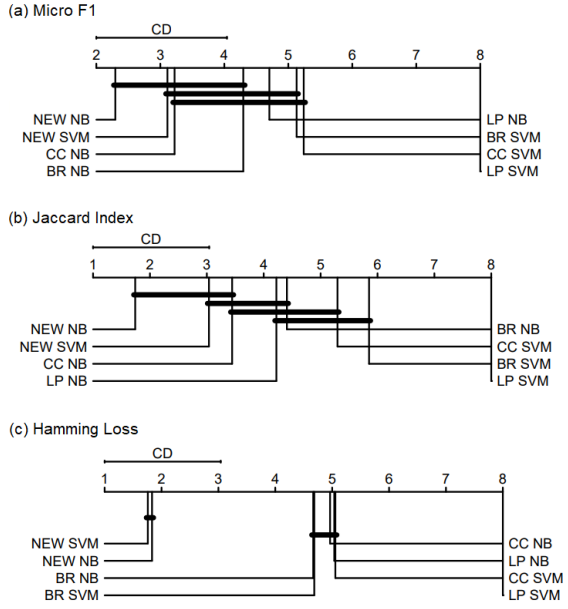

Figure 3: Nemenyi test

algorithm. For the Jaccard Index, the method was also superior to the three methods with the SVM algorithm and the LP and BR with the NB algorithm. Finally, for the Hamming Loss metric, the new method obtained results statistically superior to all other methods used. This shows that although in some cases the presented method obtained a result equivalent to CC and BR with NB, the performance of the method was superior when considering the three metrics.

In relation to the induction algorithms, we observed that the performance of NB and SVM present variations in relation to the problem transformation method and the evaluation metric. For the proposed method, the induction algo-

rithm used generated statistically similar results in all the metrics used. In the tests performed, the SVM algorithm generated worse results than the NB and the combination of this algorithm with the LP method presented the worst performances.

Conclusion

In this paper, a new approach to the task of classifying multi-label emotions or reactions in news texts was presented. The new method allows the creation of a class ensemble based on the correlation between the possible reactions expressed by news readers. A new corpus of news labeled with user reactions was also presented.

The experiments demonstrated that the new method generated statistically equivalent or superior results than all the possible combinations generated by the RAkEL method. Tests performed with the CC, BR and LP methods also demonstrated that the new method generated statistically equivalent or higher results, using less classifiers than the BR and CC methods and less new classes than the LP method. The reduction in the number of classifiers and classes allows the use of fewer computational resources.

As future work we intend to extend the presented method, replacing the LP method for a new classification method based on the division of sentences. This new approach aims to reduce the number of new classes required for classification.

Acknowledgments

We would like to thank Siemens Ltd and CAPES (Coordination for the Improvement of Higher Level Personnel) for partially supporting this work.

References

Almeida, A. M.; Cerri, R.; Paraiso, E. C.; Mantovani, R. G.; and Junior, S. B. 2018. Applying multi-label techniques in emotion identification of short texts. *Neurocomputing*.

Curi, Z.; Britto Jr, A. d. S.; and Paraiso, E. C. 2018. Multi-label classification of user reactions in online news. *arXiv preprint arXiv:1809.02811*.

Desmet, P. 2003. Measuring emotion: Development and application of an instrument to measure emotional responses to products. In *Funology*. Springer. 111–123.

Dosciatti, M. M.; Ferreira, L. P. C.; and Paraiso, E. C. 2015. Anotando um corpus de notícias para a análise de sentimentos: um relato de experiência (annotating a corpus of news for sentiment analysis: An experience report). In *Proceedings of the Brazilian Symposium in Information and Human Language Technology*, 121–130.

Gharroudi, O.; Elghazel, H.; and Aussem, A. 2015. Calibrated k-labelsets for ensemble multi-label classification. In *International Conference on Neural Information Processing*, 573–582. Springer.

Liu, S. M., and Chen, J.-H. 2015. A multi-label classification based approach for sentiment classification. *Expert Systems with Applications* 42(3):1083–1093.

Lo, H.-Y.; Lin, S.-D.; and Wang, H.-M. 2014. Generalized k-labelsets ensemble for multi-label and cost-sensitive classification. *IEEE Transactions on Knowledge and Data Engineering* 26(7):1679–1691.

Madjarov, G.; Kocev, D.; Gjorgjevikj, D.; and Džeroski, S. 2012. An extensive experimental comparison of methods for multi-label learning. *Pattern recognition* 45(9):3084–3104.

Nasierding, G.; Kouzani, A. Z.; and Tsoumakas, G. 2010. A triple-random ensemble classification method for mining multi-label data. In *IEEE International Conference on Data Mining Workshops*, 49–56. IEEE Computer Society.

Sorower, M. S. 2010. A literature survey on algorithms for multi-label learning. *Oregon State University, Corvallis* 18.

Tsoumakas, G., and Vlahavas, I. 2007. Random k-labelsets: An ensemble method for multilabel classification. In *European Conference on Machine Learning*, 406–417. Springer.

Tsoumakas, G.; Katakis, I.; and Vlahavas, I. 2009. Mining multi-label data. In *Data mining and knowledge discovery handbook*. Springer. 667–685.

Tsoumakas, G.; Katakis, I.; and Vlahavas, I. 2011. Random k-labelsets for multilabel classification. *IEEE Transactions on Knowledge and Data Engineering* 23(7):1079–1089.

Wu, Y.-P., and Lin, H.-T. 2017. Progressive random k-labelsets for cost-sensitive multi-label classification. *Machine Learning* 106(5):671–694.

Ye, L.; Xu, R.-F.; and Xu, J. 2012. Emotion prediction of news articles from reader's perspective based on multi-label classification. In *International Conference on Machine Learning and Cybernetics*, volume 5, 2019–2024. IEEE.

Zhang, M.-L., and Zhou, Z.-H. 2014. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering* 26(8):1819–1837.

Zhang, X.; Li, W.; and Lu, S. 2017. Emotion detection in online social network based on multi-label learning. In *International Conference on Database Systems for Advanced Applications*, 659–674. Springer.