

Opinion Mining and Active Learning: a Comparison of Sampling Strategies

Douglas Vitório^{1,2}, Ellen Souza², Adriano L. I. Oliveira¹

¹Centro de Informática – Universidade Federal de Pernambuco (CIn-UFPE)
Recife – PE – Brazil

²MiningBR Research Group – Universidade Federal Rural de Pernambuco (UFRPE)
Serra Talhada – PE – Brazil

Abstract. *There are two main problems when performing Opinion Mining (OM) with data streams: the lack of labeled data and the need to update the learning model. The most used OM techniques cannot deal well with these challenges, so, an alternative is to use semi-supervised methods, such as the Active Learning, which is a method to label only selected data rather than the entire data set; however, it requires the choice of a sampling strategy to select the data to be labeled. In this paper, we evaluated eight strategies in ten data sets, in order to identify the best ones for OM with Twitter streams. According to our experiments, the Entropy strategy showed the best results, but it selects a large number of instances to be labeled, requiring further investigation.*

1. Introduction

The growth of social media and the Web 2.0 has made the understanding of people's thoughts and opinions fundamental for decision making, especially when comments are voluntarily expressed [Firmino Alves et al. 2014]. However, it is impossible for humans to fully understand the user-generated content efficiently, which is why there has been an increasing interest in the scientific community to create systems capable of extracting information from this kind of data [Balazs and Velásquez 2016].

To solve this problem, the most common approaches are based on Opinion Mining (OM). OM, also known in the literature as Sentiment Analysis, is, according to [Liu and Zhang 2012], the field of study that analyzes people's sentiments, opinions, evaluations, attitudes, and emotions about entities, such as products, services, organizations, individuals, issues, events, topics, and their attributes, expressed in textual input. This analysis is made by the classification of the opinion contained in a document or sentence, or even related to a particular feature, into categories, such as *positive*, *negative*, or *neutral*.

Nevertheless, the OM applications have had, in the last decades, their focus on static and well-known domains, such as movie reviews, not being able to deal efficiently with the characteristics of continuous *data streams* and the challenges brought by them [Guerra et al. 2014]. The analysis of *data streams* from social media, also known as *social streams*, is important because of the possible changes in people's opinions about certain entities, as new information arrives [Wang et al. 2013].

In Machine Learning, the change in a given concept that occurs over time is known as *concept drift* [Widmer and Kubat 1996]. In a similar way and according

to [Wang et al. 2013], a change in an opinion that occurs over time may be considered an *opinion drift* and the detection of these drifts are relevant for the OM results.

Applications that need to deal with *data streams* and are sensitive to drifts must face two main challenges: the limited availability of labeled data and the need to constantly update the learning model, due to *data streams*' evolutionary nature. Though, the most commonly used OM models are strongly based on Supervised Learning [Ravi and Ravi 2015], which intensifies the first problem; whereas the second problem lies in vocabulary changes and, mainly, in the aforementioned nature of the streams [Guerra et al. 2014].

Due to these problems, an alternative is the use of Semi-supervised Learning, which does not require that all data is labeled, but only a part of it. A common semi-supervised approach in Data Mining is Active Learning (AL) [Zhu et al. 2007], which can deal well with problems where labeled data is costly to obtain [Zimmermann et al. 2015]. The AL technique labels only the most valuable data rather than labeling the entire data set. However, a sampling strategy is needed to decide whether the label of an instance should be requested or not, so it can be used to feed the model's training set [Žliobaitė et al. 2011].

Only a few recent papers were found applying Opinion Mining with AL [Smailović et al. 2014, Zimmermann et al. 2015, Kranjc et al. 2015], and they used different strategies to select the instances. Also, they did not provide a complete comparison between these strategies. Thus, the goal of this study is to present an evaluation of Active Learning sampling strategies in the Opinion Mining scenario. Eight strategies were evaluated, of which two are proposed by the authors inspired by other strategies found in literature: *Variable Entropy* and *Variable Randomized Entropy*.

The rest of this paper is structured as follows: Section 2 shows the background of Active Learning sampling strategies for Opinion Mining and discusses the related work. Section 3 details the experimental setup. The findings are reported and discussed in Section 4. Finally, Section 5 draws the conclusions.

2. Background and Related Work

Active Learning is a Semi-supervised approach, commonly used for Data Mining with *data streams* due to its competence to deal with problems in which labeled data is costly to obtain [Zhu et al. 2007, Zimmermann et al. 2015].

This method uses an initial seed of labeled documents as the first training set of the model, then, a sampling strategy is used to occasionally request the label of the most suitable documents that are not present in the training set, in order to update the model with these new instances [Zimmermann et al. 2015].

Therefore, the main challenge of the AL area is to identify the sampling strategy which achieves the greater performance, while keeping the training set the smaller as possible [Zhu et al. 2007]. Besides that, this strategy must be able to balance the labeled data set through time and detect changes (drifts) [Žliobaitė et al. 2011]. A general practice is to employ some heuristics, or rules, to select the most valuable instances [Zhu et al. 2010].

The next subsections explain strategies that have been used in Text Mining and Opinion Mining applications.

2.1. Random Sampling

The *Random Sampling* is a very simple strategy that selects the instances at random, based on a probability B , where B is the budget. The budget represents the portion of the instances that will be added to the model.

This strategy was used by [Žliobaitė et al. 2011] for Text Mining and in the Opinion Mining applications of [Smailović et al. 2014, Zimmermann et al. 2015, Kranjc et al. 2015].

2.2. Uncertainty

The *Uncertainty* strategy selects the instances for which the current classifier is less certain. The implementation from [Žliobaitė et al. 2011] was considered in this study, which uses a threshold (θ) value. In this case, an instance is selected if the classifier's certainty for it is below $1 - \theta$.

[Smailović et al. 2014] used an SVM classifier and combined the advantages of this strategy and the *Random Sampling* one, besides using them separately. The authors' goal was to find the best querying strategy for Active Learning to Sentiment Analysis on financial Twitter *data streams*, concluding that, by using the AL approach, the prediction power of the sentiment classifier in the stock market application was improved. While [Kranjc et al. 2015] created a framework for Sentiment Analysis with AL, using *Uncertainty* and *Random Sampling* strategies, although the authors did not evaluate or compare the strategies performances.

Finally, [Žliobaitė et al. 2011] also made two modifications in the *Uncertainty* technique, based on dynamic allocation of labeling efforts over time and randomization of the search space. The first one, called *Variable Uncertainty*, changes the threshold using an adjusting step (s) to adapt it depending on the incoming data, expanding and contracting the threshold. The second one, called *Variable Randomized Uncertainty*, also uses the adjusting step, besides randomizing the threshold for every instance, multiplying it by a normally distributed random variable that follows $\mathcal{N}(1, \delta)$. As stated by the authors, these strategies react well to changes (drifts) that can occur anywhere in the instance space, being suitable for *data streams*. They performed experiments using benchmark stream prediction data sets, as well as textual data from the IMDb and Reuters; however, they did not perform Sentiment Analysis.

2.3. Information Gain

[Zimmermann et al. 2015] proposed an AL sampling strategy based on information gain, which is specific for Text Mining, since it uses the word-class distribution of words of a document and the distribution after considering the predicted label for that document. With this strategy, the documents which provide a gain in information, considering the actual model, are selected.

The authors used this strategy, and the *Uncertainty* one, to select the documents (*tweets* and product reviews) that will update the model and the words added to the vocabulary. They compared the results with the *Random Sampling* strategy, with an incremental approach that requires all arriving document labels, and a non-adaptive method. The AL sampling strategies performed better than the others, and the *Information Gain* showed good performance on all data sets, considering the *kappa* statistic metric.

2.4. Entropy

According to [Yang and Loog 2018], the entropy can be used as a measure of uncertainty. The authors performed a comparison between AL strategies for logistic regression and showed that the usage of entropy as a measure of uncertainty can produce the best accuracy in a large number of data sets, although they did not use textual data.

In this study, the entropy implementation from [Lewis and Gale 1994] is considered and the *Entropy* AL strategy follows the same idea of the *Uncertainty* one, but, in this case, a document is selected if its entropy is greater than the threshold θ .

However, it is worth mentioning that uncertainty-based sampling approaches can get stuck at sub-optimal solutions, due to their lack of exploration, continuously selecting instances which will not improve the current classifier [Huang et al. 2014].

3. Experimental Setup

To achieve the goal of this study, the six sampling strategies presented above (*Random Sampling*, *Uncertainty*, *Variable Uncertainty*, *Variable Randomized Uncertainty*, *Information Gain*, and *Entropy*) were evaluated on the same scenarios, which are explained in the following subsections.

In addition, we proposed two other strategies based on entropy and inspired by the modifications in *Uncertainty* brought by [Žliobaitė et al. 2011]: ***Variable Entropy*** and ***Variable Randomized Entropy***, which are basically the same presented in [Žliobaitė et al. 2011], but using the entropy measure [Lewis and Gale 1994] instead of the classifier’s certainty. The purpose is to investigate whether the benefits of randomization and the use of a variable threshold for drift reaction appear also with different uncertainty measures.

3.1. Data sets

Two publicly available Twitter stream data sets were used: Sentiment140 [Go et al. 2009] and Sanders [Sanders 2011]. Studies performing OM with *data streams* were found using the former [Smailović et al. 2014, Zimmermann et al. 2014a, Zimmermann et al. 2014b, Zimmermann et al. 2015, Wagner et al. 2015] and the latter [Aston et al. 2014a, Aston et al. 2014b].

The Sentiment140 *training* data set contains 1,600,000 automatically annotated *tweets*, which would be extremely costly to process; so, we took subsets of 10,000, 5,000, 2,500, and 1,000 *tweets*, in order to evaluate the strategies on different sized corpora. In addition, we also used the Sentiment140 *test* data set, which contains 497 *tweets* manually annotated.

On the other hand, the Sanders data set contains *tweets* about four different companies, so we used the *tweets* mentioning each company as a subset, besides the *full* corpus, containing all *tweets*. This partition was done in order to provide a bigger number of corpora in which we could evaluate the strategies, besides the fact that a *data stream* usually refers to only one entity. Then, we had a total of ten corpora, whose details are presented in Table 1. As can be observed, the corpora differs from each other by size, number of classes, domain, and unbalance.

Table 1. Details of the corpora used.

Name	Size	Positive	Negative	Neutral	Details
Sentiment140_test	497	182	177	138	-
Sentiment140_10000	10,000	5,812	4,188	-	<i>tweets</i> 25,000 - 35,000
Sentiment140_5000	5,000	2,970	2,030	-	<i>tweets</i> 35,000 - 40,000
Sentiment140_2500	2,500	1,461	1,039	-	<i>tweets</i> 40,000 - 42,500
Sentiment140_1000	1,000	579	421	-	<i>tweets</i> 42,500 - 43,500
Sanders_apple	1,002	164	316	522	-
Sanders_google	838	202	57	579	-
Sanders_microsoft	864	91	132	641	-
Sanders_twitter	719	62	67	590	-
Sanders_all	3,423	519	572	2,332	-

3.2. Pre-processing

The implementation was done using the Python programming language and the Natural Language Toolkit (NLTK)¹ and Scikit-learn² libraries. To structure the data into a Vector Space Model (VSM), we used the CountVectorizer from the Scikit-learn library. The TweetTokenizer, from the NLTK, was chosen to perform the tokenization of the documents, due to its specificity. We opted to use simple pre-processing techniques and focus only on the evaluation of the AL strategies.

3.3. Processing

The classifier chosen for this study was the Multinomial Naïve Bayes (MNB), which is particularly suitable for opinion streams, since it adjusts to changes in the probabilities of the words [Wagner et al. 2015]. Besides that, the MNB proved to be the most efficient in the paper of [Žliobaitė et al. 2011], which used Active Learning with textual data. The MNB was also used in the study of [Zimmermann et al. 2015] and, according to [Souza et al. 2016], it is one of the most used classifiers for Opinion Mining. We have used the *Scikit-learn* MNB implementation.

To implement the Active Learning sampling strategies, we used the modAL Python library [Danka and Horvath 2018]. This library has the implementation of the uncertainty and entropy measures. We adapted the strategies based on these two measures using the algorithms presented in [Žliobaitė et al. 2011]. The *Random Sampling* strategy was manually implemented, as well as the *Information Gain*, which followed the details as demonstrated in [Zimmermann et al. 2015]. In this paper, the Active Learning process was iterative, so, each time a new instance was selected, the model was updated.

As hyperparameters, for the *Random Sampling* strategy we used $B = 0.3$, so, about 30% of the instances were added to the model; the papers researched [Žliobaitė et al. 2011, Zimmermann et al. 2015] evaluated different values of B , then, we chose 0.3 as an average value. In a similar way, we also used $\theta = 0.3$ for *Uncertainty* and *Entropy*; so, the instances with a classifier’s confidence below than 70% were selected by the former, while the instances with an entropy greater than 0.3 were chosen

¹<https://www.nltk.org>

²<https://scikit-learn.org>

by the latter. Finally, in the same way as [Žliobaitė et al. 2011], we used $s = 0.01$ and $\delta = 1$ for both *Uncertainty* and *Entropy* variations, with the goal to perform a fair comparison between them.

3.4. Evaluation

Due to the continuous nature of the streams, we opted not to use the *cross-validation* technique in the evaluation step. Instead, we used the *holdout* technique, separating the last 30% of the stream to test the performance of the classifier and the others 70% for training. The performance measure was *accuracy*.

For all the techniques, we started with an initial seed containing the first 10% of the stream, while the sampling strategies decided which of the documents, from the remaining 60% of the data separated for training, would be used to update the model.

4. Results and Discussion

Table 2 describes the accuracy of each sampling strategy: *Random Sampling* (RAND), *Uncertainty* (UNC), *Variable Uncertainty* (UNCV), *Variable Randomized Uncertainty* (UNCVR), *Information Gain* (IG), *Entropy* (ENT), *Variable Entropy* (ENTV), and *Variable Randomized Entropy* (ENTVR); while Table 3 shows the quantity of instances selected by each strategy. The best results per data set are shown in bold.

Table 2. The accuracy of each sampling strategy.

Data set	RAND	UNC	UNCV	UNCVR	IG	ENT	ENTV	ENTVR
Sanders_apple	0.583	0.583	0.613	0.620	0.560	0.626	0.623	0.603
Sanders_google	0.793	0.741	0.773	0.777	0.793	0.757	0.793	0.765
Sanders_microsoft	0.703	0.687	0.683	0.683	0.687	0.687	0.695	0.679
Sanders_twitter	0.758	0.763	0.758	0.753	0.753	0.763	0.758	0.749
Sanders_all	0.748	0.720	0.721	0.694	0.750	0.727	0.710	0.718
Sentiment140_test	0.604	0.584	0.590	0.617	0.402	0.637	0.610	0.570
Sentiment140_1000	0.580	0.570	0.663	0.663	0.656	0.676	0.670	0.616
Sentiment140_2500	0.730	0.729	0.749	0.737	0.692	0.742	0.716	0.740
Sentiment140_5000	0.681	0.680	0.703	0.710	0.674	0.708	0.700	0.694
Sentiment140_10000	0.719	0.710	0.724	0.723	0.704	0.725	0.712	0.729
Mean	0.690	0.676	0.698	0.698	0.667	0.705	0.699	0.686
Average ranking	4	5.7	3.7	4.1	5.6	2.3	3.7	5.5
Win counts	2	1	1	1	2	4	1	1

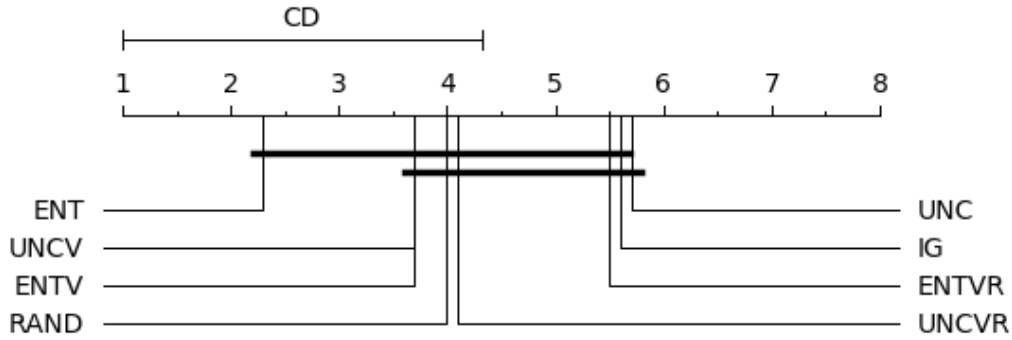
The Friedman and the Nemenyi post-hoc tests were applied. The Friedman test was used to observe whether the strategies performance presented a statistical difference (considering a p-value = 0.05). The test resulted in a p-value = 0.019, demonstrating that there is a statistical difference between some of the strategies. Then, the Nemenyi post-hoc test was used to determine which techniques show this difference; Figure 1 presents the CD diagram obtained from the post-hoc test, in which the strategies connected with a black bar are statistically similar. The SciPy³ Python library was used for the former test and the Orange⁴ library for the latter.

³docs.scipy.org/doc/scipy-0.14.0/reference/index.html

⁴<http://docs.orange.biolab.si/3/data-mining-library/index.html>

Table 3. Number of instances selected by each strategy.

Data set	RAND	UNC	UNCV	UNCVR	IG	ENT	ENTV	ENTVR
Sanders_apple	183	129	288	296	172	327	300	307
Sanders_google	153	128	234	239	176	282	247	248
Sanders_microsoft	145	5	240	242	227	32	242	244
Sanders_twitter	135	1	197	199	291	8	197	200
Sanders_all	583	320	1,005	1,010	560	817	1,010	1,012
Sentiment140_test	92	105	141	143	79	188	166	159
Sentiment140_1000	177	144	291	296	190	384	319	311
Sentiment140_2500	451	445	737	741	506	1,005	758	755
Sentiment140_5000	875	787	1,482	1,486	1,112	1,878	1,501	1,503
Sentiment140_10000	1,815	1,815	2,980	2,981	2,432	4,136	3,005	2,994

**Figure 1. Nemenyi post-hoc test comparison of the eight sampling strategies.**

The results demonstrated that the eight strategies had similar performance in the same scenarios. However, the *Entropy* strategy achieved a statistically better performance than the *Uncertainty* one, and was numerically superior to the other techniques. This result was similar to [Yang and Loog 2018], which showed that the usage of entropy, as a measure of uncertainty, can produce the best accuracy in a large number of non textual and non stream data sets.

Although the *Information Gain* strategy achieved the best results in [Zimmermann et al. 2015], it did not performed well in our experiments. Besides that, it was the most complex technique evaluated in this study, as it maintains the vocabulary at a high computational cost.

It was observed that the changes in the *Uncertainty* strategy from [Žliobaitė et al. 2011] have proved to be useful for OM with Twitter streams: the *Variable Uncertainty* and the *Variable Randomized Uncertainty* showed substantially better results than the fixed *Uncertainty* strategy. Though, this was not observed with the entropy replacing the uncertainty measure. The *Entropy* strategy without randomization and/or the adjusting step achieved the best results.

This latest observation may be explained by the fact that, while the number of selected instances increased in all scenarios with the changes in the *Uncertainty* technique, it dropped with the changes in the *Entropy* strategy, as shown in Table 3.

Although the *Entropy* strategy achieved good accuracy in most scenarios, its large number of selected documents makes it less efficient in terms of computational cost. Moreover, one of the benefits of the Active Learning method is that it requires only a few labeled instances rather than the entire data set, so, a strategy that selects a fewer number of instances is preferable.

Finally, the fixed *Uncertainty* technique selected less instances in almost all data sets, but, in our experiments, this strategy reached the worst results. This may indicate that the low number of instances selected was not enough to build a classifier that can deal well with *drifts*, and that there is a relationship between the accuracy and the number of selected instances. In this case, an alternative is to increase the value of the *Uncertainty* θ threshold.

5. Conclusion

In this study, a document level sentiment analysis with *data streams* from Twitter was performed, in order to identify the best sampling strategies for Active Learning with this kind of data. To do this, we implemented and evaluated eight AL sampling techniques, two of them proposed by us and inspired by existing strategies in the literature.

Another contribution of this paper is the evaluation of AL sampling strategies in Opinion Mining scenarios, as in the case of *Entropy* and the *Uncertainty* methods proposed by [Žliobaitė et al. 2011], which had not been used in this area until then.

According to the experiments, the *Entropy* strategy achieved the best results when compared to the other techniques, mainly *Uncertainty*. However, it needs a large number of documents to be labeled, making it less efficient in terms of computational cost.

And, by noticing that the best strategy (*Entropy*) selected a large number of documents, while the worst strategy (*Uncertainty*) selected a very small one, we also observed that there may be a correlation between the accuracy and the number of selected instances.

As future work, we plan to extend the results presented in this paper by increasing the number of data sets and evaluating other scenarios. The Sentiment140 corpus is extremely large, so we plan to take subsets of data from different parts of the stream, as we only took the data from the initial part in this study. Also, we plan to perform experiments with data from other sources (e.g., Facebook, review sites, etc.).

Another scenarios to evaluate are: (1) fixing the number of selected instances; (2) establishing a stopping criteria; and (3) selecting all instances before updating the model, rather than updating it every time a new instance is selected.

Finally, some further analysis might be useful: measuring the impact of different values for the hyperparameters (budget, θ , s , and δ); and evaluating classifier models other than Multinomial Naïve Bayes, such as SVM or Logistic Regression. The increase of the θ threshold value, for instance, may improve the *Uncertainty* accuracy.

Acknowledgment

Douglas Vitório and Adriano L. I. Oliveira are supported by CNPq (Brazilian Council for Scientific and Technological Development).

References

- Aston, N., Liddle, J., and Hu, W. (2014a). Twitter sentiment in data streams with perceptron. *Journal of Computer and Communications*, 2(03):11.
- Aston, N., Munson, T., Liddle, J., Hartshaw, G., Livingston, D., and Hu, W. (2014b). Sentiment analysis on the social networks using stream algorithms. *Journal of Data Analysis and Information Processing*, 2(02):60.
- Balazs, J. A. and Velásquez, J. D. (2016). Opinion Mining and Information Fusion: A survey. *Information Fusion*, 27:95–110.
- Danka, T. and Horvath, P. (2018). modAL: A modular active learning framework for Python. available on arXiv at <https://arxiv.org/abs/1805.00979>.
- Firmino Alves, A. L., Baptista, C. d. S., Firmino, A. A., Oliveira, M. G. a. d., and Paiva, A. C. d. (2014). A comparison of svm versus naive-bayes techniques for sentiment analysis in tweets: A case study with the 2013 fifa confederations cup. In *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web, WebMedia '14*, pages 123–130, Nova York, NY, EUA. ACM.
- Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12).
- Guerra, P. C., Meira, Jr., W., and Cardie, C. (2014). Sentiment analysis on evolving social streams: How self-report imbalances can help. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM '14*, pages 443–452, New York, NY, USA. ACM.
- Huang, S., Jin, R., and Zhou, Z. (2014). Active learning by querying informative and representative examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10):1936–1949.
- Kranjc, J., Smailović, J., Podpečan, V., Grčar, M., Žnidaršič, M., and Lavrač, N. (2015). Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the clowdfloWS platform. *Information Processing & Management*, 51(2):187 – 203.
- Lewis, D. D. and Gale, W. A. (1994). A sequential algorithm for training text classifiers. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '94*, pages 3–12, New York, NY, USA. Springer-Verlag New York, Inc.
- Liu, B. and Zhang, L. (2012). *A Survey of Opinion Mining and Sentiment Analysis*, pages 415–463. Springer US, Boston, MA, EUA.
- Ravi, K. and Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89:14–46.
- Sanders, N. J. (2011). Twitter sentiment corpus.
- Smailović, J., Grčar, M., Lavrač, N., and Žnidaršič, M. (2014). Stream-based active learning for sentiment analysis in the financial domain. *Inf. Sci.*, 285(C):181–203.
- Souza, E., Vitória, D., Castro, D., Oliveira, A. L. I., and Gusmão, C. (2016). Characterizing opinion mining: A systematic mapping study of the portuguese language. In

- Silva, J., Ribeiro, R., Quaresma, P., Adami, A., and Branco, A., editors, *Computational Processing of the Portuguese Language*, pages 122–127, Cham. Springer International Publishing.
- Wagner, S., Zimmermann, M., Ntoutsis, E., and Spiliopoulou, M. (2015). Ageing-based multinomial naive bayes classifiers over opinionated data streams. In *Proceedings of the 2015th European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part I, ECMLPKDD'15*, pages 401–416, Switzerland. Springer.
- Wang, D., Feng, S., Wang, D., and Yu, G. (2013). Detecting opinion drift from chinese web comments based on sentiment distribution computing. In Lin, X., Manolopoulos, Y., Srivastava, D., and Huang, G., editors, *Web Information Systems Engineering – WISE 2013*, pages 72–81, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Widmer, G. and Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Mach. Learn.*, 23(1):69–101.
- Yang, Y. and Loog, M. (2018). A benchmark and comparison of active learning for logistic regression. *Pattern Recognition*, 83:401 – 415.
- Zhu, X., Zhang, P., Lin, X., and Shi, Y. (2007). Active learning from data streams. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 757–762.
- Zhu, X., Zhang, P., Lin, X., and Shi, Y. (2010). Active learning from stream data using optimal weight classifier ensemble. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(6):1607–1621.
- Zimmermann, M., Ntoutsis, E., and Spiliopoulou, M. (2014a). Adaptive semi supervised opinion classifier with forgetting mechanism. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing, SAC '14*, pages 805–812, New York, NY, USA. ACM.
- Zimmermann, M., Ntoutsis, E., and Spiliopoulou, M. (2014b). A semi-supervised self-adaptive classifier over opinionated streams. In *2014 IEEE International Conference on Data Mining Workshop*, pages 425–432.
- Zimmermann, M., Ntoutsis, E., and Spiliopoulou, M. (2015). Incremental active opinion learning over a stream of opinionated documents. *arXiv preprint arXiv:1509.01288*.
- Žliobaitė, I., Bifet, A., Pfahringer, B., and Holmes, G. (2011). Active learning with evolving streaming data. In Gunopulos, D., Hofmann, T., Malerba, D., and Vazirgiannis, M., editors, *Machine Learning and Knowledge Discovery in Databases*, pages 597–612, Berlin, Heidelberg. Springer Berlin Heidelberg.