Resampling approaches to improve news importance prediction

Nuno Moniz[1], Luís Torgo[1], and Fátima Rodrigues[2]

[1] LIAAD-INESC TEC / FCUP-DCC, University of Porto
[2] GECAD - ISEP/IPP, ISEP-DEI, Polytechnic of Porto
nmmoniz@liaad.up.pt, ltorgo@dcc.fc.up.pt, mfc@isep.ipp.pt

Abstract. The methods used to produce news rankings by recommender systems are not public and it is unclear if they reflect the real importance assigned by readers. We address the task of trying to forecast the number of times a news item will be tweeted, as a proxy for the importance assigned by its readers. We focus on methods for accurately forecasting which news will have a high number of tweets as these are the key for accurate recommendations. This type of news is rare and this creates difficulties to standard prediction methods. Recent research has shown that most models will fail on tasks where the goal is accuracy on a small sub-set of rare values of the target variable. In order to overcome this, resampling approaches with several methods for handling imbalanced regression tasks were tested in our domain. This paper describes and discusses the results of these experimental comparisons.

1 Introduction

The Internet is becoming one of the main sources for users to collect news concerning their topics of interest. News recommender systems provide help in managing the huge amount of information that is available. A typical example of these systems is Google News, a well-known and highly solicited robust news aggregator counting several thousands of official news sources. Although the actual process of ranking the news that is used by Google News is not known, official sources state that it is based on characteristics such as freshness, location, relevance and diversity. This process, based on the Page Rank algorithm explained in Page et al. [16] and generally described in Curtiss et al. [4], is of the most importance as it is responsible for presenting the best possible results for a user query on a set of given terms. However, some points have been questioned such as the type of documents that the algorithm gives preference to and its effects, and some authors conclude that it favours legacy media such as print or broadcast news "over pure players, aggregators or digital native organizations" [6]. Also, as it seems, this algorithm does not make use of the available information concerning the impact in or importance given by real-time users in an apparent strategy of deflecting attempts of using its capabilities in ones personal favour.

This paper describes some initial attempts on a task that is part of a larger project that tries to merge the news recommendations provided by two types

of sources: (i) official media that will be represented by Google News; and (ii) the recommendations of Internet users as they emerge from their social network activity. Concerning this latter source, the idea is to use Twitter as the source of information for checking which news are being shared the most by users. We will use the number of tweets of a given news as a kind of proxy for its impact on consumers of news, on a given topic[3]. The workflow solved by the system we plan to develop within this project is the following:

1. At a time t a user asks for the most interesting news for topic X
2. The system returns a ranking of news (a recommendation) that results from aggregating two other rankings:
   - The ranking provided by the official media (represented by Google News)
   - The ranking of the consumers of news (represented by the number of times the news are tweeted)

This workflow requires that at any point in time we are able to anticipate how important a news will be in Twitter. If some time has already past since the news publication this can be estimated by looking at the observed number of times this news piece was tweeted. However, when the news is very recent, the number of already observed tweets will potentially under-estimate the attributed importance of the news item. In this context, for very "fresh" news we need to be able to estimate their future relevance within Twitter, i.e. their future number of tweets. Moreover, given that we are interested in using this estimated number of tweets as a proxy for the news relevance, we are only interested in forecasting this number accurately for news with high impact, i.e. we are interested in being accurate at forecasting the news that will have a high number of tweets. This is the goal of the work presented in this paper. We describe an experimental comparison of different approaches to forecasting the future number of tweets of news, when the goal is predictive accuracy for highly popular news. This latter aspect is the key distinguishing aspect of our work when compared to existing related work, and it is motivated by the above-mentioned long-term project of news recommendation integrating different types of rankings. In this context, the main contribution of this work is a study and proposal of approaches to the problem of predicting highly popular news upon their publication.

2 Previous Work

In our research, although we did not find work that deals with the prediction of rare cases of highly tweeted news events, we did find important work focused on the general prediction of the number of tweets a news events will obtain in a given future.

[3] We are aware that this may be a debatable assumption, still this is something objective that can be easily operationalised, whilst other alternatives would typically introduce some subjectivity that would also be questionable.

Leskovec et al. [15] suggests that popular news take about four days until their popularity stagnates. Our own research using the data collected for this paper suggests that tweeting of a news item very rarely occours after two days.

Related to the subject in this paper, Asur and Huberman [1] use Twitter to forecast the box-office revenues for movies by building linear regression models, as well as demonstrate the utility that sentiment analysis has in the improvement of such objectives.

In Bandari et al. [2] classification and regression algorithms are examined in order to predict popularity, translated as the number of tweets, of articles in Twitter. The distinguishing factor of this work from others [19, 13, 20, 11, 14] that attempt to predict popularity of items, is that this work attempts to do this prior to the publication of the item. To this purpose the authors used four features: source of the article, category, subjectivity in the language and named entities mentioned. Furthermore, the authors conclude that the source of a given article is one of the most important predictors.

Regression, classification and hybrid approaches are used by Gupta et al. [7] also to predict event popularity. However, in this work the authors use data from Twitter, such as the number of followers/followees. The objective is the same in the work of Hsieh et al. [9], but the authors approach the problem by improving crowd wisdom with the proposal of two strategies: combining crowd wisdom with expert wisdom and reducing the noise by removing "overly talkative" users from the analysis.

Recently, a Bayesian approach was proposed by Zaman et al. [25] where a probabilistic model for the evolution of the retweets was developed. This work differs from the others in a significant manner as it is focused on the prediction of the popularity of a given tweet. The authors conclude that it is possible to predict the popularity of a given tweet after 10 minutes of its publication. They state that the number of tweets after two hours of its publication should improve by roughly 50% in relation to the first ten minutes. The test cases include both famous and non-famous twitter accounts. This work is preceded by others also using the retweet function as predictor having as the objective result an interval [8] or the probability of being retweeted.

3 Problem Description and Approach

This work addresses the issue of predicting the number of tweets of very recent news events with a focus on the predictive accuracy at news that are highly tweeted. This is a numeric prediction task, where we are trying to forecast this number based on some description of the news. However, this task has one particularity: we are only interested in prediction accuracy at a small sub-set of the news - the ones that are tweeted the most. These are the news that the public deems as highly relevant for a given topic and these are the ones we want to put at the top of our news recommendation. The fact that we are solely interested on being accurate at a low frequency range of the values of the target variable (the number of tweets) creates serious problems to standard prediction methods.

In this paper we describe and test several approaches that try to improve the predictive performance on this difficult task.

3.1 Formalization of the Data Mining Task

Our goal of forecasting the number of tweets of a given news is a numeric prediction task, usually known as a regression problem. This means that we assume that there is an unknown function that maps some characteristics of the news into the number of times this news is tweeted, i.e. $Y = f(X_1, X_2, \cdots, X_p)$, where Y is the number of tweets in our case, $X_1, X_2, \cdots, X_p$ are features describing the news and $f()$ is the unknown function we want to approximate. In order to obtain an approximation (a model) of this unknown function we use a data set with examples of the function mapping (known as a training set), i.e. $D = \{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^{n}$.

The standard regression tasks we have just formalized can be solved using many existing algorithms, and most of them try to find the model that optimizes a standard error criterion like the mean squared error. What sets our specific task apart is the fact that we are solely interested in models that are accurate at forecasting the rare and high values of the target variable Y, i.e. the news that are highly tweeted. Only this small sub-set of news is relevant for our overall task of providing a ranking of the most important news for a given topic. In effect, predictive accuracy at the more common news that have a small number of tweets is completely irrelevant because only the top positions of the ranking of recommended news are really relevant for the user, and these top positions are supposed to be filled by the news that have a very high number of tweets.

3.2 Handling the Imbalanced Distribution of the Number of Tweets

Previous work [17, 22, 23] has shown that standard regression tools fail dramatically on tasks where the goal is accuracy at the rare extreme values of the target variable. The main goal of the current paper is to compare some of the proposed solutions to this type of imbalanced regression tasks in the particular problem of forecasting the number of tweets of news.

Several methodologies were proposed for addressing this type of tasks. Resampling methods are among the simplest and most effective. Resampling strategies work by changing the distribution of the available training data in order to meet the preference bias of the users. Their main advantage is that they do not require any special algorithms to obtain the models - they work as a preprocessing method that creates a "new" training set upon which one can apply any learning algorithm. In this paper we will experiment with two of the most successful resampling strategies: (i) SMOTE [3] and (ii) under-sampling [12]. These methods were originally developed for classification tasks where the target variable is nominal. The basic idea of under-sampling is to decrease the number of observations with the most common target variable values with the goal of better balancing the ratio between these observations and the ones with the interesting target values that are less frequent. SMOTE works by combining

under-sampling of the frequent classes with over-sampling of the minority class. Namely, new cases of the minority class are artificially generated by interpolating between existing cases. Recently, Torgo et al. [23, 24] extended these methods for regression tasks as it is the case of our problem. We have used the work of these authors to create two variants of each of our datasets. The first variant uses the SMOTEr algorithm [23] to create a new training set by over-sampling the cases with extremely large number of tweets, and under-sampling the most frequent cases, thus balancing the resulting distribution of the target variable. The second variant uses the under-sampling algorithm proposed by the same authors to decrease the number of cases with low number of tweets, hence the most common, once again resulting in a more balanced distribution. In our experiments we will apply and compare these methodologies in order to check which one provides better results in forecasting accurately the number of tweets of highly popular news items. As explained in detail in Section 4.3 these resampling strategies are framed in an utility-based regression framework (Torgo and Ribeiro [22] and Ribeiro [17]) which maps the values of the target variable into a $[0, 1]$ scale of relevance. By defining a relevance threshold, this scale is then used to define the sub-range of target variable values as rare and common. The SMOTEr and under-sampling strategies use this information to over- and under-sampling the values to balance the distribution.

4 Materials and Methods

4.1 The Used Data

The experiments that we will describe are based on news concerning four specific topics: economy, microsoft, obama and palestine. These topics were chosen due to two factors: its actual use and because they report to different types of entities (sector, company, person and country). For each of these four topics we have constructed a dataset with news mentioned in Google News during a period of 13 days, between 2013-Nov-15 and 2013-Nov-28. Figure 1 shows the number of news per topic during this period.

For each news obtained from Google News the following information was collected: title, headline and publication date. For each of the four topics a dataset was built for solving the predictive task formalized in Section 3.1. These datasets were built using the following procedure. For obtaining the target variable value we have used the Twitter API[4] to check the number of times the news was tweeted in the two days following its publication. These two days limit was decided based on the work of Leskovec et al. [15] that suggests that after a few days the news stop being tweeted. Despite Leskovec et al. [15] statement that this period is of four days, some initial tests on our specific data sets have shown that after a period of two days the number of tweets is residual, and therefore we chose this time interval. In terms of predictor variables used to describe each news we have selected the following. We have applied a standard bag of words

[4] Twitter API Documentation: https://dev.twitter.com/docs/api

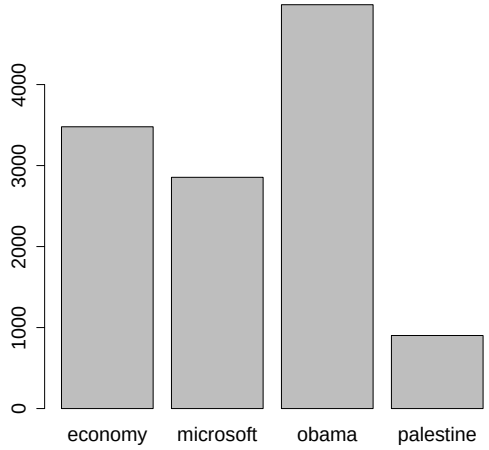

Fig. 1. Number of news per topic

approach to the news headline to obtain a set of terms describing it[5]. Some initial experiments we have carried out have shown that the headline provides better results than the title of the news item. We have not considered the use of the full news text as this would require following the available link to the original news site and have a specific crawler to grab this text. Given the wide diversity of news sites that are aggregated by Google News, this would be an unfeasible task. To this set of predictors we have added two sentiment scores: one for the title and the other for the headline. These two scores were obtained by applying the function `polarity()` of the R package **qdap** [18] that is based on the sentiment dictionary described by Hu and Liu [10]. Summarizing, our four datasets are built using the information described on Table 1 for each available news.

As expected, the distribution of the values of the target variable for the four obtained datasets is highly skewed. Moreover, as we have mentioned our goal is the accuracy at the low frequency cases where the number of tweets is very high. We will apply the different methods described in Section 3.2 to our collected data. This will lead to 12 different datasets, three for each of the selected topics: (i) the original imbalanced dataset; (ii) the dataset balanced using SMOTEr; and (iii) the dataset balanced using under-sampling. The hypothesis driving the current paper is that by using the re-sampled variants of the four original datasets we will gain predictive accuracy at the highly tweeted news, which are the most relevant for providing accurate news recommendations.

[5] We have used the infra-structure provided by the R package **tm** [5].

Variable	Description
NrTweets	The number of times the news was tweeted in the two days following its publication. This is the target variable.
$T_1, T_2, \cdots$	The term frequency of the terms selected through the bag of words approach when applied to all news headlines.
SentTitle	The sentiment score of the news title.
SentHeadline	The sentiment score of the news headline.

Table 1: The variables used in our predictive tasks

4.2 Regression Algorithms

In order to test our hypothesis that using resampling methods will improve the predictive accuracy of the models on the cases that matter to our application, we have selected a diverse set of regression tools. Our goal here is to try to make sure our conclusions are not biased by the choice of a particular regression tool.

Table 2 shows the regression methods and tools that were used in our experiments. To make sure our work can be easily replicable we have used the implementations of these tools available at the free and open source R environment. All tools were applied using their default parameter values.

ID	Method	R package
RF	Random forests	randomForest
LM	Multiple linear regression	stats
SVM	Support vector machine	e1071
MARS	Multivariate adaptive regression splines	earth

Table 2: Regression algorithms and respective R packages

4.3 Evaluation Metrics

It is a well-known fact that when the interest of the user is a small proportion of rare events, the use of standard predictive performance metrics will lead to biased conclusions. In effect, standard prediction metrics focus on the "average" behaviour of the prediction models and for these tasks the user goal is a small and rare proportion of the cases. Most of the previous studies on this type of problems was carried out for classification tasks, however, Torgo and Ribeiro [22] and Ribeiro [17] have shown that the same problems arise on regression tasks when using standard metrics like for instance the Mean Squared Error. Moreover, these authors have shown that discretizing the target numeric variable into a nominal variable followed by the application of classification algorithms is also prone to problems and leads to sub-optimal results.

In this context, we will base our evaluation on the utility-based regression framework proposed in the work by Torgo and Ribeiro [22] and Ribeiro [17]. The metrics proposed by these authors assume that the user is able to specify what is the sub-range of the target variable values that is most relevant. This is done by specifying a relevance function that maps the values of the target variable into a $[0, 1]$ scale of relevance. Using this mapping and a user-provided relevance threshold the authors defined a series of metrics that focus the evaluation of models on the cases that matter for the user. In our experiments we have used as relevance threshold the value of 0.9, which leads to having on average 7% to 10% of the cases tagged as rare (i.e. important in terms of number of tweets) depending on the topic.

In our evaluation process we will mainly rely on two utility-based regression metrics: a variant of the mean squared error weighed by relevance, and the F-Score. The variant of the mean squared error (`mse_phi` in our tables of results) is calculated by multiplying each error by the relevance of the true number of tweets. This means that the errors on the most relevant news will be amplified. The main problem of this metric is that it does not consider situations were the models forecast a high number of tweets for a news that ends up having a low number of tweets, i.e false positives. On the contrary, the F-Score is able to take into account both problems. This is a composite measure that integrates the values of precision and recall according to their adaptation for regression described in the above mentioned evaluation framework.

5 Experimental Comparison

5.1 Experimental Methodology

Our data (news items) has a temporal order. In this context, one needs to be careful in terms of the process used to obtain reliable estimates of the selected evaluation metrics. This means that the experimental methodology should make sure that the original order of the news is kept so that models are trained on past data and tested on future data to avoid over-optimistic estimates of their scores. In this context, we have used Monte Carlo estimates as the experimental methodology to obtain reliable estimates of the selected evaluation metrics for each of the alternative methodologies. This methodology randomly selects a set of points in time within the available data, and then for each of these points selects a certain past window as training data and a subsequent window as test data, with the overall train+test process repeated for each point. All alternative approaches are compared using the same train and test sets to ensure fair pairwise comparisons of the obtained estimates. Our results are obtained through 10 repetitions of a Monte Carlo estimation process with 50% of the cases used as training set and 25% used as test set. This process is carried out in R using the infra-structure provided by the R package **performanceEstimation** [21].

5.2 Results

Our results contemplate four topics, as referred before: economy, microsoft, obama and palestine. Tables 3 and 4 present a summary of our results, with other results ommited due to space economy reasons though the general trends are similar. For each regression algorithm the best estimated scores are denoted in italics, whilst the best overall score is in bold. Table 3 presents all estimated metric scores for the palestine topic data set. The mean squared error estimates are provided for highlighting once again that this type of metrics may mislead users when the focus is accuracy on rare extreme values of the target variable. The last four metrics are the most interesting from the perspective of our application, particularly the last three (precision, recall and F-measure) because they also penalise false positives (i.e. predicting a very high number of tweets for a news that is not highly tweeted). These results clearly show that in most setups all algorithms are able to take advantage of resampling strategies to clearly boost their performance. The results obtained with both random forests and SVMs are particularly remarkable, moreover taking into account that all methods were applied with their default parameter settings. With precision scores around 60% we can assume that if we use the predictions of these models for ranking news items by their predicted number of tweets, the resulting rank will match reasonably well the reading preferences of the users.

	mse	mse_phi	prec	rec	F1
lm	*4622.81*	*1651.31*	0.28	0.14	0.18
lm+SMOTE	763343.67	115671.61	0.50	0.13	0.20
lm+UNDER	240787.13	36782.38	*0.57*	*0.16*	*0.24*
svm	**1681.23**	1543.91	0.00	0.00	0.00
svm+SMOTE	7845.29	662.88	*0.59*	**0.57**	*0.57*
svm+UNDER	7235.22	*645.46*	0.57	0.56	0.56
mars	*2334.96*	*1490.63*	0.41	0.10	0.16
mars+SMOTE	17238.67	1971.88	0.43	0.25	0.31
mars+UNDER	15291.91	1514.40	*0.49*	*0.36*	*0.41*
rf	*1770.16*	1438.93	0.23	0.04	0.06
rf+SMOTE	6309.61	636.77	0.49	0.50	0.49
rf+UNDER	8170.99	**618.81**	*0.61*	*0.56*	**0.58**

Table 3: Prediction Models Results - Topic Palestine.

Table 4 shows the overall estimated precision, recall and F1 scores for all alternatives in the four topics. Once again we confirm that using resampling strategies provides very good results in the task of predicting highly tweeted news for these four diverse topics. These results are also in accordance with the findings by Torgo et al. [23] where it was reported that similar gains were obtained with both under-sampling and SMOTEr. These results show that in all experimental settings we have considered, the use of resampling was able to

clearly improve the precision of the models at identifying the news that will be highly tweeted by users.

	economy			microsoft			obama			palestine		
	prec	rec	F1	prec	rec	F1	prec	rec	F1	prec	rec	F1
lm	0.15	*0.12*	*0.13*	0.30	*0.11*	*0.16*	0.15	*0.07*	*0.10*	0.28	0.14	0.18
lm+SMOTE	0.30	0.02	0.04	*0.60*	0.04	0.07	**0.55**	0.04	0.07	0.50	0.13	0.20
lm+UNDER	*0.41*	0.02	0.04	0.58	0.03	0.05	0.38	0.03	0.05	*0.57*	*0.16*	*0.24*
svm	0.00	0.00	0.00	0.37	0.03	0.05	0.00	0.00	0.00	0.00	0.00	0.00
svm+SMOTE	*0.41*	**0.51**	*0.45*	0.46	0.39	0.42	*0.50*	**0.56**	**0.53**	*0.59*	**0.57**	*0.57*
svm+UNDER	0.37	0.47	0.41	*0.48*	*0.41*	*0.44*	0.48	0.55	0.51	0.57	0.56	0.56
mars	0.16	0.10	0.12	0.35	0.10	0.15	0.09	0.04	0.06	0.41	0.10	0.16
mars+SMOTE	0.40	0.20	0.27	0.56	0.23	0.32	0.40	0.17	0.24	0.43	0.25	0.31
mars+UNDER	*0.42*	*0.33*	*0.36*	*0.58*	*0.31*	*0.40*	*0.41*	*0.27*	*0.32*	*0.49*	*0.36*	*0.41*
rf	0.11	0.05	0.07	0.29	0.06	0.10	0.02	0.01	0.01	0.23	0.04	0.06
rf+SMOTE	0.45	0.48	0.46	0.54	0.46	0.50	0.34	0.43	0.38	0.49	0.50	0.49
rf+UNDER	**0.54**	*0.49*	**0.51**	**0.63**	**0.52**	**0.56**	*0.46*	*0.47*	*0.46*	**0.61**	*0.56*	**0.58**

Table 4: The F1 estimated scores for all topics.

Overall, the main conclusion from our comparisons is that resampling methods are very effective in improving the predictive accuracy of different models for the specific task of forecasting the number of tweets of highly popular news. These methods are able to overcome the difficulty of these news being infrequent. This is particularly important within our application goal that requires us to be able to accurately identify the news that are more relevant for the users in order to be able to improve the performance of news recommender systems.

6 Conclusions

This paper describes an experimental analysis of different methods of predicting the number of times a news item will be tweeted. Being able to forecast accurately this number for news that will be highly tweeted is very important for effective news recommendation. These news are rare and this poses difficult challenges to existing prediction models. We evaluate recently proposed methods for addressing these problems in our particular task.

The results of our experimental comparisons clearly confirm the hypothesis that using resampling methods is an effective and simple way of addressing the task of predicting when a news item will be highly tweeted. Our results, under different experimental settings and using different prediction algorithms, clearly indicate that resampling is able to boost the accuracy of the models on cases that are relevant for this application. In particular, we have observed a marked increase of the precision of the models, which means that most of the times when they forecast that a news will be highly tweeted, that will happen. This is very important for this application as it means that rankings produced based

on these predictions will be useful for users as they will suggest news items that are effectively interesting for them.

Future work will include the addition of more information such as the source of the news article and the use of the predictions of the different alternative ways of forecasting the number of tweets to produce actual news rankings. Moreover, we will use methods for comparing rankings in order to correctly measure the impact of the use of resampling methods on the quality of news recommendation.

7 Acknowledgments

This work is supported by SIBILA Project "NORTE-07-0124-FEDER-000059", financed by the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT). The work of N. Moniz is supported by a PhD scholarship of the Portuguese government (SFRH/BD/90180/2012).

References

1. S. Asur and B. A. Huberman. Predicting the future with social media. In *Proc. of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '10, pages 492–499. IEEE Computer Society, 2010.
2. R. Bandari, S. Asur, and B. A. Huberman. The pulse of news in social media: Forecasting popularity. *CoRR*, 2012.
3. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *JAIR*, 16:321–357, 2002.
4. M. Curtiss, K. Bharat, and M. Schmitt. Systems and methods for improving the ranking of news articles, March 17 2005. US Patent App. 10/662,931.
5. I. Feinerer, K. Hornik, and D. Meyer. Text mining infrastructure in r. *Journal of Statistical Software*, 5(25):1–54, 2008.
6. F. Filloux and J. Gassee. Google news: The secret sauce. *Monday Note*, 2013. URL `http://www.mondaynote.com/2013/02/24/google-news-the-secret-sauce/`.
7. M. Gupta, J. Gao, C. Zhai, and J. Han. Predicting future popularity trend of events in microblogging platforms. *ASIS&T 75th Annual Meeting*, 2012.
8. L. Hong, B. Dom, S. Gurumurthy, and K. Tsioutsiouliklis. A time-dependent topic model for multiple text streams. In *Proc. of the 17th ACM SIGKDD*, KDD '11, pages 832–840. ACM, 2011.
9. C. Hsieh, C. Moghbel, J. Fang, and J. Cho. Experts vs. the crowd: examining popular news prediction performance on twitter. In *Proc. of ACM KDD conference*, 2013.
10. M. Hu and B. Liu. Mining opinion features in customer reviews. In *Proc. of the 19th National Conference on Artificial Intelligence (AAAI'04)*, 2004.

11. S. Kim, S. Kim, and H. Cho. Predicting the virtual temperature of web-blog articles as a measurement tool for online popularity. In *Proc. of the 2011 IEEE 11th International Conference on Computer and Information Technology*, CIT '11, pages 449–454. IEEE Computer Society, 2011.
12. M. Kubat and S. Matwin. Addressing the curse of imbalanced training sets: One-sided selection. In *Proc. of the 14th Int. Conf. on Machine Learning*, pages 179–186, Nashville, TN, USA, 1997. Morgan Kaufmann.
13. J. G. Lee, S. Moon, and K. Salamatian. An approach to model and predict the popularity of online contents with explanatory factors. In *Proc. of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '10, pages 623–630. IEEE Computer Society, 2010.
14. K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proc. of the 19th international conference on World Wide Web*, WWW '10, pages 621–630. ACM, 2010.
15. J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 497–506. ACM, 2009.
16. L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University, 1998.
17. R. Ribeiro. *Utility-based Regression*. PhD thesis, Dep. Computer Science, Faculty of Sciences - University of Porto, 2011.
18. T. W. Rinker. *qdap: Quantitative Discourse Analysis Package*. University at Buffalo/SUNY, 2013.
19. G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Commun. ACM*, 53(8):80–88, August 2010.
20. A. Tatar, J. Leguay, P. Antoniadis, A. Limbourg, M. D. de Amorim, and S. Fdida. Predicting the popularity of online articles based on user comments. In *Proc. of the International Conference on Web Intelligence, Mining and Semantics*, WIMS '11, pages 67:1–67:8. ACM, 2011.
21. L. Torgo. *An Infra-Structure for Performance Estimation and Experimental Comparison of Predictive Models*, 2013. URL `https://github.com/ltorgo/performanceEstimation`.
22. L. Torgo and R. Ribeiro. Utility-based regression. In Springer, editor, *PKDD'07: Proc. of 11th European Conf. on Principles and Practice of Knowledge Discovery in Databases*, pages 597–604, 2007.
23. L. Torgo, R. P. Ribeiro, B. Pfahringer, and P. Branco. Smote for regression. In L. Correia, L. P. Reis, and J. Cascalho, editors, *EPIA*, volume 8154 of *Lecture Notes in Computer Science*, pages 378–389. Springer, 2013.
24. L. Torgo, P. Branco, R. Ribeiro, and B. Pfahringer. Re-sampling strategies for regression. *Expert Systems*, (to appear), 2014.
25. T. Zaman, E. B. Fox, and E. T. Bradlow. A Bayesian Approach for Predicting the Popularity of Tweets. Technical Report arXiv:1304.6777, Apr 2013.