The 4th International Conference on Ambient Systems, Networks and Technologies (ANT 2013)

Bridging the gap between the least and the most influential Twitter users *

J. del Campo-Ávila, N. Moreno-Vergara and M. Trella-López
Dpto. Lenguajes y Ciencias de la Computación,
University of Málaga, Spain
E-mail: {jcampo,moreno,trella}@lcc.uma.es

Abstract

Social networks play an increasingly important role in shaping the behaviour of users of the Web. Conceivably Twitter stands out from the others, not only for the platform's simplicity but also for the great influence that the messages sent over the network can have. The impact of such messages determines the influence of a Twitter user and is what tools such as Klout, PeerIndex or TwitterGrader aim to calculate. Reducing all the factors that make a person influential into a single number is not an easy task, and the effort involved could become useless if the Twitter users do not know how to improve it. In this paper we identify what specific actions should be carried out for a Twitterer to increase their influence in each of above-mentioned tools applying, for this purpose, data mining techniques based on classification and regression algorithms to the information collected from a set of Twitter users.

1 Introduction

Twitter [1] is an information network made up of 140-character messages called tweets. It is an easy way to discover the latest news (i.e. *what's happening*) related to subjects or people that are important to you. The reasons why users are interested in this free platform are as varied and different as the user profiles themselves but maybe many of them share a common goal: to become influential Twitter users. Therefore a question arises, what does it mean to be an influential Twitter user? In this regard, there does not seem to be any consensus on what exactly *influence* means in the Twittersphere. Several definitions try to determine what aspects are relevant to explaining the impact (trust, power, authority, reach, connection, value,...) of a user in the community. At the simplest level, Twitter's influence can be defined as *a measure of the ability to cause desirable and measurable actions and outcomes* [2, 3].

Reducing influence to a number is certainly a difficult simplification that a considerable number of analytical tools such as Klout [4], PeerIndex [5] or TwitterGrader [6] try to compute. Each one of these tools implements its own proprietary algorithm to estimate influence based upon attributes related to both user contact network topology (number of followers, followers/following ratio, frequency with which the user is mentioned, and so on) and the traffic tweets (number of retweets, replies and mentions, among other things). Nevertheless, none of these tools are very transparent about how the score is calculated since the algorithm is a competitive asset and disclosure would inevitably encourage people to manipulate the system. However, without that knowledge, the value computed by these systems is insignificant because, on the one hand, no guidelines are available as to which tool should be used in each case and, on the other hand, the score does not tell us anything about how to improve the ranking. So, in this context, could we identify which parameters are more relevant for each tool and its corresponding estimated ranking?

Several recent efforts have been made to track influence on Twitter [3, 7, 8, 9]. Some of them have focused on the Twitter network topology only and others have more directly focused on the Twitterers activity as an indicator of influence. Taken together, it can be concluded that the static graph is, at best, a mediocre indicator of who is actually influential on Twitter and, although, analyzing activity is a key step in identifying influential users–taken in isolation–it is not enough as what users need are a set of specific actions they could carry out to increase their own influence.

*This article is the Accepted Author Manuscript (AAM).

Paper published in Procedia Computer Science, Volume 19, 2013, Pages 437-444, ISSN 1877-0509
DOI: http://dx.doi.org/10.1016/j.procs.2013.06.059

Addressing this problem requires advanced techniques to derive hidden correlations between the information that Twitter stores about users and their particular degrees of influence. That is one of the tasks that data mining and, in particular Web data mining [10, 11, 12] can perform. They use statistics and machine learning algorithms in order to discover relevant knowledge from a large amount of data by applying different computation strategy types such as classification, regression or segmentation algorithms.

In this paper, we analyze different Twitter influence metrics and tools (Section 2) and perform a statistical study. After that, in Section 3, we apply data mining techniques to a dataset built with information collected from Twitter users in order to: (1) identify the actions which can increase the influence of a user depending on the most popular Twitter influence analytical tool and (2) discover if we can improve influence in more than one way. Finally, we conclude with directions for further research in Section 4.

2 Analytical Tools to Calculate the Twitter Influence

As we have already mentioned, there is no one way to calculate "influence". Up until now, maybe the most comercial of the existent measurement tools is Klout (which is even being used in scientific papers [13]). Other sources of social media metrics measurement commonly used are PeerIndex and TwitterGrader [13, 14]. The three of them assign the user a score in the rage from 1 to 100, higher scores represent a wider and stronger sphere of influence. As we have mentioned in the previous Section, the algorithms used to estimate the influence level are not public. The owner companies provide imprecise descriptions of them. The users can know that in one way or another they use data provided by the Twitter's API (number of followers, number of retweets, etc.). As well as these data, Klout and PeerIndex, use some derived metrics and information from other social networks.

2.1 Tools

Klout. The final Klout Score is a representation of how successful a user is at engaging their audience and how big of an impact their messages have on people. Recently, Klout added LinkedIn, Foursquare, Blogger, Tumblr, Flickr, Last.fm, Instagram, Google+ and YouTube data to its algorithm to calculate Klout's particular derived metrics. They are: *true reach*(i.e., the size of one's engaged audience based on followers and friends who actively pay attention to and react to messages), the *amplification probability* (i.e., the likelihood that one's messages will generate actions such as retweets, likes or comments), and *network score* (i.e., how influential the people who retweet, mention, list of who follows you are) .

PeerIndex. PeerIndex ranking tries to reflect the underlying value of what people say and who cares about what they are saying – if they are a VIP or otherwise (this does not really matter). It takes into account the relationships they build up on various social media platforms since the impact of those relationships also affects their authority exhibited on the web to such an extent that authority on a subject is affirmed when the content that is shared is approved. In order to do that they use data from Twitter, Facebook and LinkedIn. The score is broken down into five sub-scores: (1) *Authority*: it measures reliability and trust; (2) *Topic Resonance*: it measures influence on certain topics the user influences; (3)*Audience*: it measures how people respond to all posts; (4) *Activity*: it measures how much content is posted about a topic; (5) *Realness*: it determines whether the account is an actual person, a feed, or a spam account.

TwitterGrader. TwitterGrader checks the power of a twitter profile compared to millions of other users that have been graded. The TwitterGrader team makes the factors that go into the algorithm readily available although those factors are itemized in no particular order and which one gets the bigger chunk of the influence remains unknown. Specifically, the grading system takes into account the following parameters: number of followers, power of followers, updates frequency, follower/following ratio and engagement.

2.2 Comparation

Twitter User List. To compare and contrast the tools that are considered in this study, we needed to compile a list of Twitter users as well as their corresponding rank values according to such systems. In particular, the list has been created trying to make it as heterogeneous as possible (domain, popularity, participation, etc.). Thus, starting from an initial subset of twitterers chosen from lists with the top twitterers, well-known companies, categories in

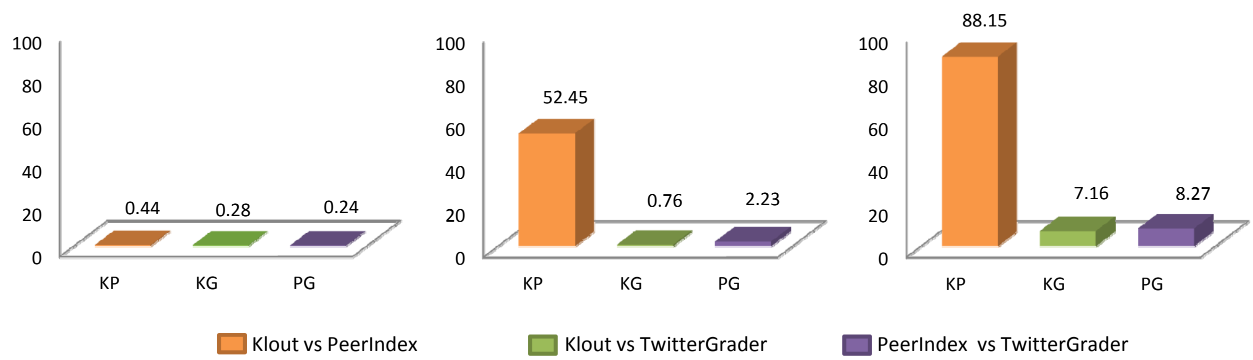


Figure 1: Percentage of users whose influence matches in the same quantile. From left to right, percentiles, deciles and quartiles

Table 1: Correlation ratio

	Klout	PeerIndex	TwitterGrader
Klout	1		
PeerIndex	0.7671	1	
TwitterGrader	0.0110	-0.0111	1

Twitter, etc., we have automatically gathered the rest by adding people following users in the initial subset until reaching a dataset with some tens of thousands of users.

Using the corresponding APIs of Klout, TwitterGrader and PeerIndex, we collected information about their calculated rankings. In fact, although this last step was automated, different problems arose during the collection (fail whale error, cancelation because of time delay, users without information, etc.) and not every rank was available for every user. Therefore, the final dataset does not have registered information for all the initial users, but, even with that limitation, the dataset was composed of approximately five thousand users. Due to restrictions derived from the terms of use of some tools when using their APIs, we are not able to distribute the final dataset, although the scripts used to automate this task, and the final list of twitterers can be downloaded from [15].

Correlation ratio between tools. Two of the three tools evaluated are in direct competition to become the facto standard for measuring the influence of social networks: Klout and PeerIndex. Despite the fact their algorithms evaluate distinct parameters and assign them different weights, only 0,44% of users obtain exactly the same score. This percentage rises to 52,45% for values that fall within the same decil. As is to be expected, this tendency keeps rising, up to 85,15% when we talk about values which coincide in the first quartile as shown in Figure 1.

Particularly for Klout and PeerIndex (see Table 1), the coefficient of correlation between both tools is of 0,76 that corresponds to a positive correlation in which the high values of influence as calculated by Klout are associated with high values of influence calculated with PeerIndex. Not so for the pairs of tools TwitterGrader-Klout and TwitterGrader-PeerIndex whose correlations are effectively zero, that is, there is not any linear relation between these systems. In no less than 7,16% Klout and TwitterGrader return values in the same quartile, an amount which rises no more than 1% rising to 8,27% between TwitterGrader and PeerIndex. However, maintaining TwitterGrader in this study can provide very useful information and can widen the variety of actions focused to increase the influence.

None of these three tools provide the answer on how to increase your influence and improve your scores other than the obvious: put out good content, engage people who are influential, pick a shot set of subjects that you consistently share opinions/content/links about and be highly responsive to those that comment back to you. However, not all such actions have the same impact and none level is suggested as a threshold that differentiates between the least and the most influential Twitter user. In the next Section we explain how the use of data mining techniques can help Twitter users to derive concrete rules which guide them in order to increase their influence.

3 Using Data Mining to Identify Dominant Parameters

When there are large datasets and extracting new relevant knowledge is required, data mining techniques become very useful [16]. Data mining is a discipline that uses statistics and machine learning algorithms in order to discover models, patterns or relationships in the data that are being studied. Bearing this in mind, we want to know there is any relationship between the simple parameters collected from Twitter, the specific metrics calculated by some of the tools mentioned in Subsection 2.1 and the influence estimated by those tools in order to: (a) identify actions to increase the influence for a particular analytical tool and (b) discover rules that can be applied in more than one tool to simultaneously increment the influence.

3.1 Datasets Definition and Data Mining Tasks

We have used the list of Twitter users described in Subsection 2.2 and we have augmented it with more information. Thus, we have defined three datasets (one dataset for each tool) that incorporate attributes obtained from Twitter such as *following count*, *follower count*, etc. and the ranking values estimated by the three analytical tools. Additionally, two extended datasets (marked with * in Table 2 have also been created by including the specific metrics calculated by Klout and PeerIndex (for example *true_reach* or *authority*).

Considering these datasets, in the context of machine learning, multiple algorithms can be used to develop different data mining tasks. In our case we are interested in a dual approximation: quantitative and qualitative. In the first case regression techniques are the most appropriate for our purposes since the class attributes with information to be learned from are numerical (ranking values taken from the studied tools) and we want to give a detailed approximation. On the other hand, we are also interested in simplifying the problem (from the point of view of the twitter user) and we reduce the numeric ranking to three discrete levels of influence (low, medium and high). In this case, classification techniques are suitable for the task of extracting knowledge with a qualitative perspective. Among all the possible methods and models available to represent the extracted knowledge, we have used those that induce decision trees because their rules are understandable by humans and can be easily translated into actions. Specifically we have used the regression (REPTREE and M5P [17, 18]) and classification (C4.5 [19]) algorithms implemented in Weka [20].

Above mentioned algorithms use variance reduction and information gain ratio as splitting criteria to select the most relevant attributes to expand the tree. But the knowledge contained in the model surpasses the simply detection of relevant attributes, and go beyond the feature subset selection problem, detecting how the values of those attributes are related (different attributes in different branches, thresholds for every attribute, etc.). Anyway, the usage of the attributes selected in the models does not discard the possibility that other attributes could have similar (but slightly lower) importance.

3.2 Analysis of Induced Models

The models generated usually reach good quality results it terms of accuracy as can be seen in Table 2. We show some performance metrics achieved by the different algorithms for every dataset (10-fold cross validation has been conducted). All the outputs (including the models and many more performance metrics) generated in the experimental process are available at [15].

How to increase the influence according to Klout. The regression models that have been induced, using REPTREE or M5P, achieve similar levels of quality (correlation coefficient and relative absolute error). Although the regression trees generated by M5P seem simpler (smaller trees), this is because part of the regression complexity is in the rules created at the M5P's leaves, so REPTREE and M5P offer similar results in terms of complexity. As expected, the precision of the models increases when the specific measures calculated by the tools are considered (marked with * in Table 2). This reveals that those measures are really used by Klout and they hide part of the mechanism used to calculate user rankings.

In Figure 2 (left) the subtree for more influential users calculated by REPTREE for Klout is represented. Something that it is common to the entire tree can be seen here: the most important attributes to decide the ranking level are the *true_reach* and the *amplification* which are specific metrics calculated by Klout. The third metric, *network_score* also appears in the tree but in deeper levels (lower importance) and combined with other network attributes like the number of followers and followings. This suggests that the network structure is not the most important component for Klout to decide the ranking assigned to a twitter user, although it has some relevance.

Table 2: Metrics achieved by regression and classification algorithms (10-fold cross validation). Datasets without * only consider twitter attributes and the ranking attribute measured by the tool. Datasets marked with * include specific attributes calculated within the corresponding tool.

		Klout*	Klout	PeerIndex*	PeerIndex	TwitterGrader
REPTREE	Tree size (leaves)	425	263	341	119	229
	Correlation coef.	0.997	0.842	0.967	0.816	0.949
	Relat. abs. error (%)	6,326	49,776	11.562	56.201	9.9
M5P	Tree size (leaves)	60	28	28	10	73
	Correlation coef.	0.996	0.857	0.975	0.825	0.952
	Relat. abs. error (%)	5,592	47,318	7.263	55.156	8.479
C4.5	Tree size (leaves)	18	42	17	47	12
	Accuracy	0.962	0.683	0.953	0.654	0.978
	ROC area (average)	0.992	0.841	0.986	0.827	0.993

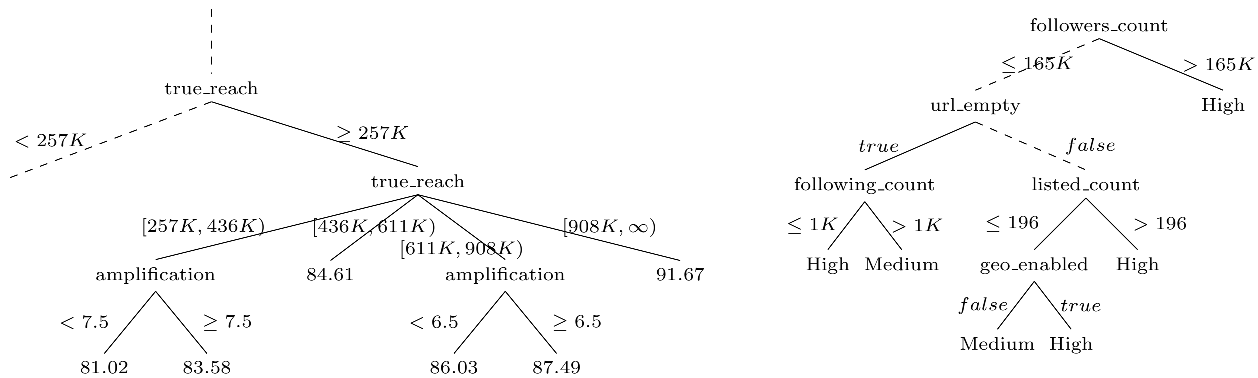
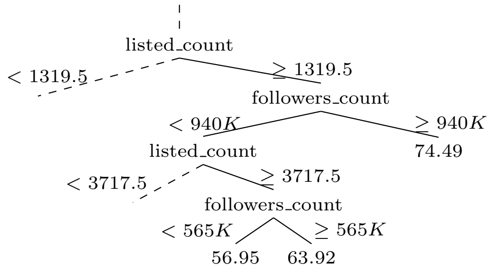


Figure 2: Decision tree models representing the most influential users for Klout *Left*: Regression subtree induced by REPTREE (dataset with specific Klout measures and Twitter attributes) *Right*: Classification subtree induced by C4.5 (dataset only with Twitter attributes) where it can be seen how personal information can influence the final ranking assigned.

When considering *classification trees*, it is also clear that the inclusion of specific metrics help to achieve more accurate models too. When variations of the dataset that do not contain specific metrics from Klout are used to train the models, the performance decreases, and the reason is the high dependency that the tools have with their own metrics. Anyway, the models achieve reasonably high levels of performance and some conclusions can be extracted. In the absence of those specific attributes, the important ones are, primarily, the number of lists where a user is included (*listed_count*) and the number of followers (*followers_count*): the greater the values of these attributes, the greater the ranking assigned. Other attributes, not as important as those mentioned previously, but that can produce some improvement in the final rank are related with the personal information available for a given user. It is observed (an example is given on right side of Figure 2) that enabling geolocation, or personal information (like URL or profile image) clearly acts on the increase of the influence level of a twitterer. So, it seems that the best way to improve the Klout influence score (or perhaps the fastest way) is by incrementing the set of interactions with other Twitter users using both Twitter platform and other social networks. In other words, we are increasing the *true_reach*.

How to increase the influence according to PeerIndex. Applying regression algorithms, PeerIndex also calculated the ranking value giving a great weight to their own specific metrics: *authority* and *audience*. To the contrary, *activity*, the third metric calculated, only appears in deeper levels of the tree. We would like to point out how this tool barely uses the number of followers and followings, which suggests (as occurred with Klout) that the relation between users is not a key component for PeerIndex to decide whether a user's rank should be in a low, medium or high level, but has some importance to tune up the final ranking inside that macro level.

When the dataset without PeerIndex metrics is used to train the models, the performance of the models decreases. Despite that aspect worsening, the models keep a considerable level of confidence and we can use them to extract some interesting knowledge. As can be seen in Figure 3 (left), the most important attributes for PeerIndex to assign




```
Influence-calculated-by-PeerIndex =
...
+ 0.9313 * verified=true
- 4.8796 * default_profile=false
...
+ 77.053
```

Figure 3: Regression subtree and rule representing the most influential users for PeerIndex. *Left*: Regression subtree induced by REPTREE (dataset only with Twitter attributes). *Right*: Regression rule at a leaf of a M5P tree where it can be seen how personal information can influence the final ranking assigned.

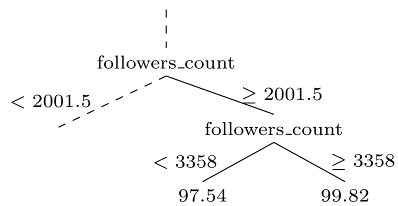


Figure 4: Regression subtree induced by REPTREE for TwitterGrader. It shows the importance of the number of followers to gaining high influence levels

the ranking to a user are the number of lists where he/she is labeled (*listed_count*) and the number of followers (*followers_count*). Other attributes also take part in the process, but to a lesser degree. However, the calculations made by PeerIndex could be influenced by some other attributes not yet mentioned. Thus, it seems that personal information can increase the final ranking in same cases. In this sense, Figure 3 (right) shows the regression rule induced in one of the leaves of a M5P tree. It can be seen how getting the verified tag or using a personal profile despite the default one increase the level of influence.

Analyzing the models of classification generated when training with the discretized datasets, we reach similar conclusions for regression models: the inclusion of specific metrics calculated by the tool increase the quality of the model. But, when we try to extract some other patterns, forcing the non use of those specific metrics, we get more general ideas about what is really important. Concretely, in absence of those attributes, the relevant ones are the number of lists where a user is included (*listed_count*), the number of followers (*followers_count*) and the number of messages (*statuses_count*). In this case, some attributes appear related with the information in the personal profile (*URL*, *location*, etc.) too. In summary, one of the best way to increase your influence according to PeerIndex is by sharing a lot of content and trying that such content gets commented upon or retweeted, i.e., incrementing your *authority*.

How to increase the influence according to TwitterGrader. TwitterGrader is a tool that seems to use simpler process to assign the influence level to a twitter user. The performance of the models induced by regression or classification learning algorihtms (see Table 2) is high and similar to those obtained for previous tools. Note that these high levels of quality are achieved using only the attributes extracted from Twitter (and rank calculated by TwitterGrader), while those levels were only reached when we included the specific metrics that the other tools calculate (with Klout or PeerIndex).

Figure 4 shows a subtree of the model induced by REPTREE. We can say that only six attributes are used. The number of followers and followings are the most important characteristics, because they are repeatedly used from the root of the tree to the leaves. The other attributes that are rarely used are the *statuses_count*, the *listed_count*, the *friends_count* and the *geo* label. Taking this into account we can say that the calculation made by Twitter-Grader focuses on the relationships between users in a direct way. In particular, having more followers than people you follow and trying to have high-grade followers will raise your score. This can be difficult to achieve, and there is no guaranteed way to get influential people to follow you, though retweeting them and carrying on conversations through Twitter may help to increase your influence according to TwitterGrader.

How to simultaneously increase the influence considering the three tools. It has been previously asserted (see Subsection 2.2) that TwitterGrader shows a different behaviour than that of Klout and PeerIndex: the ranking values calculated by TwitterGrader are very high for almost every user (usually greater than 90). So, finding actions to increase the influence calculated by Klout and PeerIndex will be enough to increase the influence in all the tools at the same time.

When we combine the datasets that include specific attributes of Klout and PeerIndex and search for models that lead to high influence levels, we can detect some relevant attributes that concern the influence calculation in the three tools simultaneously: the *true_reach* (from Klout) and the *authority* (from PeerIndex) are the only attributes used in the rules that are associated with very influential users. Consequently, in order to reach high levels of influence, a user needs to have followers that follow him/her actively (respond to his/her messages, share them, etc.). It is not so important the brute number of followers (which includes spam, bots, and other inactive users), but rather the number of active followers (indirectly measured by *true_reach*). We have identified that medium influence levels can be achieved with more than 2 000 active followers, while more than 50 000 are needed if a user wants to be considered as highly influential. All the mentioned thresholds arise from the complete models that have been induced (they are available at [15]). Inside this second group of very influential users, those that have a higher *authority* reach more intense levels of influence.

Some other actions that are common to high influential users are related with the configuration of the personal profile. All the Twitter users with a high influence have unprotected accounts (allowing anyone to follow them) and have set a default profile image. In addition, major use of the URL field (giving direct access to their own webpages) is notable.

4 Conclusions and Future Work

In this work we have analyzed and compared three of the most popular Twitter influence metrics and tools namely Klout, PeerIndex and TwitterGrader. All of them summarize the influence of a Twitter user to a number that, in theory, should help us to measure how influential such a user is. However, in practice, this score is not really useful if the Twitterer does not know how to modify it, that is, what specific actions he could carry out to increase his influence depending on a tool.

In this sense, applying data mining techniques our study has arisen interesting conclusions such as: (1) the attributes used by existing analytical tools to measure the influence vary from one to another; (2) Klout and PeerIndex use a more sophisticated process to calculate the influence based fundamentally on their specific metrics and not on the network structure; (3) TwitterGrader applies a simpler mechanism to do that task based mainly on characteristics of his network topology; (4) detailing the personal profile (including description, url, profile image, etc.) can increase the influence; (5) keeping a constant level of participation on the social network or trying to have high-grade followers will also raise your score.

Furthermore, although the concrete algorithms used by the tools analyzed are kept secret, the classification and regression models that we have generated reveal the weight that each variable has in the final formula of the algorithm. So, the models inferred from our study can be directly translated into action rules (keeping a constant level of participation or detailing the personal profile). Besides, although some actions could suggest the same strategy, their cost can be different from one tool to another. For example, reaching the highest score in TwitterGrader by increasing the number of followers is easier than doing it in PeerIndex, since we just need around two thousand followers in the first case as opposed to the five hundred thousand that are required by the second tool.

In the future, we plan to design interactive models where the final user could guide the learning algorithm towards some specific action (removing the most expensive actions) also allowing a twitterer the possibility of discarding those recommendations that he/she considers unfeasible.

Acknowledgements

This work has been partially founded by the European Commission Project "SiSOB: An Observatorium for Science in Society based in Social Models" (http://sisob.lcc.uma.es) (Contract no.: FP7 266588), "Sistemas Inalámbricos de Gestión de Información Crítica" (with code number TIN2011-23795 and granted by the MEC, Spain) and "3DTUTOR: Sistema Interoperable de Asistencia y Tutoría Virtual e Inteligente 3D" (with code number IPT-2011-0889-900000 and granted by the MINECO, Spain)

References

[1] Twitter. [link].
URL http://twitter.com/

[2] A. Leavitt, E. Burchard, D. Fisher, S. Gilbert, The influentials: New approaches for analyzing influence on twitter, Webecology Project.

[3] M. Cha, H. Haddadi, F. Benevenuto, K. P. Gummadi, Measuring user influence in twitter: The million follower fallacy, in: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM), 2010.

[4] Klout. [link].
URL http://klout.com/

[5] Peer Index. [link].
URL http://www.peerindex.com/

[6] Twitter-Grader. [link].
URL http://twitter.grader.com/

[7] J. Weng, E.-P. Lim, J. Jiang, Q. He, Twitterrank: finding topic-sensitive influential twitterers, in: Proceedings of the third ACM international conference on Web search and data mining, WSDM '10, ACM, New York, NY, USA, 2010, pp. 261–270.

[8] E. Bakshy, J. M. Hofman, W. A. Mason, D. J. Watts, Everyone's an influencer: quantifying influence on twitter, in: Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11, ACM, New York, NY, USA, 2011, pp. 65–74.

[9] J. Lin, A. Kolcz, Large-scale machine learning at twitter, in: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD '12, ACM, New York, NY, USA, 2012, pp. 793–804.

[10] B. Liu, Web Data Mining, Springer, 2011.

[11] M. Kimura, K. Saito, R. Nakano, H. Motoda, Extracting influential nodes on a social network for information diffusion, Data Mining and Knowledge Discovery 20 (2010) 70–97.

[12] C. C. Aggarwal (Ed.), Social Network Data Analytics, Springer, 2011.

[13] I. Anger, C. Kittl, Measuring influence on Twitter, in: Proc. 11th International Conference on Knowledge Management and Knowledge Technologies, ACM, 2011, pp. 31:1–31:4.

[14] J. Serrano-Puche, Herramientas web para la medicin de la influencia digital: anlisis de klout y peerindex, El profesional de la informacin 21 (3) (2012) 298–303.

[15] ATIM. [link].
URL http://iaia.lcc.uma.es/~jcampo/software/ATIM2/

[16] I. Kononenko, M. Kukar, Machine learning and data mining: Introduction to principles and algorithms, Horwood Publishing, 2007.

[17] J. R. Quinlan, Learning with continuous classes, in: Proceedings of the Australian Joint Conference on Artificial Intelligence, 1992, pp. 343–348.

[18] Y. Wang, I. H. Witten, Induction of model trees for predicting continuous classes, in: Poster papers of the 9th European Conference on Machine Learning, Springer, 1997.

[19] J. R. Quinlan, C4.5: Programs for machine learning, Morgan Kaufmann, 1993.

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. Witten, The WEKA data mining software: an update, SIGKDD Explorations Newsletter 11 (1) (2009) 10–18.