A Computer System for Automatic Evaluation of Researchers' Performance

Ashkan Ebadi[1] and Andrea Schiffauerova[2]

[1] a_ebad@encs.concordia.ca, [2] andrea@ciise.concordia.ca
Concordia Institute for Information Systems Engineering (CIISE), Concordia University,
1515 Ste-Catherine Street West, Montreal, Quebec H3G 2W1 (Canada)

Abstract

The increasing number of researchers and the limited financial resources has caused a tight competition among scientists to secure research funding. On the other side, it has become even harder for funding allocation organizations to evaluate the performance of researchers and select the best candidates. However, it seems that the current evaluation methods are highly correlated with subjective criteria. In addition, the subjective nature of peer-review as one the most common methods in scientific evaluation calls itself for an accurate complementary quantitative method to help the decision makers. This paper proposes an automatic computer system, which is based on machine learning techniques for predicting the performance of researchers. The proposed system uses various features of different types as the input to a complex machine learning module to predict the performance of a researcher in a given year. The method provides the decision makers with fair comparative results regardless of any subjective criteria. Our results show the high accuracy of the proposed system in predicting the performance of researchers.

Conference Topic

Methods and techniques, Science policy and research assessment

Introduction

Research grants is known as one of the crucial drivers of scientific activities that can influence the size and efficiency of R&D sector and its productivity (Jacob & Lefgren, 2011). It can also affect the performance of researchers through providing them with a better access to the research resources (Lee & Bozeman, 2005). In the meantime, policies on R&D activities have evolved over the past fifty years (Elzinga & Jamison, 1995; Sanz-Menendez & Borras, 2000). Funding agencies put a lot of efforts on selecting the best candidates for allocating grants as well as on evaluating the performance of researchers in regards to the amount of funding that they have been receiving. On the other hand, the growing number of researchers worldwide has made the competition for securing the limited financial resources even harder. For example, according to Polster (2007) the contest for receiving research funding is on the rise in Canada especially among the academic researchers mainly due to the changes in federal funding policies, lack of university operating budgets, and increasing research costs. The researchers' demand for funding cannot be fully satisfied by the finite financial capacity of the funding agencies. However, the case could be even worse for the young researchers since the senior researchers are more known within their scientific community that might help them in getting money for research.

Peer review is the oldest measure that has been being used for evaluating researchers' performance and their proposals. Most of the funding agencies use a committee of independent researchers to review the researchers' proposals for funding and select the most appropriate researcher(s) through a competitive process. However, the peer review process has been widely criticized in the literature due to the potential biases since the accuracy of the procedure is highly dependent on the selected experts. For example, preferences of peers can affect the final decision or it can act as a gatekeeper for new research interests since peers may not come into an integrated conclusion (King, 1987). Despite the aforesaid drawbacks, the great advantage of peer review process is that the impact of the proposed research could be assessed quite easily and accurately (Allen et al., 2009). For this important reason it has still remained as one of the most popular techniques in scientific evaluation. Though, the current trend is to combine the expert review with quantitative performance indicators (Butler, 2005; Hicks et al., 2004) in order to achieve a more balanced evaluation since it cannot be reliable enough as a single indicator. For this purpose, citation and publication counts based indicators are commonly used as the quantitative indicators of researchers' performance.

One of the reasons that scientists publish their work in the form of scientific papers is that in this way they can secure their priority in discoveries (De Bellis, 2009). According to the review of literature done by Tan (1986), performance evaluation of individual researchers and research departments are in most cases based on publication counts measures (at least partially). For the quality of publications, citation counts based indicators, first introduced by Gross and Gross in 1927, are commonly accepted as a proxy for the impact of a scientific publication (Gingras, 1996). In general, they count the number of citations received by an article after the date it is published; hence, papers with higher number of citations are assumed to have higher impact.

Invention of the Internet and availability of the digital data have made it feasible to extract and collect data in a very large scale. In addition, the rapid advancement in the field of computer science has made new ideas and algorithms available to the data scientists. Therefore, large scale digital data and complex algorithms provide researchers with novel opportunities to explore new directions of the information science as well as scientific evaluation. This paper presents an integrated highly accurate automatic productivity prediction system that can assist decision makers (and peers) to detect the most appropriate researchers for funding allocation. The remainder of the paper proceeds as follows: *Data and Methodology* section describes the data gathering procedure in detail while explaining the methods and methodologies that were used; the *Results* section presents the performance evaluation results and interpretations for the proposed system; the paper concludes in *Discussion* section; and limitations and future research directions are stated in the last section of the paper.

Data and Methodology

Data

We decided to focus on performance of the researchers who have been funded by the Natural Sciences and Engineering Research Council (NSERC)[1] of Canada. The main reasons for choosing NSERC was its role as the main federal funding organization in Canada, and the fact that almost all the Canadian researchers in natural sciences and engineering receive at least a basic research grant from NSERC (Godin, 2003). Therefore, as the first stage information about the funded researchers was collected from NSERC[2]. In the next phase, Elsevier's Scopus[3] was used to gather all the information about the funded researchers. The data spans from information about the authors themselves (*e.g.* Scopus ID, their affiliation, number of publications in a given year, *etc.*) to their articles (*e.g.* year of publication, authors of the paper, keywords, *etc.*).

The time interval of the research was set to the period of 1996 to 2010 since the data coverage of Scopus was better after 1996. Moreover, to have a proxy of the quality of the papers we

[1] For more information, see: http://www.nserc-crsng.gc.ca/index_eng.asp

[2] Students were excluded from the data as the goal of the paper is evaluating the performance of researchers.

[3] Scopus is a commercial database of scientific articles that has been launched by Elsevier in 2004. It is now one of the main competitors of Thomson Reuter's Web of Science.

used SCImago[4] to collect the impact factor information of the journals in which the articles were published. SCImago was chosen for two main reasons. Firstly, it provides annual data of the journal impact factors that enables us to perform a more accurate analysis since we are considering the impact factor of the journal in the year that an article was published not its impact in the current year. Secondly, SCImago is powered by Scopus that makes it more compatible with our publications database.

In the next phase of data preparation, we calculated several bibliometric features such as amount of funding received by a researcher in a given year, his/her career age, average number of co-authors, average number of publications, average number of citations, *etc.* In addition, using Pajek[5] software social network analysis techniques were employed to construct the collaboration networks of the researchers within the examined time interval. The created networks were used to calculate various network structure properties (*e.g.* betweenness centrality, eigenvector centrality, and clustering coefficient) of the researchers at the individual level. All the calculated features were integrated in a MySQL[6] dataset. The final database contains 117,942 records of researchers. In the next section, methodologies are discussed in more detail.

Methodology

Several features of various types and from different sources were selected for this study. Funding is acknowledged in the literature as one of the main drivers of scientific activities where a three-year (*e.g.* Payne & Siow, 2003) or a five-year (*e.g.* Jacob & Lefgren, 2007) time window is mostly considered for the funding to take effect. In this paper a three-year time window was considered for all the bibliometric variables, *e.g.* for assessing the productivity of a given researcher in year 1999 his/her amount of funding was summed up for the period of 1996 to 1998 (*sumFund3*). Intuitively, productive researchers are expected to at least maintain their performance level. Various past productivity features were hence included in the model reflecting the quality and quantity of the publications. As a proxy for the rate of publications, number of publications in a three-year time window (*noArt3*) was considered. Two indicators were used as proxies for the quality of publications, *i.e.* average number of citations in a three year time window (*avgCit3*) and the average impact factor of the journals in which the articles were published in a three year time interval (*avgIf3*). Both of the mentioned features can serve as a proxy for quality, but with a slightly different meaning. Impact factor indicates the respectability of the journal, *i.e.* the quality and the level of contribution perceived by the authors and the reviewers of the paper, whereas citation counts show the impact of the article on the scientific community and on the subsequent research.

A multi-level feature representing the scientific field of the researcher (*discip*) was also used in the model since publication and citation habits can be different in various scientific fields. For example, citing habits and the rate of citations may vary across different scientific fields in a way that in some scientific fields authors publish articles more frequently or the published papers contain more references (MacRoberts & MacRoberts, 1996; Phelan, 1999). It is argued in the literature that older researchers in general can be more productive (Merton, 1973; Kyvik & Olsen, 2008) due to several reasons (*e.g.* better access to the funding and expertise sources, more established collaboration network, better access to modern equipments). Hence, the career age of the researcher (*careerAge*) was included in the model representing the time difference between the date of his/her first article in the database and the given year. As a common indicator of the scientific collaboration, the average number of co-authors per paper was also included in the prediction model (*teamSize*). It is expected that

---

[4] For more information, see: http://www.scimagojr.com

[5] Social network analysis software, for more information see: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

[6] Open source relational database management system, for more information see: http://www.mysql.com/

researchers who have on average higher number of co-authors have more connections that might result in relatively higher number of projects or future publications, hence this feature was also considered as one of the influencing factors.

As discussed in the previous section, social network analysis was used to construct the collaboration networks and to measure the structural network properties of researchers. In particular, four network structure indicators were calculated namely betweenness centrality (*bc*), clustering coefficient (*cc*), eigenvector centrality (*ec*), and degree centrality (*dc*). Betweenness Centrality (*bc*) is an indicator of the important players (researchers) in a network who have a control over the flow of knowledge and resources. These players, who are also called as *gatekeepers*, are able to bridge different communities. Theoretically, betweenness centrality of the node *k* is measured based on the share of times that a node *i* reaches a node *j* via the shortest path passing from node *k* (Borgatti, 2005) and is calculated as follows (σ_{ij} is the total number of shortest paths from node *i* to j and $\sigma_{ij}(k)$ is the number of shortest paths from node *i* to node *j* that contains node *k*):

$$bc_k = \sum_{i \neq k \neq j} \frac{\sigma_{ij}(k)}{\sigma_{ij}} \qquad (1)$$

Clustering Coefficient (*cc*), also called *cliquishness*, indicates the tendency of researchers to cluster with other researchers in the network. Hence, researchers with high clustering coefficient may have a relatively high number of connections with the other team members who are collaborating in a tightly knit group. Therefore, this indicator was selected to represent the tight collaboration impact on the overall performance of the team. Theoretically, clustering coefficient of node *i* (cc_i) is defined based on the number of triangles (inter-connected sub-network of three nodes) that contains the node *i* (t_i) normalized by the maximum number of triangles in the given network (Watts & Strogatz, 1998). Let n_i denotes number of neighbors of the node *i*, hence:

$$cc_i = \frac{2t_i}{n_i(n_i - 1)} \qquad (2)$$

Degree Centrality (*dc*) that was also considered as one of the network variables is defined based on the number of ties that a node has (degree) in an undirected graph. Hence, researchers with high degree centrality should be more active since they have higher number of ties (links) to other researchers (Wasserman, 1994). Moreover, in co-authorship networks it can be regarded as the number of direct partners or team members of a given researcher. Hence, it is expected to have an influence on the scientific activities. Degree centrality for node *i* (dc_i) is thus defined based on the node's degree (deg_i) and then the values are normalized between 0 and 1 (dividing by the highest degree in the network) to be able to compare the centralities:

$$dc_i = \frac{deg_i}{deg_{high}} \qquad (3)$$

Eigenvector Centrality (*ec*) takes the importance of a node and its connections into the account. Hence, a researcher has high eigenvector centrality if he/she is connected with other important actors who are themselves occupying central positions in the network. These researchers can be identified as *leaders* in the scientific networks since they are connected with too many other influential and highly central researchers, and it is hence expected that they shape the collaborations and play an important role in setting priorities in scientific projects that might affect the performance of researchers. A complete list of the selected features is shown in Table 1.

Table 1. List of attributes for the prediction models.[7]

No	*Attribute*
1	Scientific area in which the researcher is working (*discip*)
2	Total amount of funding received by each researcher in a 3 year time window (*sumFund3*)
3	Total number of publications of each researcher in a 3 year time window (*noArt3*)
4	Average number of citations received by researcher's articles in a 3 year time window (*avgCit3*)
5	Average impact factor of the journals in which researcher's articles were published in a 3 year time window (*avgIf3*)
6	Average betweenness centrality of each researcher in a 3 year time window (*btwn3*)
7	Average degree centrality for each researcher in a 3 year time window (*deg3*)
8	Average clustering coefficient of each researcher in a 3 year time window (*clust3*)
9	Average eigenvector centrality of each researcher in a 3 year time window (*eigen3*)
10	Average number of authors per paper for each researcher (*teamSize*)
11	Career age of the researcher (*careerAge*)

The mentioned features were used as an input to the prediction model. Figure 1 shows the whole process of the researchers' performance prediction. Number of publications was considered as the target variable for the performance prediction task. As it can be seen, data is first preprocessed and cleaned. For this purpose, several JAVA programs were coded to check the data for redundancy, out of range values, impossible combinations, errors, and missing values and then data was filtered based on the records that contained all the required data. The resulted data containing all the mentioned features was fed into the data preparation block where at first all the features were normalized to a value between 0 and 1. This was a crucial step since the features were of different units and scales. Local Outlier Factor (LOF) algorithm was then implemented to detect the outliers. LOF that was proposed by Breunig et al. (2000) is based on the local density concept in which the local deviation of a given data is measured with respect to its *k* nearest neighbors. A given data is outlier if it has a substantial different density from its *k* neighbors. The final step of the data preparation step was optimizing the attributes' weights. For this purpose we used an evolutionary attributes weights optimizer that employed genetic algorithm to calculate the weights of the attributes. The weighting procedure improved the accuracy of the system by giving more value to the most influential attributes. The resulted data was integrated into a single data repository named as the target data.

[7] The initial list of the selected features was prepared as a result of an intensive statistical analyses performed on the target data. The list was then refined and weighted within the proposed system.

After making the data ready for the analysis, a stratified 10-fold cross validation design was used for the model validation. Cross validation is an analytics tool that is used to design and develop fine tune models. In other words, the data is split into two disjoint sets where one part is used for training and fitting a model (training set) while the other part is employed for estimating the error of the model (test set) (Weiss & Kulikowski, 1991). We used a nested 10-fold cross validation in which the data is split into 10 disjoint subsets in a way that union of the 10 folds results the original data. The method runs 10 times and in each time one fold is considered as the test data while the rest are regarded as the training data.

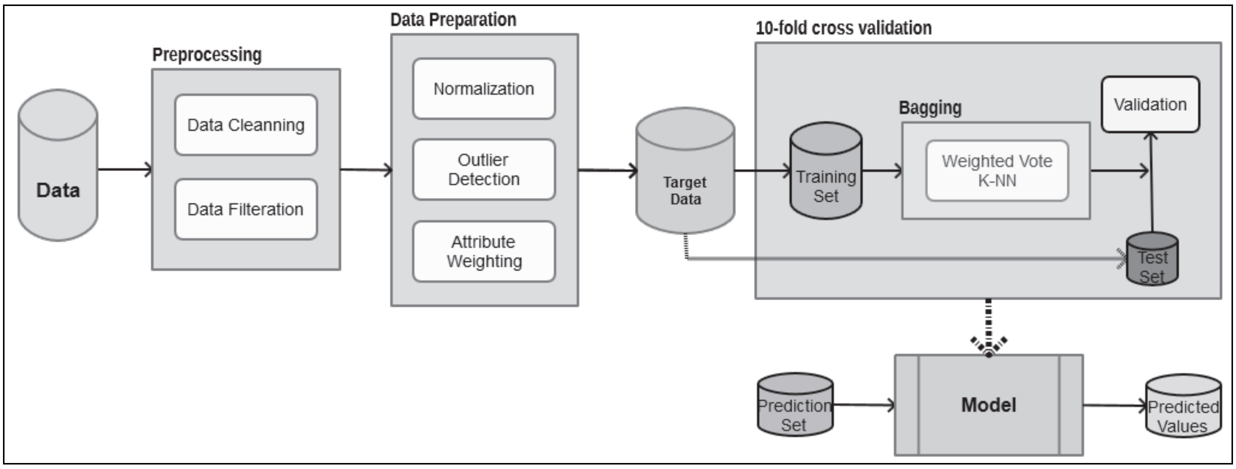

Figure 1. Proposed model for automatic evaluation of researchers' performance.

As mentioned earlier, number of publications was considered as the target variable. To further improve the accuracy of the prediction the ensemble meta-algorithm was employed. For this purpose, bootstrap aggregating (bagging) approach was used. Bagging is an ensemble method that makes random subsets of the data and trains them separately where the final result is obtained by averaging over the results of the separated models (Breiman, 1996). Bagging is a nested module in which we used weighted vote 10-Nearest Neighbor (10-NN) algorithm to train the data and to create the model. In weighted vote 10-NN the distance of the neighbors to the given data is considered as a weight in the prediction in a way that neighbors that are closer to the given data get higher weights. This particularly helped to increase the accuracy of the prediction. Data in the range of 1996 to 2009 was used to train and build the model while a separate disjoint data for 2010 (prediction set) was used for testing the accuracy of the prediction model. The final output of the proposed automatic computer system was the predicted number of publications for the researchers in the prediction set.

Results

In this section the results of the performance evaluation of the proposed automatic computer system (PACS) is presented. As discussed earlier, the model was trained on the data from 1996 to 2009 and a disjoint dataset for 2010 was used for the prediction and the accuracy tests. The accuracy of the proposed model was compared with several well-known machine learning algorithms, however, in this paper the results are presented and compared for the PACS model as well as two other algorithms that showed the highest accuracy in predicting the target variable.

Figure 2 shows the prediction errors of PACS, linear regression, and polynomial regression of degree three[8]. We considered three error measures for comparing the performance of the

[8] Other algorithms (*e.g.* decision trees) were also tested but these listed algorithms were the top two ones with the highest accuracy.

mentioned algorithms. Root mean squared error is one of the main measures for comparing the accuracy of the prediction models and is defined as the square root of the average of the squares of errors. According to Figure 2, PACS is predicating the number of publications of researchers with 1.451 average deviation between the predicted value and the real number of publications. Normalized absolute error is the absolute error (difference between the predicted value and the real value) divided by the error made if the average would have been predicted. The root relative squared error takes the average of the actual values as a simple predictor to calculate the total squared error. The result is then normalized by dividing it by the total squared error of the simple predictor and square root is taken to transform it to the same dimension as the predicted value. As it can be seen PACS is performing better in all the three measures where the degree 3 polynomial fit is the worst.

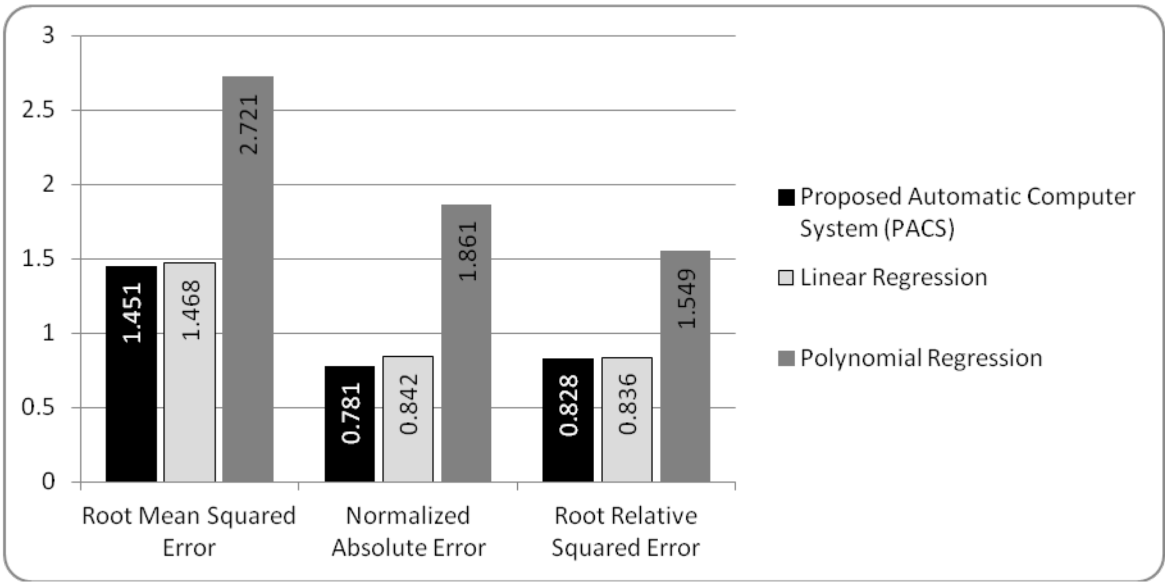

Figure 2. Accuracy test, PACS vs. other two top performing algorithms.

Table 2. Prediction results.

No	*Predicted no of articles*	*noArt*	*sum Fund3*	*avg If3*	*avg Cit3*	*teamSize*	*btwn3*	*clust3*	*deg3*	*eigen3*	*careerAge*	*discip*	*noArt3*
1	0.361	0	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.737	2	0
2	1.102	0	0.013	0.279	0.028	0.000	0.000	1.000	0.005	0.000	0.632	3	1
3	3.865	7	0.044	0.054	0.005	0.001	0.059	0.125	0.027	0.000	0.737	1	13
4	1.103	0	0.010	0.068	0.083	0.000	0.000	1.000	0.007	0.000	0.737	3	1
5	1.206	1	0.072	0.132	0.020	0.002	0.016	0.409	0.020	0.000	0.526	0	6
6	6.703	4	0.167	0.246	0.080	0.002	0.055	0.158	0.039	0.000	0.737	1	26
7	1.030	4	0.032	0.115	0.017	0.001	0.018	0.455	0.018	0.000	0.737	0	6
8	4.120	3	0.061	0.136	0.041	0.002	0.185	0.109	0.134	0.000	0.737	1	15
9	0.000	0	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.263	0	0
10	5.047	3	0.137	0.141	0.041	0.001	0.133	0.163	0.050	0.000	0.684	0	15
11	1.128	1	0.010	0.091	0.062	0.003	0.003	0.333	0.007	0.000	0.526	1	1
12	1.964	1	0.010	0.113	0.009	0.004	0.053	0.192	0.022	0.018	0.737	1	7
13	12.228	7	0.095	0.399	0.028	0.010	0.197	0.042	0.075	0.000	0.684	0	31
14	2.112	2	0.190	0.228	0.091	0.001	0.011	0.182	0.020	0.000	0.737	1	6
15	2.233	3	0.299	0.230	0.051	0.002	0.013	0.457	0.035	0.000	0.737	0	7
16	3.577	4	0.198	0.259	0.055	0.002	0.042	0.145	0.059	0.000	0.579	4	12
17	11.308	9	0.329	0.309	0.116	0.002	1.000	0.062	0.148	0.000	0.737	1	40
18	4.841	4	0.093	0.458	0.051	0.001	0.027	0.117	0.037	0.000	0.737	0	19
19	5.752	4	0.116	0.253	0.055	0.123	0.003	0.823	0.940	1.000	0.737	1	20
20	7.421	8	0.193	0.270	0.077	0.002	0.153	0.079	0.082	0.000	0.737	1	26

A randomly selected sample of the predictions is presented in Table 2. Each row represents a distinct researcher's profile in 2010 for whom several indicators have been calculated and used in the PACS model as the input features. The real number of articles is shown in *noArt* column that was not fed into the prediction model. Based on the other attributes the proposed system automatically predicted the number of publications of a researcher in 2010, *i.e.* column named *Predicted no of articles* in Table 2 and is highlighted in dark grey. As it can be seen using several features of different types and employing various techniques for data gathering (*e.g.* bibliometrics, social network analysis) and preparation provides the system with highly accurate high-dimensional input data that led to a low error rate and good predictions. Interestingly, it seems that the system successfully considered the differences between various scientific fields in performing scientific activities. According to the results, although the profile of the researchers numbered 1 and 9 in Table 2 are relatively similar, the predicted performance differs as they do not belong to the same scientific field. Hence, the results confirm the importance of the scientific disciplines in predicting the performance of researchers. In addition, comparison of the researchers numbered 6 and 7 highlights the importance of the past productivity as well as the quality of publications in predicting the number of publications.

Discussion

In this paper we used various bibliometric as well as network structural property features to build a model to predict the performance of researchers. Machine learning techniques and availability of the digital data has made it possible to use complex algorithms on high dimensional large scale data. This provides scientometrists with an opportunity to go beyond the current border of using common indicators or simple statistical analyses. Although some researchers recently worked on citation prediction using machine learning algorithms (*e.g.* Fu & Aliferis, 2010; Lokker et al., 2008) to our knowledge this is the first study that focused on the prediction of researchers' productivity using input features of different types and at the individual level of the researchers.

The attribute weighting method to rank features based on their importance that was implemented in the proposed model as well as the outlier detection module for data filtration increased the accuracy of the predictions significantly. Results of the attribute weighting module can also shed light on the most influential attributes in predicting the scientific activities of the target researchers. Another unique approach that was employed in designing the proposed system was using several features of similar nature in building the model that reinforced the prediction power of the system. For example, average number of citations and average impact factor of the journals were used to represent the quality of the paper. Another example is the degree centrality and scientific team size as the former represents the number of direct connections of a researcher while the latter indicates the average number of his/her co-authors. These attributes of similar nature surely empowered the accuracy of the model by providing it with more dimension and flexibility.

To conclude, as it was observed complex computer algorithms can be used to design automatic evaluation systems and prediction tools to evaluate different aspects of scientific activities of researchers. It is obvious that peer reviewing cannot be completely replaced by such tools. However, such systems can help decision makers in setting both long-run and short-term strategies in regard to the funding allocation and/or analyzing researchers' productivity. In addition, the availability of high-dimensional large scale data (in our case, a large dataset spanning from 1996 to 2010) that is intensively cleaned and preprocessed for learning the model will surely contribute to highly accurate predictions that are not based on a limited criteria or a limited feature set. Therefore, this can also help to establish a fairer funding allocation or scientific evaluation system.

Limitations and Future Work

We were exposed to some limitations in this paper. Firstly, Scopus was selected for gathering information about the funded researchers' articles. Since Scopus and other similar databases are English biased, hence, non-English articles are underrepresented (Okubo, 1997). Secondly, due to the better coverage of Scopus before 1996, the time interval of 1996 to 2010 was selected for the analysis. Although Scopus is confirmed in the literature to have a good coverage of articles, as a future work it would be recommended to focus on other similar databases to compare the results.

Furthermore, we were exposed to some limitations in measuring scientific collaboration among the researchers where we used the network structure properties. In particular, we were unable to capture other links that might exist among the researchers like informal relationships since these types of connections are never recorded and thus cannot be quantified. In addition, there are also some drawbacks in using co-authorship as an indicator of scientific collaboration since collaboration does not necessarily result in a joint article (Tijssen, 2004). An example could be the case when two scientists cooperate together on a research project and then decide to publish their results separately (Katz & Martin, 1997). For assessing the quality of the papers based on citation counts we did not account for self citations, negative citations, or special inter-citation patterns among a number of researchers. Although we also used another proxy (average impact factor of journals) to overcome this limitation, it can be addressed in the future works.

References

Allen, L., Jones, C., Dolby, K., Lynn, D., & Walport, M. (2009). Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLoS One*, *4*(6), e5910.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55-71.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123-140.

Breunig, M. M., Kriegel, H., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. *ACM Sigmod Record*, *29*(2), pp. 93-104.

Butler, L. (2005). What happens when funding is linked to publication counts? Handbook of quantitative science and technology research (pp. 389-405), Springer.

De Bellis, N. (2009). *Bibliometrics and citation analysis: From the science citation index to cybermetrics*. Scarecrow Press.

Elzinga, A. & Jamison, A. (1995). *Changing policy agendas in science and technology. Handbook of Science and Technology Studies* Ed.by Sheila Jasanoff et al. London: Sage.

Fu, L. D. & Aliferis, C. F. (2010). Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. *Scientometrics*, *85*(1), 257-270.

Gingras, Y. (1996). Bibliometric analysis of funded research. A feasibility study. *Report to the Program Evaluation Committee of NSERC.*

Godin, B. (2003). The impact of research grants on the productivity and quality of scientific research. No. 2003. *INRS Working Paper.*

Gross, P., & Gross, E. (1927). College libraries and chemical education *Science, 66*(1713), 385-389.

Hicks, D., Tomizawa, H., Saitoh, Y., & Kobayashi, S. (2004). Bibliometric techniques in the evaluation of federally funded research in the United States. *Research Evaluation*, *13*(2), 76-86.

Jacob, B. A. & Lefgren, L. (2011). The impact of research grant funding on scientific productivity. *Journal of Public Economics*, *95*(9), 1168-1177.

Katz, J. S. & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1-18.

King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science*, *13*(5), 261-276.

Kyvik, S. & Olsen, T. B. (2008). Does the aging of tenured academic staff affect the research performance of universities? *Scientometrics*, *76*(3), 439-455.

Lee, S. & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673-702.

Lokker, C., McKibbon, K. A., McKinlay, R. J., Wilczynski, N. L., & Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: Retrospective cohort study. *BMJ* (Clinical Research Ed.), *336*(7645), 655-657.

MacRoberts, M. H. & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics*, *36*(3), 435-444.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press.

Okubo, Y. (1997). Bibliometric indicators and analysis of research systems: Methods and examples. No. 1997/1. *OECD Publishing*.

Payne, A. A. & Siow, A. (2003). Does federal research funding increase university research output? *Advances in Economic Analysis & Policy*, *3*(1).

Phelan, T. (1999). A compendium of issues for citation analysis. *Scientometrics*, *45*(1), 117-136.

Polster, C. (2007). The nature and implications of the growing importance of research grants to Canadian universities and academics. *Higher Education*, *53*(5), 599-622.

Sanz Menéndez, L., & Borrás, S. (2000). Explaining changes and continuity in EU technology policy: The politics of ideas.

Tan, D. L. (1986). The assessment of quality in higher education: A critical review of the literature and research. *Research in Higher Education*, *24*(3), 223-265.

Tijssen, R. J. (2004). Is the commercialisation of scientific research affecting the production of public knowledge? Global trends in the output of corporate research articles. *Research Policy*, *33*(5), 709-733.

Wasserman, S. (1994). Social network analysis: Methods and applications. *Cambridge university press*.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440-442.

Weiss, S., & Kulikowski, C. (1991). *Computer Systems that Learn*: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. San Francisco: Morgan Kaufmann.