ERS: Evaluating Reputations of Scientific Journals

Émilie Samuel, Colin de la Higuera

ECAI 2008
M. Ghallab et al. (Eds.)
IOS Press, 2008

doi:10.3233/978-1-58603-891-5-817

ERS: Evaluating Reputations of Scientific Journals

Émilie Samuel and **Colin de la Higuera**[1]

Abstract. Current methods for evaluating research are based on counting the number of citations received for publications. Thus, the more an article is cited and the more its impact is considered as important. In this article, we propose a new method for assessing the reputation of scientific journals, based on a Web application in which are gathered the votes of expert researchers. The voting results indicate degrees of preference for one journal over another. Our system uses, in addition, the publications of an expert in order to quantify his expertise in specific fields. These values are coupled with those of votes to determine the relevance, according to the field, of each journal in each topic. An iterative process of transferring values given to journals by experts to values of the experts themselves given their publications has been implemented.

1 Key concepts

The system *ERS* manages bibliographic data formed by journals, researchers and themes. The journals and themes are bound by a relation called *relevance*. Each journal publishes articles more or less relevant to some research topic. Journals and researchers are directly connected by publications (relation *has published*). From these we hope to measure the *expertise* of each researcher in relationship with each theme. Finally, the votes of the researchers, depending on the expertise of the latter, will influence the relevance of the themes for journals. The phenomenon is recursively cyclic: the relevance influences the expertises, which in turn affect the relevances through the votes. The influence of researchers grows with their *global confidence*.

We summarise these relationships through the diagram represented in Figure 1. Entities **Experts**, **Themes** and **Journals** are connected by relations *relevance*, *expertise*, *has published* and *vote*. Associated attributes are not represented.

1.1 Relevance of a journal for a theme

Each journal is more or less relevant to each theme. The relevance reflects the reputation for this topic, of the articles published by the journal. Its value can be interpreted as the probability that the community of researchers in the field advise someone searching for literature in the given area, this journal.

1.2 Global confidence and expertise of a voter in a topic

The computation of the expertise in a theme depends directly on the relevance of the journals in which the expert has published for that theme. If, for example, the expert has published several times in journals recognised by the system itself as being relevant to the theme

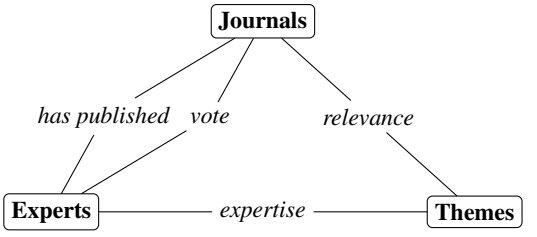

Figure 1. The general model

databases, then the expert will be deemed to be an expert in this area.

Thus, the calculation of expertise in each subject, which is based on the journals in which the author has published, depends, for each of them:

- in the number of publications;
- in the sum of all relevances of the journal, which reflects its importance;
- in the likelihood of the theme, given the journal;
- in the belief in the relevance of the theme for this journal.

However, comparisons between different researchers should be avoided. One can, for example, consider that a group of individuals with similar profiles have interests for similar research fields. By contrast, a researcher with expertise of 10 % in *information retrieval* can not be considered twice as recognised in this area as a researcher with expertise of 5%. It may, in fact, be the case that the publications of one are less diversified than the second, which would then generate higher expertise, but in fewer topics.

1.3 Interrogating the experts

For each expert, a list of journals to be evaluated automatically is defined. This list consists of journals in which he has published, and of journals that are judged by the system, close to his expertise. This list can also include journals in which his co-authors have published or journals on which the system has little information.

The method of *paired comparisons* is used, whose application to ranking has been addressed since [2]. This method is intended to indicate a degree of preference, and lets one get a partial order by comparing journals two by two. It is then possible, from several partial orders resulting from expert opinions, to establish a total order of all the journals in each theme. Our approach is related to that shown in [3], where the authors propose to build clusters of total orders, corresponding to the opinions collected about movies.

The expert must answer questions such as " If you were to choose an article by one of these two journals, which would you choose?". We call this process between two journals a *match*. A series of matches (until interruption by the expert) is organised, each match

[1] Laboratoire Hubert Curien - UMR CNRS 5516 - 18 rue Benoît Lauras, 42000 Saint-Étienne - email: {emilie.samuel, cdlh}@univ-st-etienne.fr

being randomly drawn, where the journals in which the expert has published have higher probability to appear.

The results of the matches are then analysed following the methodology employed by the *Elo* classification, used to rank chessplayers [1]. This classification assigns each player a rating based on his performance in competition. The rating of a player evolves over time with his results. When two players meet, a predicted result for each is calculated, the highest ranked player being supposed to beat his weaker opponent. The greater the difference in rating between the two players and the higher the probability that the best player wins. Following the match between the two players, their ratings are updated according to the following principle: if a player has achieved a better result than expected, it means that he was underestimated, and his rating is therefore increased, and vice versa. A rating can therefore rise or diminish, and the adjustment takes place proportionally with the difference between the true outcome and the presumed outcome.

2 Operational aspects

A *beta* version of the system *ERS* has been running since July 2007[2]. Its ergonomics and aesthetics are subject to change. We are seeking a more attractive, user-friendly and interactive platform while retaining its ease of use. Initially we used a limited list of 16 themes, to which was added one smaller theme (*grammatical inference*) for testing purposes.

The operationalisation required an initialisation phase, each journal being allocated an initial relevance in each theme. To do this, we chose an initial set of themes, and associated with each theme a list of keywords. For example, can be associated to the theme *machine learning* words like *pattern recognition*, or *classification*, or *reinforcement*. We then computed a frequency (*term frequency*) for each keyword appearing in the titles of journal articles. Thus, the more a journal publishes articles with these words in their titles and the more its relevance to the corresponding theme increases. The confidence in the relevance matches, is obtained as a computation of the *inverse document frequency*, which is a function increasing with the specificity of the keywords.

3 Convergence of the system

The update of the system is done daily, in a batch mode. Thus, the data of the system constituted by relevance, global confidence and expertise are changing continuously, and are recomputed iteratively. The convergence of these values occurs as soon as relevance remains stable from one iteration to the next.

The first phase of the experimental validation of the convergence of relevances consisted in the initialisation (random) and normalisation of the relevances for 570 journals and for 17 themes. In order to constitute a panel of experts, 2000 researchers were then randomly selected from those identified in DBLP. Their global confidence and expertise in each subject were computed according to their publications in journals. Thereafter, a simulation of votes by these experts took place. This consisted of generating randomly 28500 votes, so as to reach an average of 50 per journal. The algorithm was finally run on this repeatedly until convergence of the values of relevance, global confidence and expertise.

The convergence results of three experiments respecting this protocol are shown in Figure 2. The variation distance L1 was used to measure the value of the difference between the relevance of an iteration to the next. As can be seen, the computation converges in a small number of iterations, each carried out in an average of 2 seconds.

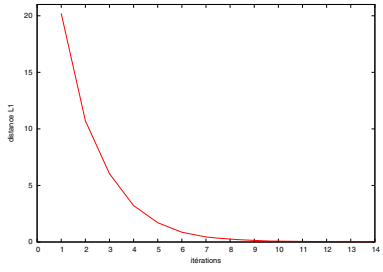

Figure 2. Convergence of the relevances during the batch computations

4 Conclusion and perspectives

System *ERS* permits a different evaluation of scientific journals, directly based on the opinions of scholars. This system, which we hope to render attractive, simple and efficient, offers an assessment protocol for comparing journals two by two. Following the processing of votes, the results indicate, according to the subject, which are the journals best recognised by the community of the area.

A number of perspectives are being looked into. In addition to those cited in this article, the first is to work on the actually very reduced list of themes: ideally the list should be dynamic: new communities or sub-communities should be detected by the system, and the corresponding keywords should be automatically computed. The computation of the expertise and confidence of the researchers could involve a more complex analysis, taking into account (again in an automatic way) the date on which his articles were published, or other information beyond DBLP and obtained by *Web mining* techniques. The interrogation scenario should also be considered as being improvable. Using better the results is another possible task: a profile for a journal (as a vector of quantities over themes) can be easily computed, and a similar profile can be computed for a researcher. One can therefore query the system with questions like "which journal is the closest to my way of doing research?".

In addition, the identification of researchers at the registration remains an important point on which further work is necessary. Finally, the evaluation of conferences is a logical evolution of the system, which requires additional attention and so is the even more ambitious task of adapting the system to other fields of research.

Acknowledgements

System *ERS* was developed in the laboratory Hubert Curien with the help of Fabrice Muhlenbach, Baptiste Jeudy and François Jacquenet. The expertise of Thierry Murgue has solved many engineering problems. Students from the department of computer science at Saint-Etienne have and continue to develop different modules for the system.

REFERENCES

[1] A. E. Elo, *The rating of chessplayers, past and present*, Arco, 1978.
[2] H. Joe, 'Rating systems based on paired comparison models', *Statistics & Probability Letters*, **11**, 343–347, (1991).
[3] A. Ukkonen and H. Mannila, 'Finding outlying items in sets of partial rankings', in *Knowledge Discovery in Databases: PKDD 2007*, volume 4702 of *LNCS*, pp. 265–276. Springer, (2007).

[2] http://labh-curien.univ-st-etienne.fr/ERS/