Toward the more effective identification of journals with anomalous self-citation

Tian Yu[1]*, **Guang Yu**[2], **Yan Song**[1] and **Ming-Yang Wang**[3]
[1]School of Economics and Management,
Harbin Engineering University,
145 Nantong Street, Nan Gang District, Harbin, 150001, CHINA
[2]School of Management, Harbin Institute of Technology, Harbin, 150001, CHINA
[3]College of Information and Computer Engineering,
Northeast Forestry University, Harbin, 150040, CHINA.
e-mail: *yutian@hrbeu.edu.cn; yug@hit.edu.cn; songyan@hrbeu.edu.cn;
wangmingyang@nefu.edu.cn.

ABSTRACT

Because of its important evaluative function, journal impact factors began to be manipulated by anomalous self-citations. To deal with this scientific misconduct and its undesirable influences, in this paper, an automatic classification model for journals with anomalous self-citation was constructed based on previous research. First, a training journal set and three test journal sets of normal journals and abnormal journals were established and four features were selected from a feature set. Then, a classification model was learnt using the Deep Belief Network (DBN) method, which was successfully able to identify abnormal journals in the data sets. Third, Logistic Regression and Support Vector Machine were employed to learn the classification models, the classification performances for which were then compared with the DBN model. Finally, 1138 journals in twelve subject areas from the journal Citation Report (JCR) in 2014 were chosen as empirical journal samples for the DBN model, from which 6.9 percent of empirical journals were identified as suspect journals with anomalous self-citation.

Keywords: Scientific journal; Journal Impact Factor; Anomalous self-citation; Classification model; Journal manipulation.

INTRODUCTION

Since Garfield first proposed and developed the impact factor concept in 1955 (Garfield 1955; Garfield and Sher 1963), it has had an increasingly important role in journal evaluation as it gives the average citation ratio for each article published by the journal. Despite the shortcomings (Vanclay 2012), the impact factor has become an important criterion for judging the quality of scientific publications over the years, and has influenced the evaluation of institutions and individual researchers worldwide.

https://doi.org/10.22452/mjlis.vol23no2.2

Because of its importance in the journal evaluation system, the impact factor has been favoured by journal publishers, editors, rating agencies and researchers, and governments have ever used it to rank universities and research institutions, who in turn, use it to assess researchers for employment, promotion and grant approvals; therefore, researchers also seek to submit manuscripts to high impact factor journals (Simons 2008). Since the evaluation index has become an important academic research measure, the impact factor now has significant power and influence (Falagas and Alexiou 2008; Arnold 2009), which has put pressure on the journal publishers and editors whose main aim is to increase the annual impact factor. Generally, improving a journal's impact factor takes several years; however, it can be quickly improved through covert manipulation (Simons 2008). The impact factor is defined as: citation counts in one year to a journal's contents in the two previous years, divided by the number of citable items in that journal in the two previous years. Due to the operability of its defining equation, practices attempt to either increase the nominator or decrease the denominator in the impact factor to tamper with the impact factor (Martin 2016). Journal manipulation methods include: (a) anomalous self-citation, also known as coercive self-citation, namely requiring manuscripts to cite at least one paper published in their own journal to increase the cumulative citation count (Arnold 2009; Falagas and Alexiou 2008; Mavrogenis et al. 2010; Wilhite and Fong 2012; Opthof 2013); (b) limiting the total number of research articles and increasing the number of reviews that are more likely to be cited (Hemmingsson et al. 2002a; Kurmis 2003); and (c) rising publication delays (Hemmingsson et al. 2002b; Tort et al. 2012; Heneberg 2013). Anomalous self-citation is scientific misbehavior that not only contradicts the normal rules of scientific progress and seriously undermines journal the integrity and impartiality, but also means that the citation relationship between the journals is false. Clearly, this is neither fair nor ethical.

As anomalous self-citation is negatively affecting academic research, it is necessary to find a way to monitor such misbehavior. Researchers focus attention on two research directions, one is to put forward new alternative evaluation indicators to measure the quality of journals (Rijcke and Rushforth 2015; Larivie`re et al. 2016; Chorus and Waltman 2016; Maity and Hatua 2016; Yu and Yu 2016), the other is to identify abnormal journals with anomalous self-citation from the journal sets by machine learning algorithm.

As well as constantly detecting and adjusting the list of journals covered by Journal Citation Report (JCR) to identify abnormal journals, Clarivate Analytics is constantly monitoring six indicators: total citations, journal impact factor, rank in category, percentage of journal self-citations in the Journal Impact Factor numerator, proportional increase in Journal Impact Factor with/without journal self-citations and the effect of journal self-citations on rank in the category by Journal Impact Factor (refer to http://wokinfo.com/media/pdf/jcr-suppression.pdf.). These indicators have been reasonably effective in identifying abnormal journals with high self-citations; however, previous research has found that the editors of some abnormal journals appear to target specific manuscripts and authors (Wilhite and Fong 2012).

There have been some attempts to identify anomalous self-citation in journals. Yu et al.

(2011) constructed a journal classification model using the k-nearest neighbor (KNN) method to automatically identify abnormal journals with impact factor manipulation. However, due to the unbalanced number of normal journals and abnormal journals with anomalous self-citation in JCR database, the accuracy of the classification model needs to be further improved. Subsequently, Yu et al. (2014) established a journal classification model using logistic regression method to improve the classification accuracy, but the precision of the model needed to be further improved. Through the use of surface learning methods, the classification performance of the obtained models was unsatisfactory. Therefore, in this paper, a more effective method is developed to detect undesirable behavior and ensure that the actual performance of scientific journals is honestly presented.

In this paper, a Deep Belief Network (DBN) method is used to learn the classification model so as to more effectively identify those journals with anomalous self-citation. The following section explains the methodology used in this research. In the other section, a classification model is constructed using the DBN method, the performance for which is then compared with a Logistic Regression (LR) model and a Support Vector Machine (SVM) model. To identify suspect journals, the empirical sample for the DBN model involved 1138 journals in twelve subject areas from the 2014 JCR.

METHOD

The goal of this research was to develop a classification model that could effectively identify abnormal journals with anomalous self-citation. First, a journal set was established that included both normal journals and abnormal ones after which the differences between them were evaluated to select relevant features, which then formed the basis for the development of the journal classification model. The experimental steps are shown in Figure 1.

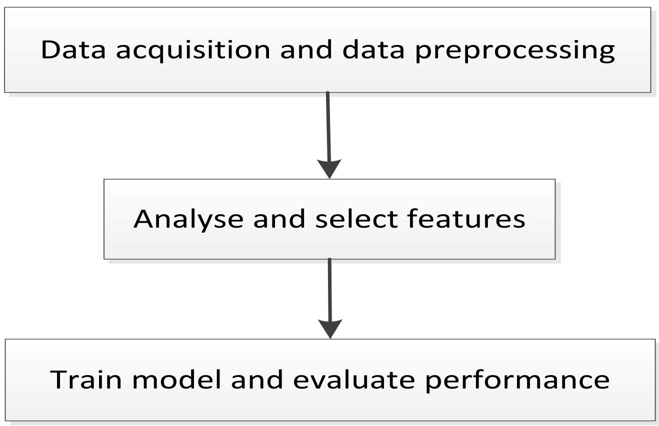


Figure 1: Experimental Steps for the Development of the Journal Classification Model

Establishing a Feature Space for Describing Journal Self-Citation Behavior

Journal anomalous self-citation is a kind of citation that contradicts the normal rules of scientific progress. The publishers and editors of some journals usually implicitly or directly require that a manuscript cites at least one paper published in their own journal to increase the cumulative citation count, or else the paper will be rejected. This is called journal anomalous self-citation behavior (Chang et al. 2013; Chorus and Waltman 2016; Opthof 2013; Wilhite and Fong 2012). Therefore, in this section, we attempted to analyse the characteristics of this undesirable self-citation behavior and to lay the foundation for distinguishing anomalous self-citation from normal self-citation.

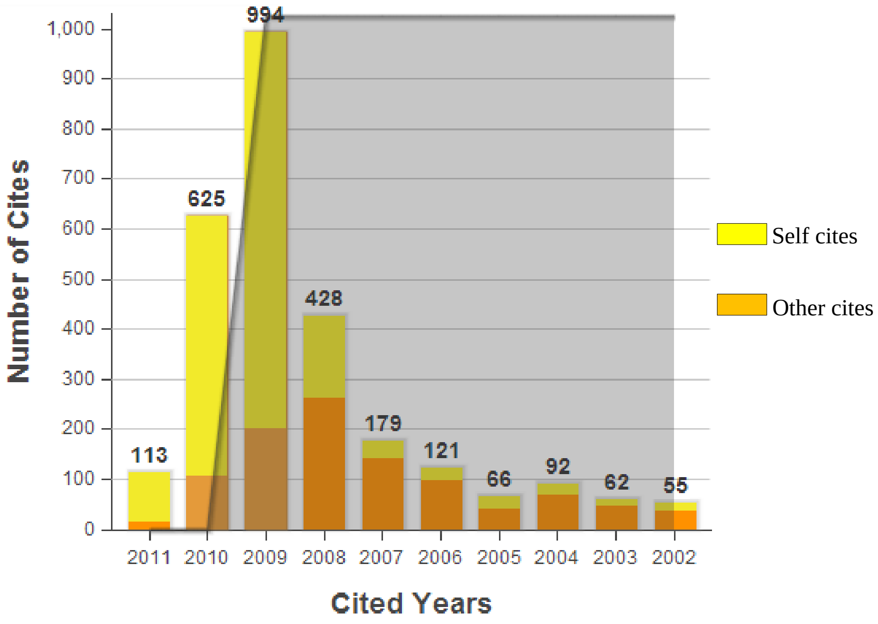
Two journals, *Pain Physician* and *Alzheimers and Dementia*, were selected for a more detailed description of anomalous self-citation behavior. They were included in the same JCR category of Clinical Neurology. The journal *Pain Physician* had been indexed in JCR since 2010 and its impact factor in 2011 was very high. However, the journal was suppressed by JCR in 2012 due to anomalous self-citation pattern found in the citation data. Another journal *Alzheimers and Dementia* was similar in citable item and subject ranking to the journal *Pain Physician* in 2011, as shown in Table 1.

Table 1: Indicators of Two Journals in JCR Category of Clinical Neurology in 2011

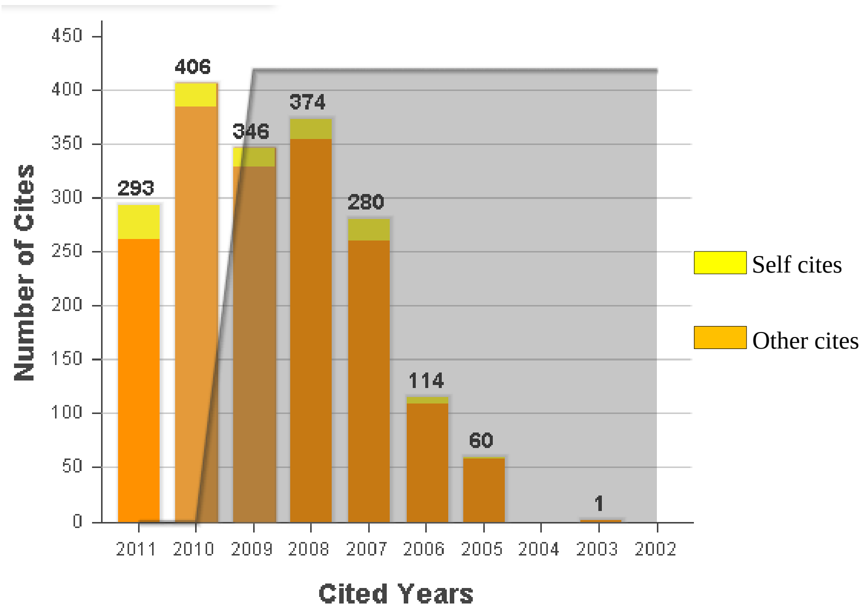
Journal name	**JIF**	**Citable Items**	**Rank**
Pain Physician	10.722	72	Q1 (4/192)
Alzheimers & Dementia	6.373	62	Q1 (11/192)

These anomalous self-citation pattern results in a significant distortion of the journal impact factor and rank that does not accurately reflect the journal's citation performance in the literature. So it shows that journals with anomalous self-citation have three abnormal characteristics: the total self-citation rate, the citation distribution, and the fluctuation in their impact factors.

Figures 2 and 3 show the citation distributions of two journals in 2011 and their impact factor trend. Clearly, journal self-citation is a common form of citation that agrees with academic rules and the demand for journal development, which is characterized by normal citation behavior. However, the anomalous self-citation produces abnormal signs. On one hand, abnormal journals with anomalous self-citation usually include one or more additional and irrelevant citations published by the same journals, and they usually have higher total self-citation rates than the average for the research field (Garfield 1997). One the other hand, the self-citation rates of the normal journals are higher in the first 3 years than those in other years. The self-citation distributions of the normal journals are similar to their overall citation distributions. Thus, many citations are accompanied by a substantial number of self-citations. However, the abnormal journals have a higher number of self-citations in the first 2 years compared with other citations, which results in an unusually high average self-citation rate (Yu et al. 2014). Furthermore, abnormal journals often have an unusually high number of self-citations and total citations in the first 3 years of the citation distribution as a large number of self-citations can result in large differences between impact factors and impact factors without self-citations (Campanario 2011).

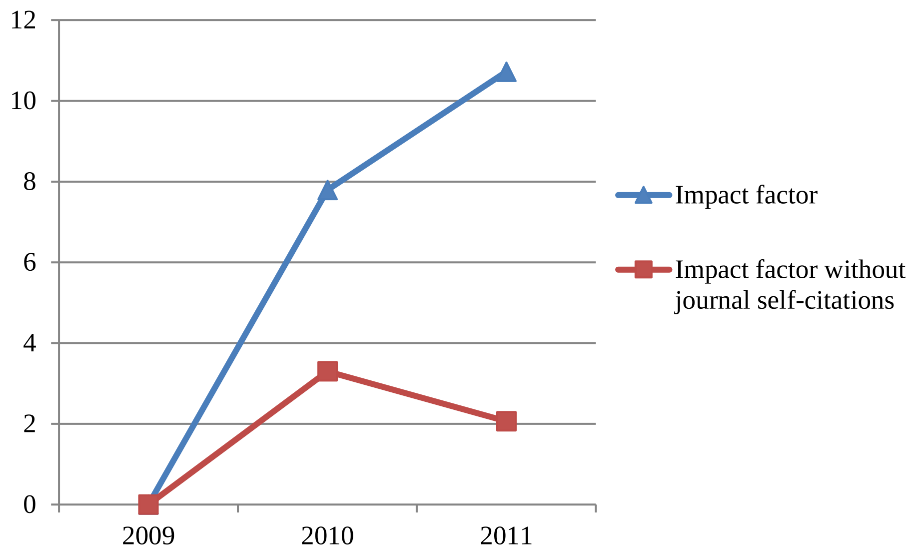


(a) Abnormal Journal *Pain Physician*

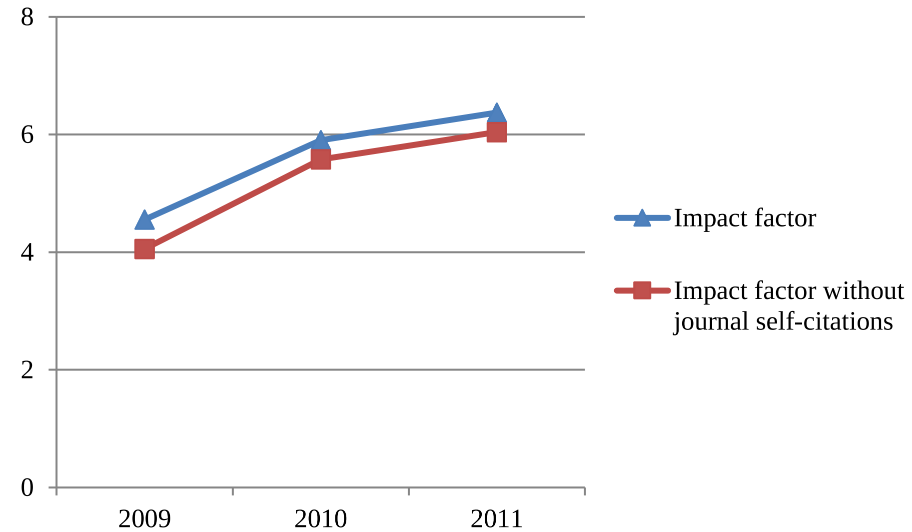


(b) Normal Journal *Alzheimers & Dementia*

Figure 2 : Citation Distribution Charts of Two Journals from the JCR Database in 2011



(a) Abnormal Journal *Pain Physician*



(b) Normal Journal *Alzheimers & Dementia*

Figure 3 : Impact Factor Trend Graph of two journals from the JCR database in 2011

Based on the difference between the self-citation behaviors of normal and abnormal journals, we selected relevant features for describing the citation behavior as comprehensively as possible. First, we selected indicators that represented the citation condition of a journal in a given year, i.e., the total citations, the total self-citations, the citations in the given year, and the self-citations in the given year. Second, the journal impact factor will change rapidly within a short period if it is disrupted by anomalous self-citation behavior, so the citations in the previous 2 years, the self-citations in the previous 2 years, the self-citation rate in the previous 2 years, the self-citation rate in the previous 3 years, and the total self-citation rate were calculated. In addition, the anomalous self-citation behavior is generally used to manipulate the impact factor, so we introduced two other attributes, i.e., the number of papers published in the journal in the given year and the reference count published in the previous 2 years per paper published in the same journal in the given year, which equaled the self-citation count in the previous 2 years divided by the number of papers published in the journal in the given year. Thus, 11 features were used to construct the feature space to describe the citation behavior.

Constructing the Sample Set

To develop an effective classification model, a journal set of different classes needed to be established, which is known as the training set. The samples in the training set included normal journals and abnormal ones with anomalous self-citation, with the normal journals being labeled Class 1 and the abnormal journals being labeled Class 2. However, because it is very difficult to detect anomalous self-citation as it generally occurs secretly between journal editors and authors, abnormal journals can only be identified from author complaints and traditional surveys and from the list that JCR publishes each year of removed journals with anomalous citations patterns. We obtained 48 journals with anomalous self-citation through a survey of the researchers around and selected 54 abnormal journals in the journal suppression list published by JCR each year. The final training sample set had 98 normal samples and 102 abnormal samples; therefore, there were two classes with 200 instances in the training set. The specific training samples can be found in the Appendix.

After learning the classification model based on the training set, another journal set, referred to as the test set was required to test the model's reliability and to verify its generalizability. The test samples and the training samples were independent as they used different data. In this study, journals in three subject areas from the JCR database: biology, mathematics and chemistry, applied, in the JCR from 2002 to 2014 were used as the test sets to validate the classification model.

In this research, the data used were obtained from JCR, which is a basic and comprehensive journal evaluation resource tool. Data for all journals were downloaded from JCR in January 2016.

Feature Selection

Based on the analysis described above, eleven journal features closely related to journal citations were identified: total citations, total self-citations, citations in the given year, self-citations in the given year, citations in the previous 2 years, self-citations in the previous 2 years, total self-citation rate, self-citation rate in the previous 2 years, self-citation rate in the previous 3 years, the number of papers published in the journal in the given year, and the reference count published in the previous 2 years per paper published in the same journal in the given year. In order to build an effective classification model, feature selection was needed to optimise the specific system index and improve the performance of the learning algorithm.

The feature selection method combined GA-Wrapper with RelifF was used to reduce the number of irrelevant and redundant features and to obtain the optimal features (Yu et al. 2014). After feature selection, the following four features obtained were used as the inputs for the classification model:

(1) Total self-citation rate (TSR)

$$\text{TSR}=\frac{\#\text{Total self-citations}}{\#\text{Total citations}}$$

(2) Self-citation rate in the previous two years (SR2Y)

$$\text{SR2Y}=\frac{\#\text{Self-citations in the previous two years}}{\#\text{Citations in the previous two years}}$$

(3) Self-citation rate in the previous three years (SR3Y)

$$\text{SR3Y}=\frac{\#\text{Self-citations in the previous three years}}{\#\text{Citations in the previous three years}}$$

(4) Reference count published in the previous two years per paper published in the same journal in this year (S2Y per paper)

$$\text{S2Y per paper}=\frac{\#\text{Self-citations in the previous two years}}{\#\text{The number of papers published in the journal in this year}}$$

After feature selection, four features obtained were used as the key indicators to detect journals with anomalous self-citation. Each feature could characterize the proportion of the self-citations in a certain period of time, and could distinguish normal journals and abnormal ones to a certain extent.

Classification by Deep Belief Network (DBN)

After feature selection, the journal classification model was constructed. In previous research, surface learning methods had been used to learn journal classification models. However, due to the limited data representation capability of surface learning, the performance of the classification model was difficult to improve. Therefore, in this research, DBN, an important deep learning method, was employed to perform the classification. DBN has a powerful ability to express and learn from a small number of samples and has been widely used in many areas.

Proposed by Hinton and Salakhutdinov (2006), the DBN is a deep neural network composed of multiple hidden layers of Restricted Boltzmann Machines (RBM) and a layer of back propagation neural network. By building a machine learning model with multiple hidden layers, more useful features can be included to improve classification performances.

As a learning module for the DBN, an RBM is a type of stochastic neural network model with a two-layer structure, symmetric connection and no self-feedback, as shown in Figure 4. There are two layers of units: *v* for the visible layer (the input data), which is formed by the visible units, and *h* for the hidden layer, as the feature detectors, which is formed by the hidden units. *W* is the connection weight between two two layers, which is the RBM learning result.

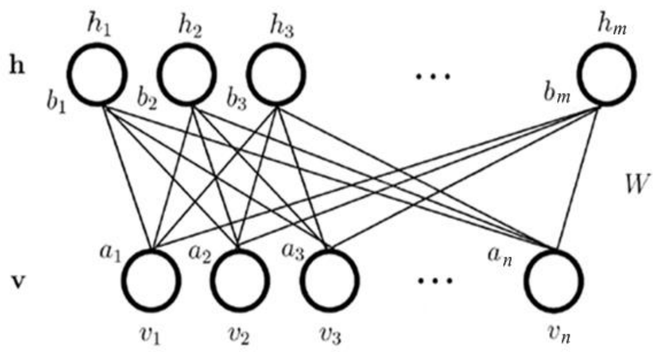


Figure 4: Diagram of a Restricted Boltzmann Machines (RBM)

The standard RBM has binary-valued hidden and visible units, that is

$$\forall i, j, v_i \in \{0,1\}, h_j \in \{0,1\}$$

The RBM consists of a matrix of weights W_{ij}, which are associated with the connection between a hidden unit h_j and a visible unit v_i, as well as the bias weights (offsets) a_i for the visible units and b_j for the hidden units. Given these, the *energy* of a configuration (pair of boolean vectors) (v,h) is defined as:

$$E(v,h\,|\,\theta) = -\sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij} v_i h_j$$

where ϑ ={W_{ij}, a_i, b_j}.

As the RBM has the shape of a bipartite graph and has no intra-layer connections, the hidden unit activations are mutually independent given the visible unit activations and conversely, the visible unit activations are mutually independent given the hidden unit activations (Carreira-Perpinan and Hinton 2005). The individual activation probabilities are given by

$$P(h_j = 1\,|\,v,\theta) = \sigma(b_j + \sum_i W_{ij} v_i)$$

$$P(v_i = 1\,|\,h,\theta) = \sigma(a_j + \sum_j W_{ij} h_j)$$

where σ denotes the logistic sigmoid, $\sigma(x) = \frac{1}{1+\exp(-x)}$.

The training process of an RBM adjusts parameter ϑ to optimise the fitting of the training data, that is, to find the parameter ϑ to maximize the probabilities of all training samples on the distribution.

After understanding the working principle and the training process of an RBM, the DBN, made up of multiple layers of RBM and a layer of back propagation neural network (Hinton 2009), can be trained, as shown in Figure 5. The training process of the DBN was divided into two steps. The first step was a pre-training process, for which each layer of RBM was trained unsupervised to ensure the eigenvectors could retain as much information as much as possible when mapped to a different feature space. The second step was a fine-tuning process. The BP network was established in the last layer of the DBN where the output of the RBMs served as its input. The BP network propagated error messages for each layer of RBM from the top to the bottom to fine-tune the DBN.

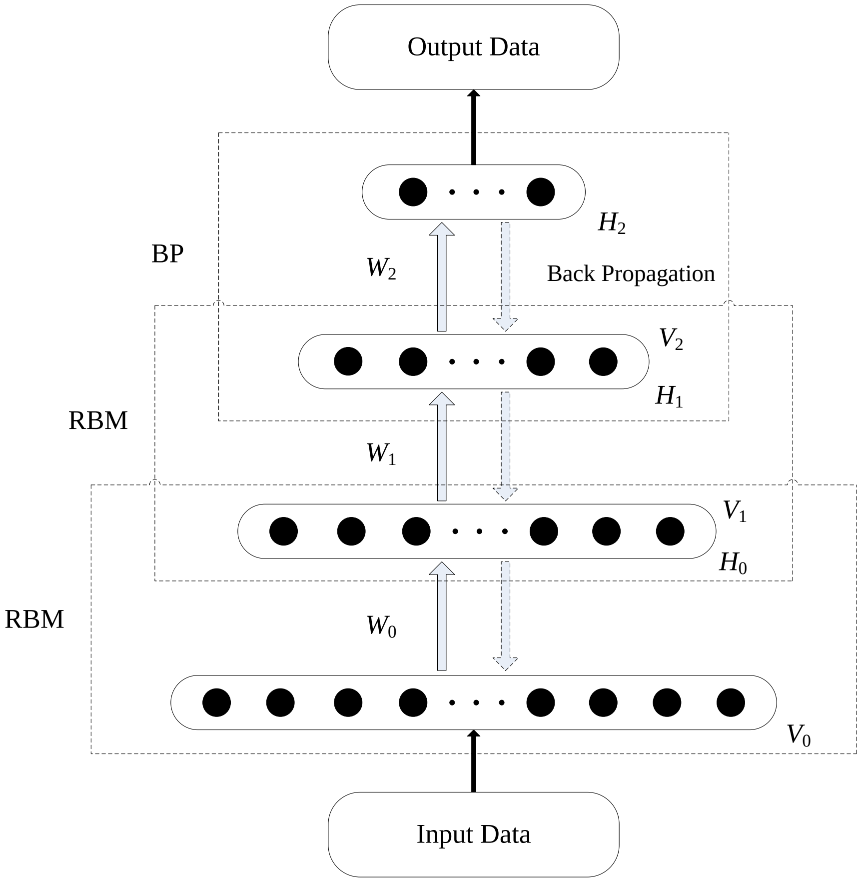


Figure 5 : Diagram of a Deep Belief Network (DBN)

The training algorithm for the DBN was as follows (Hinton, Osindero and Teh. 2006) – *X* was nominated as an input matrix, or a set of feature vectors. First, an RBM was trained on *X* to obtain a weight matrix *W* which was used as the weight matrix between the lower two layers of the network. Second, *X* was transformed by the RBM to produce a new data *X'*, either by sampling or by computing the mean activation of the hidden units. Then, this procedure was repeated with $X \leftarrow X'$ for the next pair of layers, until the top two layers of the network were reached. Finally, all the parameters of this deep architecture were fine-turned with respect to a supervised training criterion (after adding an extra classifier to convert the learned representation into supervised predictions).

In this research, a DBN model was trained using the optimal features outlined above to classify the journals into normal journals (Class 1) and abnormal journals (Class 2). For the learned model, the value of the journal features was the input, and the class of journals is the output. To eliminate learning inefficiencies, the journal feature values were normalized so that the original input data fell between [0, 1]. To achieve the classification function, a

Softmax classifier was combined with the DBN to obtain the classification results.

RESULTS AND DISCUSSION

Classification Model Performance Evaluation

The DBN method was used in this study because of its powerful data representation ability. Through a number of experiments and parameter adjustments, a suitable classification model was developed from the 200 training samples made up of three layers of RBM and a softmax classifier. There were 4, 2 and 2 RBN nodes of each layer in the DBN model.

The training set consisted of 98 normal journal samples and 102 abnormal journal samples. In order to make the samples more clearly, we extracted some training samples from the training set, as shown in Table 2.

Table 2: Value of Four Features of the Training Samples (EXTRACT)

Journal Num.	TSR	SR2Y	SR3Y	S2Y per paper	Label
1	0.028	0.029	0.016	0.056	Class 1
2	0.041	0.056	0.045	0.233	Class 1
…	…	…	…	…	Class 1
97	0.038	0.057	0.050	0.645	Class 1
98	0.078	0.143	0.120	0.353	Class 1
99	0.458	0.651	0.550	1.000	Class 2
100	0.439	0.619	0.580	1.225	Class 2
…	…	…	…	…	Class 2
199	0.127	0.448	0.347	1.302	Class 2
200	0.696	0.764	0.734	3.827	Class 2

Table 3 and Table 4 show the results and performances of classification prediction for the training set using the DBN model. The accuracy of the DBN model was 98% for the 200 training set samples as only two normal journal samples and two abnormal samples were misclassified.

Three common measures are used to evaluate a classifier's performance: classification precision, recall and the F-measure (Bataineh et al. 2011). Precision is the fraction of retrieved instances that are relevant, recall is the fraction of relevant instances that are retrieved, and the F-measure is the weighted harmonic mean of Precision and Recall. In this research, this was expressed by: F-measure = Precision * Recall * 2 / (Precision + Recall). Of these three metrics, precision is a measure of exactness and recall is a measure of completeness. In simple terms, high precision means that an algorithm returns substantially more relevant than irrelevant results, and high recall means that the algorithm returned most of the relevant results. Therefore, these three metrics were used to reveal the classification performance, with the classification results obtained from the DBN model shown in Table 4.

In the training set, the DBN model had good classification performance, with the precision and recall of Class 1 and Class 2 being about 98%; in other words, the trained model could accurately classify the normal journal samples and abnormal ones in the training set.

Table 3: Results of Class 1/2 Prediction for the Training Samples

		Manual judgment		
		Class 1	**Class 2**	**Subtotal**
DBN model judgment	Class 1	96	2	98
	Class 2	2	100	102
	Subtotal	98	102	200

Table 4: Performances of Class 1/2 Prediction for the Training Samples

	Precision	**Recall**	**F-Measure**
Class 1	0.980	0.980	0.980
Class 2	0.980	0.980	0.980

However, as the DBN model learned from the training set, the good classification performance of the training set did not necessarily indicate that the model had strong classification ability. Therefore, the test samples were used to evaluate the generalizability of the classification model. All journals published from 2002 to 2014 and indexed by JCR in three subject areas (biology, mathematics and chemistry, applied) were selected as test samples to observe the model's classification results. And the journal sets published in the areas of biology, mathematics and chemistry, applied were called as test sets 1, 2, and 3 respectively. After deleting journals with incomplete information (i.e. the journals whose citable item is zero), the test samples consisted of 4512 journals.

A class 1/2 journal classification was performed for the journal samples in the three test sets using the DBN model. To verify the classification accuracy of the model, the test samples classes were manually determined. We invited nine professors from Harbin Institute of Technology to categorize journals manually, and the classification result of class 1 or class 2 was set as the verification standard. Three experts in each of the subject areas analysed the journals and determined the manual classification. And these manual classification results were used as the validation criteria. Table 5 shows the classification results compared with the results of the manual assessments and also indicates which sample journals in the three test sets were in Class 1 or Class 2. The performance measures based on the results shown in Table 5 are shown in Table 6. After comparing the model's classification results with the manually assessed results, the classification accuracy of the model obtained for the test sets was 98.6 percent for all test samples. The accuracy of the DBN model for the three test set samples was 98.0 percent, 96.5 percent and 99.4 percent respectively, indicating that most test samples were correctly classified. The classification performance of the model on the three test sets was also good. When classifying the normal journals in the test sets, the precision and the recall of the DBN model were both higher than 0.97; however, when classifying the abnormal journals, the precision was slightly lower, but higher than 0.65, with the recall being over 0.93 for all three test sets. This suggested that the model had good classification performance and generalizability and was capable of effectively identifying suspect journals.

Table 5: Results of Class 1/2 Prediction for the Test Samples

			Manual judgment		
			Class 1	**Class 2**	**Subtotal**
DBN model judgment	Test set 1	Class 1	874	2	876
		Class 2	16	30	46
		Subtotal	890	32	922
	Test set 2	Class 1	695	1	696
		Class 2	21	48	69
		Subtotal	716	49	765
	Test set 3	Class 1	2774	1	2775
		Class 2	15	35	50
		Subtotal	2789	36	2825

Table 6: Performances of Class 1/2 Prediction for Three Test Sets

		Precision	**Recall**	**F-Measure**
Test set 1	Class 1	0.998	0.982	0.990
	Class 2	0.652	0.938	0.769
Test set 2	Class 1	0.999	0.971	0.985
	Class 2	0.696	0.980	0.814
Test set 3	Class 1	1.000	0.995	0.997
	Class 2	0.700	0.972	0.814

There are many classical classification algorithms for machine learning, such as support vector machine (SVM) (Burges 1998). SVM is a supervised learning model based on statistical learning theory that can improve the generalizability of a learning machine by minimizing structured risk. When the statistical sample size is small, it can also obtain good statistical rules. In a previous study, the logistic regression (LR) model was also found to be useful in distinguishing normal journals from abnormal journals with anomalous self-citation (Yu et al. 2014). Therefore, SVM and LR algorithms were chosen to build the journal classification model and compare the respective classification performances.

The classification results for the three models on the training set are shown in Table 7. To obtain reliable and stable results, a ten-fold cross-validation was conducted in the learning process (Kohavi 1995). The number of misclassified journal samples in the training set was 4, 5 and 4 respectively based on the three classification models, and the accuracy of the three models was over 97 percent for the 200 training set samples.

Table 7: Classification Results for the Three Models on the Training Set

	Number of Class 1 samples misclassified as Class 2	**Number of Class 2 samples misclassified as Class 1**	**Number of correctly classified samples**	**Classification accuracy**
DBN model	2	2	196	0.980
SVM model	3	2	195	0.975
LR model	2	2	196	0.980

After learning the classification models in the training set, the test sets were then used to evaluate the generalizability of the models. The purpose of this paper was to automatically identify suspect journals by establishing a journal classification model; therefore, the abnormal journals (Class 2) were the main focus of this research. Table 8 shows the classification performances of the three models for the abnormal journal samples in the test sets, for which precision, recall and F-measure were also used to measure the performance of the classifiers.

Table 8: Performances of Class 2 Prediction for the Test Sample Journals

		Precision	**Recall**	**F-Measure**
Test set 1	DBN model	0.652	0.938	0.769
	SVM model	**0.377**	0.625	**0.471**
	LR model	**0.333**	0.656	**0.442**
Test set 2	DBN model	0.696	0.980	0.814
	SVM model	0.745	0.776	0.760
	LR model	0.707	0.837	0.766
Test set 3	DBN model	0.700	0.972	0.814
	SVM model	**0.340**	**0.444**	**0.386**
	LR model	**0.452**	0.778	**0.571**

It can be seen from Table 8 that the precision and recall of the DBN model was more than 65 percent and 93 percent respectively for all three test sets, indicating that the DBN model was able to accurately identify the abnormal journal samples in the sample sets. However, the classification results from the SVM model were less than satisfactory as it did not achieve a very high classification precision for the Class 2 journal samples in test set 1 (38%) and 3 (34%) and the Class 2 recall was only 44 percent for test set 3. The LR model performance was also unsatisfactory, with the precision only above 33 percent for the three test sets (33.3%, 70.7% and 45.2%), and the recall only above 65 percent (65.6%, 83.7% and 77.8%). That the Class 2 precision was less than 50 percent indicated that the model misclassified many Class 1 samples more than the number of Class 2 samples, and that the Class 2 recall was less than 50 percent indicated that less than half of Class 2 samples were identified by the model. From this analysis, it is apparent that the DBN model was superior to either the SVM model or the LR model in terms of applicability and classification performance, and was more accurate identifying abnormal journals. Therefore, from these results, it was concluded that the DBN was a good model for the classification of normal and abnormal journals.

An Empirical Study on Abnormal Journals

Every year, hundreds of thousands of scientific papers are published in various scientific journals as journals have become the most important way to disseminate the scientific research results and to measure the performance and impact of scientific research in universities, research institutions and other research bodies. Recently, the impact factor has become a prominent indicator of a journal's standing; therefore, anomalous self-citation could seriously undermine not only the authenticity and fairness of the journal evaluation system but also scientific research development. Therefore, to deal with this academic misconduct, abnormal journals with anomalous self-citation need to be easily identified. As the DBN model was proven to be effective in classifying normal and abnormal journals, an empirical study on abnormal journals was conducted.

Since the classification results from the obtained DBN model were satisfactory in three different test sets, the model could classify the unlabelled journal samples into normal or abnormal and the classification results were meaningful and effective according to the theory of pattern recognition. In 2014, JCR indexed 11,770 journals, covering 242 subject areas. Due to limited time, 12 subject areas were selected as the empirical objects of this research, as listed alphabetically in Table 9. After excluding journals with incomplete information, the empirical research objects included 1138 journals. First, the value of four features for the 1138 journal samples were calculated based on JCR data, and then the DBN model was used to classify the journals into normal or abnormal. Afterwards, the abnormal journals classified as Class 2 by the DBN model were counted for each subject area. The classification results are shown in Table 9. It was mentioned that the journal samples classified as Class 2 were referred to as suspected journals in this empirical study.

From the 1138 journals indexed by JCR in 2014, 78 journals were identified as suspect journals with anomalous self-citation, with the total proportion of suspect journals being 6.9 percent, as shown in Table 9. For the different subject areas, the percentage of suspect journals identified by the DBN model varied widely, ranging from 2.1 percent to 13.3 percent, and there was a great deal of difference between similar subject areas in terms of proportion of suspect journals. For example, medicine, general and internal, medicine, legal and medicine, research and experimental all belonged to the category of Medicine, with the proportion of suspect journals being only 3.3 percent for medicine, research and experimental, but 13.3 percent for medicine, legal. Suspect journals were found in each empirical subject area, which suggested that academic fraud was an increasing worrying problem.

Anomalous self-citation can indeed improve a journal's impact factor in a short time, but over the long run, such manipulative behavior not only limits the rational use of research resources, but also undermines the normal development of journals. Therefore, the results described above demonstrate that deep learning methods are able to successfully classify normal and abnormal journals more accurately than surface learning methods. A classification model was constructed based on the DBN method to identify abnormal journals, which could be used to supervise the normal, orderly development of journals and ensure journal development on the right track.

Table 9: List of 12 Selected Subject Areas and the Number of Suspect Journals in 2014

Subject Area	Journals Indexed by JCR	Suspect Journals	Proportion of Suspect Journals
Computer science, artificial intelligence	121	12	9.9%
Computer science, cybernetics	22	2	9.1%
Computer science, hardware and architecture	47	1	2.1%
Engineering, areospace	28	1	3.6%
Engineering, electrical and electronic	237	20	8.4%
Engineering, mechanical	125	10	8.0%
Medicine, general and internal	148	7	4.7%
Medicine, legal	15	2	13.3%
Medicine, research and experimental	120	4	3.3%
Physics, applied	138	9	6.5%
Physics, condensed matter	63	3	4.8%
Physics, multidisciplinary	74	7	9.5%
Subtotal	1138	78	6.9%

CONCLUSION

In summary, the results in this paper suggest that the DBN method could be used to automatically recognise abnormal journals. A training journal set and three test journal sets of normal journals and abnormal journals were first established, after which four relevant and concise features were selected by analysing the differences between the normal and abnormal journals. A classification model based on the DBN method was then constructed in the training set, and the validity of the model verified using three test sets. Subsequently, the classical SVM and LR methods were compared to the DBN method, from which it was found that the DBN model had a significantly better performance in identifying abnormal journals. Finally, an empirical study on abnormal journals was performed, and it was found that 6.9% of empirical journals were suspect.

There were several limitations in this research. Most importantly, the number of training samples was limited because abnormal journals with anomalous self-citation are generally concealed. Classification performance is generally better when there are more samples in the training set because there is a more effective learning of the classification rules; therefore, the limited training set size constrained the classification performance of the developed model. In addition, while the DBN model was proven to be effective in identifying abnormal journals, it was unable to identify other types of abnormal journals, such as journals with coercive citations to certain other journals. Despite these limitations, the results of this study suggest that abnormal journals can be identified automatically and rapidly. Therefore, this method could save a great deal of the human effort needed to monitor journals, and facilitate an honest and open development of academic research.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 71704035 and 71473034), the China Postdoctoral Science Foundation (Grant No. 2017M611361) and the Fundamental Research Funds for the Central Universities (Grant No. HEUCFM180902).

REFERENCES

Arnold, D.N. 2009. Integrity under attack: the state of scholarly publishing. *SIAM news*, Vol. 42, no. 10: 2-3.

Bataineh, B., Abdullah, S.N.H.S., Omar, K. and Faidzul, M. 2011. Adaptive thresholding methods for documents image binarization. In: Martínez-Trinidad J.F., Carrasco-Ochoa J.A., Ben-Youssef Brants C., Hancock E.R. (eds) *Pattern Recognition*. MCPR 2011. Lecture Notes in Computer Science, Vol. 6718. Springer, Berlin, Heidelberg.

Burges, C.J. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, Vol. 2, no. 2: 121-167.

Campanario, J.M. 2011. Large increases and decreases in journal impact factors in only one year: The effect of journal self-citations. *Journal of the American Society for Information Science and Technology*, Vol. 62, no. 2: 230-235.

Carreira-Perpinan, M.A. and Hinton, G.E. 2005. On contrastive divergence learning. International Conference on Artificial Intelligence and Statistics (*AISTATS)*, Vol. 10: 33-40.

Chang, C.L., Mcaleer, M. and Oxley, L. 2013. Coercive journal self-citations, impact factor, journal influence and article influence. *Mathematics and Computers in Simulation*, Vol. 93, no. 10: 190-197.

Chorus, C. and Waltman, L. 2016. A large-scale analysis of impact factor biased journal self-citations. *PLoS One*, Vol. 11, no. 8. Available at: https://doi.org/10.1371/journal.pone.0161021.

Falagas, M.E. and Alexiou, V.G. 2008. The top-ten in journal impact factor manipulation. *Archivum immunologiae et therapiae experimentalis*, Vol. 56, no. 4: 223-226.

Garfield, E. 1997. Dispelling a few common myths about journal impact factors. *The Scientist*, Vol. 11, no. 3: 11.

Garfield, E. 1955. Citation indexes for science: A new dimension in documentation through association of ideas. *Science,* Vol. 122, no. 3159: 108-111.

Garfield, E. and Sher, I.H. 1963. New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, Vol. 14, no. 3: 195-201.

Hemmingsson, A., Edgren, J., Mygind, T. and Skjennald, A. 2002a. Impact factors in scientific journals. *Journal of Magnetic Resonance Imaging*, Vol. 15, no. 5: 619-619.

Hemmingsson, A., Mygind, T., Skjennald, A., Edgren, J. and Rogers, L. F. 2002b. Manipulation of impact factors by editors of scientific journals. *American Journal of Roentgenology*, Vol. 178, no. 3: 767.

Heneberg, P. 2013. Effects of print publication lag in dual format journals on scientometric indicators. *PloS One*, Vol. 8, no. 4. Available at: https://doi.org/10.1371/journal.pone.0059877.

Hinton, G.E. 2009. Deep belief networks. *Scholarpedia*, Vol. 4, no. 5: 5947.

Hinton, G.E. Osindero, S., and Teh, Y.W. 2006. A fast learning algorithm for deep belief nets. *Neural Computation*, Vol. 18, no. 7: 1527-1554.

Hinton, G.E. and Salakhutdinov, R.R. 2006. Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, no. 5786: 504-507.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence IJCAI'95,* Vol. 14, no. 2: 1137-1145.

Kurmis, A.P. 2003. Understanding the limitations of the journal impact factor. *Journal of Bone and Joint Surgery: American Volume*, Vol. 85, no. 12: 2449-2454.

Larivie`re V., Kiermer V., MacCallum C.J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S., Taylor, S, and Curry, S. 2016. A simple proposal for the publication of journal citation distributions. *bioRxiv*. Available at: https://doi.org/10.1101/062109.

Maity, B.K. and Hatua, S.R. 2016. Designing a model to evaluate scholarly publications with special reference to social sciences in India. *Scientometrics*, Vol. 109, no. 3: 2031-2048.

Martin, B.R. 2016. Editors' JIF-boosting stratagems–Which are appropriate and which not?. *Research Policy*, Vol. 45, no. 1: 1-7.

Mavrogenis, A.F., Ruggieri, P. and Papagelopoulos, P.J. 2010. Editorial: self-citation in publishing. *Clinical Orthopaedics and Related Research*, Vol. 468, no. 10: 2803-2807.

Opthof, T. 2013. Inflation of impact factors by journal self-citation in cardiovascular science. *Netherlands Heart Journal*, Vol. 21, no. 4: 163-165.

Rijcke, S. and Rushforth, A. 2015. To intervene or not to intervene; is that the question? On the role of scientometrics in research evaluation. *Journal of the Association for Information Science and Technology*, Vol. 66, no. 9: 1954-1958.

Simons, K. 2008. The misused impact factor. *Science*, Vol. 322, no. 5899: 165.

Tort, A.B., Targino, Z.H., and Amaral, O.B. 2012. Rising publication delays inflate journal impact factors. *PLoS One*, Vol. 7, no. 12. Available at: https://doi.org/10.1371/journal.pone.0053374.

Vanclay, J.K. 2012. Impact factor: outdated artefact or stepping-stone to journal certification?. *Scientometrics*, Vol. 92, no. 2: 211-238.

Wilhite, A.W. and Fong, E.A. 2012. Coercive citation in academic publishing. *Science*, Vol. 335, no. 6068: 542-543.

Yu, G., Yang, D.H. and He, H.X. 2011. An automatic recognition method of journal impact factor manipulation. *Journal of Information Science*, Vol. 37, no. 3: 235-245.

Yu, L. and Yu, H. 2016. Does the average JIF percentile make a difference?. *Scientometrics*, Vol. 109, no. 3: 1979-1987.

Yu, T., Yu, G. and Wang, M.Y. 2014. Classification method for detecting coercive self-citation in journals. *Journal of Informetrics*, Vol. 8, no. 1: 123-135.

APPENDIX

Appendix A: Training Journal Samples

Normal journal samples			Abnormal journal samples		
Journal name	**JIF**	**Year**	**Journal name**	**JIF**	**Year**
4OR-Q J OPER RES	1	2014	INT J CRASHWORTHINES	0.789	2011
ACCOUNTS CHEM RES	22.323	2014	IRAN J FUZZY SYST	1.056	2011
ACTA ADRIAT	0.655	2014	AMFITEATRU ECON	0.838	2013
BIOGEOCHEMISTRY	3.488	2014	APPL INTELL	1.853	2012
CAN ENTOMOL	0.837	2014	ARCH MIN SCI	0.608	2013
CLIN IMAG	0.81	2014	B INDONES ECON STUD	1.067	2013
CURR OPIN HIV AIDS	4.68	2014	BUS LAWYER	0.935	2012
DIFFERENTIATION	3.437	2014	CYTOJOURNAL	1.2	2012
EUR J ENTOMOL	0.975	2014	ELECTR POW COMPO SYS	0.664	2013
FASEB J	5.043	2014	EMERG MARK FINANC TR	0.468	2013
FRONT ZOOL	3.051	2014	ENTERP INF SYST-UK	9.256	2012
GENETICA	1.4	2014	ENTERP INF SYST-UK	3.684	2011
HUM BIOL	0.921	2014	ENTERP INF SYST-UK	0.786	2010
IEEE VEH TECHNOL MAG	1.75	2014	FORENSIC TOXICOL	5.756	2013
INT J AEROACOUST	0.403	2014	GEOTEXT GEOMEMBRANES	2.376	2013
IZV MATH+	0.63	2014	GEOTEXT GEOMEMBRANES	2.159	2012
J APPL LOGIC	0.576	2014	GEOTEXT GEOMEMBRANES	2.036	2011
J ECOL	5.521	2014	GEOTEXT GEOMEMBRANES	2.59	2010
J FIELD ORNITHOL	0.988	2014	GEOTEXT GEOMEMBRANES	4.039	2009
J NONLINEAR MATH PHY	0.733	2014	GEOTEXT GEOMEMBRANES	3.701	2008
J PLANT BIOL	1.208	2014	GEOTEXT GEOMEMBRANES	3.05	2007
J WILDLIFE DIS	1.355	2014	GEOTEXT GEOMEMBRANES	1.167	2006
JETP LETT+	1.359	2014	INT J COMMUN SYST	1.106	2013
MAGN RESON IMAGING	2.09	2014	INT J ELEC POWER	3.432	2012
MIS QUART	5.311	2014	INT J ELEC POWER	2.247	2011
NEUROL SCI	1.447	2014	INT J ELEC POWER	2.212	2010
NURS PHILOS	0.833	2014	INT J ELEC POWER	1.613	2009
OPER DENT	1.671	2014	INT J SUST DEV WORLD	1.771	2013
PHYSIOL BEHAV	2.976	2014	INT J SUST DEV WORLD	1.213	2012
PPAR RES	2.509	2014	INT J SUST DEV WORLD	0.965	2011
REND LINCEI-SCI FIS	0.75	2014	INTERLEND DOC SUPPLY	0.35	2013
REV MVZ CORDOBA	0.104	2014	J REAL ESTATE RES	1.439	2013
SIAM J MATH ANAL	1.265	2014	J REAL ESTATE RES	0.925	2012
SPE RESERV EVAL ENG	0.99	2014	J REAL ESTATE RES	1.075	2011
TARGET ONCOL	4	2014	J VIB CONTROL	4.355	2013

VET RES	2.815	2014	J VIB CONTROL	1.966	2012
CA-CANCER J CLIN	101.78	2011	JPC-J PLANAR CHROMAT	0.67	2013
NEW ENGL J MED	53.298	2011	LANG CULT CURRIC	0.405	2013
ANNU REV IMMUNOL	52.761	2011	MICROSURG	2.421	2013
REV MOD PHYS	43.933	2011	MICROSURG	1.623	2012
CHEM REV	40.197	2011	MICROSURG	1.605	2011
NAT REV MOL CELL BIO	39.123	2011	N AM J ECON FINANC	1.5	2013
NAT REV GENET	38.075	2011	ORGAN ENVIRON	1.386	2013
NAT REV CANCER	37.545	2011	PAK VET J	1.392	2013
ADV PHYS	37	2011	PAK VET J	1.365	2012
NATURE	36.28	2011	POLYM-PLAST TECHNOL	1.481	2012
NAT GENET	35.532	2011	POLYM-PLAST TECHNOL	1.279	2011
ANNU REV BIOCHEM	34.317	2011	STAND GENOMIC SCI	3.167	2013
NAT REV IMMUNOL	33.287	2011	STAND GENOMIC SCI	2.007	2012
NAT MATER	32.841	2011	T EMERG TELECOMMUN T	0.783	2013
SCIENCE	31.201	2011	TURK J BOT	1.6	2012
NAT REV NEUROSCI	30.445	2011	TURK J BOT	1.991	2011
INFORM PROCESS MANAG	1.119	2011	VIDEOSURGERY MINIINV	1.092	2013
J MATER CHEM	5.968	2011	VIDEOSURGERY MINIINV	0.757	2012
J EUR CERAM SOC	2.353	2011	VIDEOSURGERY MINIINV	1	2011
AAPS J	5.086	2011	INT J NONLINEAR SCI	2.345	2005
ABH MATH SEM HAMBURG	0.222	2011	INT J NONLINEAR SCI	4.386	2006
ACM T MATH SOFTWARE	1.922	2011	CHAOS SOLITON FRACT	3.315	2009
ACM T COMPUT SYST	1.188	2011	CHAOS SOLITON FRACT	2.98	2008
ACM T SENSOR NETWORK	1.808	2011	CHAOS SOLITON FRACT	3.025	2007
ACS APPL MATER INTER	4.525	2011	CHAOS SOLITON FRACT	2.042	2006
ACS NANO	11.421	2011	CHAOS SOLITON FRACT	1.938	2005
ACTA ANAESTH SCAND	2.188	2011	CHAOS SOLITON FRACT	1.526	2004
ACTA APPL MATH	0.899	2011	CHAOS SOLITON FRACT	1.064	2003
ACTA BIOCHIM POL	1.491	2011	CHAOS SOLITON FRACT	0.872	2002
ADAPT PHYS ACT Q	1.487	2011	CHAOS SOLITON FRACT	0.839	2001
ADDICT BEHAV	2.085	2011	INFORM SCIENCES	2.833	2011
ADDICTION	4.313	2011	INFORM SCIENCES	2.836	2010
ADV AGRON	5.204	2011	INFORM SCIENCES	3.291	2009
AERONAUT J	0.482	2011	INFORM SCIENCES	3.095	2008
ADV STRUCT ENG	0.324	2011	INFORM SCIENCES	2.147	2007
ALCOHOL CLIN EXP RES	3.343	2011	ENERGY EDUC SCI TECH	9.333	2010
AM J DENT	0.757	2011	INT J HYDROGEN ENERG	4.054	2011
ANN EMERG MED	4.133	2011	INT J HYDROGEN ENERG	4.057	2010
ANN LIMNOL-INT J LIM	0.93	2011	INT J HYDROGEN ENERG	3.945	2009

APPL CATEGOR STRUCT	0.6	2011	INT J HYDROGEN ENERG	3.452	2008
ATOM ENERGY+	0.077	2011	CHRONOBIOL INT	4.028	2011
AUK	2.156	2011	CHRONOBIOL INT	5.576	2010
AUST NZ J STAT	0.436	2011	CHRONOBIOL INT	3.987	2009
AUSTRALAS J DERMATOL	1	2011	CHRONOBIOL INT	3.495	2008
ADDICT BIOL	4.833	2011	CHRONOBIOL INT	3.771	2007
AUSTRIAN J FOR SCI	0.227	2011	CHRONOBIOL INT	2.517	2006
B EUR ASSOC FISH PAT	0.288	2011	CHRONOBIOL INT	2.472	2005
B BRAZ MATH SOC	0.5	2011	PLANT MOL BIOL REP	2.453	2011
B AM MUS NAT HIST	2.905	2011	STRUCT CHEM	1.846	2011
BRIT J PHARMACOL	4.409	2011	STRUCT CHEM	1.727	2010
BRIT J PSYCHIAT	6.619	2011	STRUCT CHEM	1.637	2009
BRYOLOGIST	0.902	2011	STRUCT CHEM	1.433	2008
CAN MATH BULL	0.265	2011	COMPUT-AIDED CIV INF	3.382	2011
CANCER EPIDEM BIOMAR	4.123	2011	COMPUT-AIDED CIV INF	3.17	2010
CARDIOL YOUNG	0.759	2011	COMPUT-AIDED CIV INF	1.989	2009
CFI-CERAM FORUM INT	0.051	2011	INT J COMPUT INTEG M	1.071	2011
CLASSICAL QUANT GRAV	3.32	2011	STRAHLENTHER ONKOL	3.561	2011
COMPUT APPL ENG EDUC	0.333	2011	STRAHLENTHER ONKOL	3.567	2010
SCIENTOMETRICS	1.966	2011	STRAHLENTHER ONKOL	3.776	2009
J AM SOC INF SCI TEC	2.081	2011	STRAHLENTHER ONKOL	3.005	2008
ONLINE INFORM REV	0.939	2011	STRAHLENTHER ONKOL	2.846	2000
J INF SCI	1.299	2011	MOL BIOL REP	2.929	2011
			BRIT J ORAL MAX SURG	1.95	2011
			BRIT J ORAL MAX SURG	1.89	2010
			BRIT J ORAL MAX SURG	1.327	2009
			CMES-COMP MODEL ENG	4.785	2008