Patent Evaluation Based on Technological Trajectory Revealed in Relevant Prior Patents

Sooyoung Oh[1], Zhen Lei[2], Wang-Chien Lee[1], and John Yen[3]

[1] Department of Computer Science and Engineering
[2] Department of Energy and Mineral Engineering
[3] College of Information Sciences and Technology,
The Pennsylvania State University, University Park, PA 16802, USA
{sooh,wlee}@cse.psu.edu, zlei@psu.edu, jyen@ist.psu.edu

Abstract. It is a challenging task for firms to assess the importance of a patent and identify valuable patents as early as possible. Counting the number of citations received is a widely used method to assess the value of a patent. However, recently granted patents have few citations received, which makes the use of citation counts infeasible. In this paper, we propose a novel idea to evaluate the value of new or recently granted patents using *recommended relevant prior patents*. Our approach is to exploit trends in temporal patterns of relevant prior patents, which are highly related to patent values. We evaluate the proposed approach using two patent value evaluation tasks with a large-scale collection of U.S. patents. Experimental results show that the models created based on our idea significantly enhance those using the baseline features or patent backward citations.

Keywords: patent, evaluation, ranking.

1 Introduction

Patent evaluation, including predicting a patent's future value, comparing the value of patents in a given patent set, and identifying influential patents in a field or within a company, is a challenging but important task for technology and innovation management in a firm. The *forward citations* of a patent (i.e., citations to the patent made by other patents granted in later/*forward* times), in combination with other patent information such as the number of claims, etc, have been widely used as a measure to assess the patent economic value [1, 2, 3, 4]. As highly cited patents imply a number of successful lines of innovation (that is why they are highly cited), their inventions are likely to be technologically significant and economically valuable. However, it often takes years for a patent to receive sufficient information of forward citations in order to make meaningful assessment of its value.

This paper addresses the challenging problem of evaluating patents at the early stage of their patent life when there is little information about forward citations. To highlight this challenge, we ask the following question: *Can we*

V.S. Tseng et al. (Eds.): PAKDD 2014, Part I, LNAI 8443, pp. 545–556, 2014.

evaluate patents immediately after they are granted when there is no forward citation information at all? This is not merely a theoretically interesting question stressing the limitation of forward citation-based approaches for patent evaluation, but one with practical importance. A *patent grant*, to a large extent, secures the patent protection and ascertains the scope of the right for an invention. This is often the point of time when the patent owner decides what to do with the invention, e.g., developing/incorporating it into products in house or licensing to someone else. This is also the point of time when the deal of licensing/selling the invention, if any, is made. Therefore, it is crucial to predict the value of a patent *right after* its grant.

In the absence of forward citation information, one naturally turns to the *backward citations* of the patent under evaluation, i.e., the references made by the focal patent to prior patents (which are granted in past/*backward* times), besides other useful information such as the number of claims, the number of figures, etc. Indeed, previous studies [1, 2, 3, 4] have shown that the information of patent backward citations can be used to measure the *novelty* of a patent, and thus useful for evaluating its technological and economic value of a patent as well. Patent backward citations provide information about technologically relevant prior patents for the focal patent, and we can use it to estimate how novel and in what stage of the technology trajectory the focal patent is. However, the idea of using backward citations for patent evaluation also has limitations. First, backward citations are to a large extent missing for one or two years prior to the patent grant year. The reason is simple. Applicants and examiners reference prior patents either at patent filing or during the early stage of patent examination, which tend to be one or two years earlier than patent grant. Therefore, backward citations in first one or two lagged years (presumably the most important piece of information regarding the focal patent's technology stage or novelty) are often missing. Second, patent backward citations, unlike paper references that focus on completeness, are often parsimonious and incomplete. The average number of citations is usually less than 20 and many patents have no backward citations at all.

In this paper, aiming to evaluate a newly granted patent with no forward citations and incomplete (and often quite sparse) backward citation information, we develop a novel approach to assess value of a patent by exploring *technologically relevant prior patents* as a supplement to the backward citations. The idea behind our approach is simple yet innovative. Since the key in a backward citation-based patent evaluation approach is to find relevant prior patents of a focal patent via its backward citations, we shall also identify other good (or better) technologically relevant prior patents through other means. We test this idea by identifying a set of technologically relevant prior patents based on *content similarity*, and use them to construct features representing a focal patent's novelty and stage in the technology trajectory. Identification of relevant prior patents based on content similarity seems simple, yet it works well. This is exactly the beauty of our proposed approach: simple, intuitive and working well.

Moreover, to measure the novelty of a patent, its stage of technology trajectory, or other properties related to patent value, we focus on the dynamics of relevant prior patents over past years (i.e., in form of temporal distribution). We have studied a series of effective features exhibiting discriminative temporal behaviors from diverse aspects, and identified six temporal patterns and features. We propose two prediction models and evaluate the performance of the models using features from patent backward citations, relevant prior patents, or both together with baseline features. Our experimental results show that relevant prior patents complement patent backward citations, and significantly enhance the evaluation of patent value.

Our major contributions in this work are as follows:

- **Problem of Patent Evaluation on Newly Granted Patents**: We point out the limitation in using forward citations and backward citations to evaluate newly or recently issued patents. This is a very important and practical problem for technology management in firms.

- **New Patent Evaluation Approach**: Without relying on forward patent citation information, our patent evaluation approach utilizes a set of technologically relevant prior patents identified based on content similarity to supplement information derived from backward citations.

- **Features based on Temporal Trending of Relevant Prior Patents**: We propose six sets of features measuring trending and temporal patterns in multiple subsets of technologically relevant prior patents and backward citations, which are highly related to the value of a patent.

2 Related Work

The study on patent quality evaluation using patent mining techniques has received growing interests. Hasan et al. [5] proposed a patent ranking model, called COA (Claim Originality Analysis), which evaluated the value of a patent based on the novelty (recency) and impact (influence) of the important phrases in Claim section of a patent. They used the number of patent citations received, patent maintenance status, and their own confidential ratings to evaluate the proposed model with patents related to software technology or business process. Jin et al. [6] introduced a model recommending patent maintenance decision. They proposed diverse features, which could measure the value of a patent, and built a ranking model to predict whether a patent should be maintained or abandoned when a renewal decision was made. Liu et al. [7] proposed a latent graphical model to assess patent quality using quality-related features, such as citation quality, technology relevance, claim originality, etc. They used the number of citation received, court decisions (ruled as valid or invalid), and reexamination records as patent quality measurements. Even though the court decisions may be strongly correlated with patent quality, it is hard to get enough samples to train the model. Oh et al. [8] introduced a weighted citation method. They distinguished different types of citations to rank patents by their importance. Hu et al. [9] proposed a topic-based temporal mining approach to assess the novelty and influence of a patent to discover core patents. They extracted topics from a patent document and quantified a patent's novelty and influence by analyzing the topic activeness variations along the timeline. Compared to those previous works, our study is focused on augmenting the values from patent citations and relevant prior arts. Instead of extracting diverse features from patent documents or using text analysis, our features are mainly extracted from temporal patterns and trends of patent citations or relevant patents, which make the model simpler and easier to build.

3 Patent Value Evaluation Approach

In this section, we define the patent value evaluation problem, and present our approaches to solve this problem.

Research Goal: Let $D = \langle D_1, D_2, \cdots, D_N \rangle$ be a set of newly granted N patents, which have no forward citation yet. Our goal is to evaluate these new patents, using only information available when they are granted. As discussed earlier, patent evaluation may include predicting a patent's future value, comparing the value of patents in a given patent set, and identifying influential patents in a field or within a company.

Relevant Prior Patent Based Patent Evaluation: To evaluate the value of a newly granted patent, we use information on technologically relevant prior patents. In particular, we focus on the temporal patterns and trending of the relevant prior patents that reflect the stage of the technology trajectory and/or novelty of the focal patent. For example, if the focal patent is associated with a lot of technologically relevant prior patents in the years immediately prior to its grant, its technology was likely to be new and on the rise, which would in turn suggest the focal patent likely to be novel and valuable. Inspired by existing patent backward citation based patent evaluation [1, 2, 3, 4], our first proposal is to identify the set of relevant prior patents for the focal patent based on its backward citations. However, as we pointed out earlier, patent backward citations are often incomplete overall and seriously missing in the most recent 1-2 years before patent grant year. Alternatively, we propose to identify a comprehensive set of technologically relevant prior patents based on content similarity. We then construct features that reflect the temporal patterns and trends of this set of relevant prior patents, which should be related to the value of the focal patent.

Features of Temporal Trending: Another innovative idea in this paper is that we propose to exploit the trending and the stage of a focal patent among relevant patents. We aim to reveal the technology stage, novelty and value of the focal patent from different aspects. For example, if the focal patent is associated with a lot of recent technologically relevant prior patents that are assigned to the same firm (or invented by the same inventor), this might indicate that the firm (or the inventor) has become very interested in the field of the focal patent and devoted quite a lot of resources in R&D related to the focal patent. Therefore, the focal patent and its technology field might be important to the firm (or the inventor), thus more likely to be valuable.

Patent Evaluation Models: There is no public available gold standard or benchmarks that clearly define the value of a patent in monetary terms. We consider two patent evaluation cases: i) ranking a set of patents, ii) identifying top-ranked patents. We are particularly interested in ranking patents because the patent ranking can provide a relative comparison of patent values. In these two cases, we use patent forward citations and patent maintenance status as indicators of patent values. Corresponding to the two cases, we learn prediction models to evaluate patent values. The first model is a ranking model that ranks patents according to their values. Often firms are more interested in understanding the relative values among a set of patents, rather than predicting their absolute value. For instance, when a firm decides to renew only half of its patents, or two firms decides whether a deal of cross-licensing is worthwhile to pursue, all matters is the relative ranking of the patents involved. In this model, we use the number of forward citations that a patent receives in a long time window (e.g., 12 years after grant) as an indicator of the patent's true value. The second model identifies high-valued patents. To make their technology and patent management decisions, firms often need to know which are the most valuable inventions/patents in their patent portfolio. For example, given limited financial resource, a firm may only afford to maintain the top 10% of its patents. We learn a binary classifier to identify top 10% or 20% most valuable patents based on 12-year forward citations.

4 Feature Extraction

We conjecture that the temporal distribution or trending of technologically relevant prior patents, in combination of patent backward citations, is highly related to the value of a patent. Moreover, the subsets of relevant prior patents in various relations to the focal patent might reflect patent value from different angles. In this section, we introduce the feature sets describing the temporal patterns in technologically relevant prior patents and backward citations of a newly granted patent in a variety of angles.

4.1 Feature Sources and Types

We describe main feature sources for the patent evaluation models in this work as follows: Given a patent, D_i, $\{Q_i, C_i, R_i\}$ denotes the three main feature sources, where $Q_i = \langle f_1, f_2, \cdots, f_l \rangle$ is l patent characteristic features, $C_i = \langle p_1, p_2, \cdots, p_n \rangle$ is n patent backward citations, and $R_i = \langle p_1, p_2, \cdots, p_m \rangle$ is m technologically relevant prior patents. First, we choose some features capturing the characteristics of a patent as our baseline features. These patent characteristics, including the number of claims, figures and assignees, are often used in previous studies. We also include binary indicator variables to identify the technology fields of the focal patent. These variables help to control the variance on patent value among patents in different technology fields. The patent characteristic features that we use are the number of claims, figures, inventors, assignees, foreign references, other references, USPC codes, and IPC codes.

In addition to these baseline features, we propose a variety of features based on temporal distribution pattern and trending of the subsets of relevant prior patents, such as the same assignee set, the same inventor set, or the same technological class set, etc. These features hopefully reflect or are related to the value of a focal patent, and would be used in our three patent evaluation models. The temporal distribution of relevant prior patents is defined as $T(D_i) = f_Y = \langle f_{Y_1}, f_{Y_2}, \cdots, f_{Y_n} \rangle$, where f_{Y_1} is the number (frequency) of prior patents in the set of retrieved relevant prior patents that were granted in the first year prior to the patent D_i, f_{Y_2} is the number of relevant prior patents granted in the second year prior to the patent D_i, and so on. We call the citing year gap, $Y_1, Y_2, \cdots$, as "citation lag" in this paper.

4.2 Features for Temporal Patterns

Our approach is to capture the attractiveness of the technology and the novelty of a new patent using temporal trends in the sets of relevant prior patents and backward citations. Moreover, we look at temporal patterns in the subsets of relevant prior patents or backward citations that share the same assignees, the same inventors, or the same technology class, with the focal patent. These temporal trends, revealed by the activities of inventors and assignees, or the popularity of the technology field in years prior to the grant of the focal patent, could contain further information about the novelty and the technology stage of the focal patent, from the perspectives of the assignees, inventors and technology area of the focal patent. In addition, prior patents owned by other assignees or in other technology fields might also be important to provide a comprehensive understanding of the technology position and the novelty of invention, which could be related to the value of the patent. Thus, we also construct features that characterize the temporal patterns and trending in relevant prior patents or backward citations, owned by other assignees, or in other technology fields. In total, we introduce six sets of features to characterize the temporal patterns and trending in relevant prior patents or backward citations from various perspectives.

We capture the temporal trends based on the number (frequency) of technologically relevant prior patents or backward citations in each of the 20 years prior to the grant of a new patent. The temporal patterns between patent citations and relevant prior patents tend to be quite different. For example, backward citations in most recent one or two years prior to the grant year of the focal patent are very few, due to the fact that most of backward citations are cited by applicants or examiners at the time of filing or shortly after the filing (during the process of prior art search), which is one or two years earlier than patent grant. By contrast, the number of relevant prior patents in the most recent years prior to the grant of the focal patent tends to be greatest, relative to those in other prior years. This is particularly true for patents with higher value. These patterns suggest that information about relevant prior patents, in particular those in most recent years, could be very useful for patent evaluation.

To capture a trend, we use Gaussian filters to measure diverse distribution patterns of relevant prior patents or backward citations according to their grant year. We use this Gaussian filtering technique to construct features capturing temporal trends. A feature, F_y, which characterizes the temporal trend of relevant prior patents (or backward citation) for the last N years regarding the y^{th} citation lag year, is $F_y = \sum_{i=1}^{N} f_{Y_i} e^{-\alpha(i-y)^2}, \alpha > 0$, where f_{Y_i} is the frequency of relevant prior patents (or backward citations) in the Y_i^{th} lag year relative to the grant year of the focal patent. Then, we can use K features, $F_1, F_2, \cdots, F_K$, to evaluate the value of a patent according to their temporal trends. The six sets of features for temporal patterns (trends) in technologically relevant prior patents (or backward citations) are defined as follows:

Temporal Distribution of Backward Citations and Relevant Prior Patents (C1): We first measure the novelty and technology stage of a newly granted patent using the temporal trend of all the technologically relevant prior patents (or backward citations). To capture those temporal patterns, we construct features based on Gaussian filters. In addition, we add the frequency for the most recent prior year (one year lag from the grant of the focal patent), expecting to capture patent value well, especially with those most recent relevant prior patents. Thus, the features for C1 are $C1(D_1) = \langle F_1, F_2, \cdots, F_K, f_{Y_1}, f_{Y_2}, f_{Y_3} \rangle$.

Temporal Distribution of Backward Citations to and Relevant Prior Patents in the Same Assignee (C2): This set of temporal trend features focus on the subsets of relevant prior patents and backward citations assigned to the same assignees as the focal patent. If there are many recent relevant prior patents filed by the same firm, this might suggest that the firm has a strong interest in research and innovation related to the focal patent, thus the focal patent might be important and valuable. However, the number of backward citations in most recent years prior to the grant of the focal patent is quite small, and cannot be used to evaluate patents. Thus again the information on relevant prior patents, in particular in the most recent prior years, could be very useful in patent evaluation. We construct the same features as defined in C1, but only based on temporal patterns in the subsets of relevant prior patents and backward citations that share the same assignees as the focal patent.

Temporal Distribution of Backward Citations to and Relevant Prior Patents in the Same Inventor (C3): We also look into the temporal patterns in relevant prior patents or backward citations filed by the same inventors of the focal patent. It is intuitive if there are a large number of recent relevant prior patents filed by the same inventors, this would suggest that the inventors have been devoted themselves to this line of research, which in turn suggest that the technology is important and valuable. Again here, the number of most recent backward citations that were invented by the same inventors are small and not useful to capture this intuition. However, we can use the recent relevant prior patents by the same inventors. We construct the same features as defined in C1, but only applied to the subsets of relevant prior patents and backward citations that share the same inventors as the focal patent.

Temporal Distribution of Backward Citations to and Relevant Prior Patents in the Same Technology Class (C4): The temporal patterns in relevant prior patents in the same technology field could gauge the popularity of the technology field of the focal patent. We construct the same features as defined in C1, but only based on temporal patterns in the subsets of relevant prior patents and backward citations that share the same technology class (i.e., the primary U.S. patent class) as the focal patent.

Temporal Distribution of Backward Citations to and Relevant Prior Patents in the Different Assignees (C5): The relation of a focal patent to relevant prior patents filed by other assignees (i.e., other firms) can be used to measure the attractiveness and thus the value of the patent because it reflects the interest in the associated technology by other firms. Thus, we use the assignee diversity in relevant prior patents as another feature set in predicting patent value. To capture the diversity of assignees in relevant prior patents, we use the *entropy* of the distribution in different assignees in relevant prior patents in each year prior to the grant of the focal patent. High entropy means that a focal patent is related to other previous inventions filed by a larger number of other firms. We use the entropy values in each prior year to construct this set of features. The features of C5 are defined as $C5(D_i) = \langle E_1, E_2, \cdots \rangle$, where E_y is the entropy in y^{th} citation lag year, $E_y = -\sum_{k=1}^{K} Pr(a_k) log(Pr(a_k))$, where a_k is the k^{th} assignee, K is the number of different assignees in relevant prior patents (or backward citations) for the last y years prior to the grant of the focal patent. $Pr(a_k)$ is the probability of the assignee a_k appeared in relevant prior patents (or backward citations) in that year.

Temporal Distribution of Backward Citations to and Relevant Prior Patents in the Different Classes (C6): We also look into the diversity in technology classes in relevant prior patents (or backward citations), i.e., how many different technology classes in relevant prior patents (or backward citations). This information could be useful in patent evaluation as it indicates how much the focal patent might be related to a diversity of technologies. However, the pattern in relevant prior patents is different from citations. We use the same method in C5 to construct this set of features, C6, to capture the temporal pattern in entropies in technology fields of either relevant prior patents or backward citations.

5 Experiments

We perform an empirical evaluation to validate our proposed ideas. In this section, we describe our experimental setup and discuss the results.

5.1 Experimental Setup

Data Set: We evaluate the proposed approaches using 4 million U.S. patent documents granted since 1980 until 2012. We use 14,000 patents granted on

January 2001 as our evaluation set because those patents are the most recent patents that have information about their 4th, 8th and 12th year renewal status and enough forward citations which we use as the indicator of the patent value. About 2 million patents granted between 1980 and 2000 are used as the pool of prior patents from which we retrieve technologically relevant prior patents for a given focal patent. Backward citations are also restricted to patents granted between 1980 through 2000. Patents granted between 2001 and 2012 are used to count the 12-year forward citations for the focal patent.

Content-Similarity Based Retrieval of Relevant Prior Patents: We build a prior patent retrieval/recommendation engine based on content similarity using information retrieval techniques. A patent document consists of multiple sections such as Title, Abstract, Claims, Description, etc. We extract search query terms from each section of a patent document. Xue and Croft [10] proposed a method to transform a query patent into a search query. Our approach to generate the best query terms is similar with their method. We use Indri [11] as our retrieval model. Given a search query, it returns ranked relevant prior patents with retrieval scores (relevance scores). Indri supports a weighted-term query. Xue and Croft showed a better performance when a log-scaled *TF*(Term Frequency) is used as a weight on a query term. We observed the same and thus use the a log-scaled *TF* as the weight on a query term.

Feature Set for Evaluation: We prepare four feature sets to evaluate our approaches. The first feature set (FS1) is the baseline features, involving only 8 patent characteristics features and 6 HJT-6 technological class indicators. The second feature set (FS2) uses temporal pattern and trend features extracted from backward citations, together with baseline features. FS1, FS2 is the benchmark, to which we compare our proposed approach using technologically relevant prior patents. The third feature set (FS3) uses temporal patterns and trending features extracted from relevant prior patents, together with FS1. The last feature set (FS4) combines temporal trend features based on both relevant prior patents and backward citations, together with FS1. Comparing FS4 to FS2, we can see the incremental improvement in performance from augmenting information on relevant prior patents to backward citations.

Evaluation Metrics: We use the Spearman's rank correlation coefficients to evaluate the patent ranking model. The performance of the top-ranked patent classification and the least valuable patent classification are evaluated by Precision, Recall, F-score and AUC (Area Under Curve).

5.2 Analysis of Experimental Results

We conduct experiments on the two patent evaluation tasks detailed in Section 3 for each feature set. Then, we investigate the relative significance of those six temporal pattern features we propose.

Table 1. Predicting Patent Ranks: * denotes a significant difference (p-value<0.05) from FS2. 2,3,5,10-class are divided with the same size. 5,10-class(FC) use floor(log(# of forward citations+1) as ranking values).

	FS1		FS2		FS3		FS4	
Rank	LR	SVR	LR	SVR	LR	SVR	LR	SVR
2-class	0.3095	0.3115	0.3711	0.3699	0.3593*	0.3875*	0.3974*	0.4112*
3-class	0.3494	0.3690	0.4176	0.4282	0.4075*	0.4418	0.4497*	0.4704*
5-class	0.3682	0.3890	0.4396	0.4522	0.4272*	0.4622	0.4718*	0.4948*
10-class	0.3759	0.3929	0.4495	0.4618	0.4375*	0.4709	0.4831*	0.5070*
5-class(FC)	0.3667	0.3878	0.4378	0.4504	0.4243*	0.4626*	0.4686*	0.4936*
10-class(FC)	0.3764	0.3988	0.4505	0.4651	0.4376*	0.4762*	0.4838*	0.5105*
# of FC	0.3795	0.4049	0.4534	0.4688	0.4407*	0.4794*	0.4870*	0.5121*

Predicting the Ranks of Patents: In this experiment, we evaluate the ranks of patents. Using patent ranks, we compare the relative value among patents or assess overall value for a patent portfolio. To test diverse ranking scenarios, we use several ranking approaches. We first divide the ordered evaluation set based on their 12-year forward citations (ground truth), into 2, 3, 5, or 10 classes of the same size. Thus, we prepare four types of the uniformly distributed ranked values. We also directly use the log-scaled number of forward citations. To build prediction models, we use two regression models, Linear Regression (LR) and Support Vector Regression (SVR). In all these cases the Spearman's rank correlation coefficients are used to evaluate the ranking performance. Table 1 show the performance of predicting patent ranks with diverse ranking values using LR and SVR. The results in Table 1 show that SVR is better than LR in overall performance. According to the SVR results, the feature set based on relevant prior patents (FS3) is better than that based on backward citations (FS2). Overall, the combined feature set (FS4) achieves the best performance which is significantly better than FS1 with more than 27% improvement in all these cases.

Identifying the Top-Ranked Patents: We conduct experiments to predict the top-ranked patents. We prepare two top-ranked patent data sets based on their forward citations. One is the top-10% ranked patents, and the other is the top-20% ranked patents. We build the patent evaluation model using two binary classifier, Random Forest (RF) and Support Vector Machine (SVM). Table 2 shows the results of classifying the top-10% ranked and top-20% ranked patents using RF and SVM. In both cases, the relevant patents (FS3) shows better performance than patent citations (FS2), and again, we can get the best results when we use both patent citations and relevant patents together (FS4). In general, SVM shows better results than RF. The overall results shows our approaches works for predicting the top-ranked patents.

Feature Analysis: Finally, we investigate the relative importance of the six temporal patterns and the extracted trending features discussed in Section 4.2. Table 3 shows the rank prediction results when we add or remove one of these six temporal pattern features in the models. According to the results, the temporal distribution of relevant prior patents and backward citations (C1) is the key factor among six

Table 2. Classification of Top-Ranked Patents: * denotes the significant difference (p-value<0.05) from FS2

(a) Top-10% ranked patent classification

RF					SVM				
features	Precision	Recall	F-Score	AUC	features	Precision	Recall	F-Score	AUC
FS1	0.1738	0.6410	0.2731	0.6462	FS1	0.1912	0.6541	0.2956	0.6683
FS2	0.2110	0.7186	0.3259	0.7057	FS2	0.2208	0.7167	0.3374	0.7135
FS3	0.2251*	0.7430*	0.3452*	0.7250*	FS3	0.2307	0.7730*	0.3551*	0.7387*
FS4	0.2354*	0.7613*	0.3920*	0.7389*	FS4	0.2426*	0.7725*	0.3691*	0.7482*

(b) Top-20% ranked patent classification

RF					SVM				
features	Precision	Recall	F-Score	AUC	features	Precision	Recall	F-Score	AUC
FS1	0.2976	0.6205	0.4018	0.6184	FS1	0.3292	0.6337	0.4329	0.6483
FS2	0.3440	0.6630	0.4526	0.6664	FS2	0.3627	0.6669	0.4695	0.6806
FS3	0.3608*	0.6864	0.4728*	0.6846*	FS3	0.3663	0.7024*	0.4813*	0.6927*
FS4	0.3753*	0.7167*	0.4925*	0.7027*	FS4	0.3819*	0.7110*	0.4966*	0.7055*

Table 3. Feature Analysis: * denotes a significant difference (p-value<0.05) from the baseline

	(a) Adding features			(b) Removing features		
Featues	FS2	FS3	FS4	FS2	FS3	FS4
Base	0.4128	0.4128	0.4128	0.4721	0.4908	0.5207
C1	17.83%*	15.06%*	25.33%*	-0.99%	-3.00%*	-2.21%*
C2	1.16%	2.67%*	3.81%*	0.21%	0.01%	0.31%
C3	0.73%	1.37%	1.90%	0.13%	-0.18%	-0.09%
C4	10.02%*	6.52%*	12.57%*	-0.04%	-0.67%	-0.11%
C5	14.08%*	9.62%*	19.30%*	-0.06%	-1.66%*	-1.35%*
C6	12.58%*	10.21%*	18.40%*	0.04%	-0.50%	-0.07%

temporal patterns. It shows the significant improvement (or decrease) when we add into (or remove it from) the feature sets. Other important features are the measure of assignee or technological class diversity in relevant prior patents or backward citations (C5 and C6). When we add those features, the performance is significantly improved. C2 and C3 do not show the significance, possibly because the number of relevant prior patents and backward citations assigned to the same assignee or invented by the same inventors is too small to reflect useful information regarding patent value.

6 Conclusion

In this study, we propose a novel approach to evaluate newly granted patents, using only information available at the time of patent grant. Our approach is to use the temporal patterns and trends of relevant prior patents that reflect the novelty and technology state of a patent or the attractiveness of the technology associated with a patent. The experimental results show that our approach can achieve significantly better evaluation performance by augmenting information on relevant prior patents to backward citations. The feature analysis results show that temporal patterns based on the distribution of relevant prior patents and backward citations are important features in our proposed patent evaluation models. Moreover, the measures on the dynamics regarding the diversity of other assignees and other technology classes in relevant prior patents and backward citations are highly related to patent value. Our approach based on temporal trends of relevant prior patents is the first of this kind, as an effort to assess the value of a new granted patent. Compared to previous works relying on rigorous feature extraction or text analysis, our approach is much simpler and easier to build patent evaluation models, and quite flexible to expand.

Acknowledgement. This work is supported by U.S. National Science Foundation Grant 5MA-1064194.

References

[1] Trajtenberg, M.: A penny for your quotes: patent citations and the value of innovations. The Rand Journal of Economics, 172–187 (1990)
[2] Harhoff, D., Narin, F., Scherer, F.M., Vopel, K.: Citation frequency and the value of patented inventions. Review of Economics and Statistics 81(3), 511–515 (1999)
[3] Hall, B.H., Jaffe, A.B., Trajtenberg, M.: The nber patent citation data file: Lessons, insights and methodological tools. Technical report, National Bureau of Economic Research (2001)
[4] Hall, B.H., Jaffe, A., Trajtenberg, M.: Market value and patent citations. RAND Journal of Economics, 16–38 (2005)
[5] Hasan, M.A., Spangler, W.S., Griffin, T., Alba, A.: Coa: Finding novel patents through text analysis. In: CIKM, pp. 1175–1184. ACM (2009)
[6] Jin, X., Spangler, S., Chen, Y., Cai, K., Ma, R., Zhang, L., Wu, X., Han, J.: Patent maintenance recommendation with patent information network model. In: ICDM, pp. 280–289. IEEE (2011)
[7] Liu, Y., Hseuh, P.Y., Lawrence, R., Meliksetian, S., Perlich, C., Veen, A.: Latent graphical models for quantifying and predicting patent quality. In: Proceedings of the 17th ACM KDD. ACM (2011)
[8] Oh, S., Lei, Z., Mitra, P., Yen, J.: Evaluating and ranking patents using weighted citations. In: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM (2012)
[9] Hu, P., Huang, M., Xu, P., Li, W., Usadi, A.K., Zhu, X.: Finding nuggets in ip portfolios: core patent mining through textual temporal analysis. In: Proceedings of the 21st ACM CIKM. ACM (2012)
[10] Xue, X., Croft, W.B.: Automatic query generation for patent search. In: Proceedings of the 18th ACM CIKM, pp. 2037–2040. ACM (2009)
[11] Strohman, T., Metzler, D., Turtle, H., Croft, W.B.: Indri: A language model-based search engine for complex queries. In: Proceedings of the International Conference on Intelligent Analysis, vol. 2(6) (2005)