Exploring Technological Trends for Patent Evaluation

Shuting Wang[1] Wang-Chien Lee[1] Zhen Lei[2] Xianliang Zhang[1] Yu-Hsuan Kuo[1]

[1] Department of Computer Science and Engineering, Pennsylvania State University
[2] Department of Energy and Mineral Engineering, Pennsylvania State University
sxw327@cse.psu.edu,wlee@cse.psu.edu, zlei@psu.edu, xxz149@cse.psu.edu, yzk5145@cse.psu.edu

***Abstract*—Patents are very important intangible assets that protect firm technologies and maintain market competitiveness. Thus, patent evaluation is critical for firm business strategy and innovation management. Currently patent evaluation mostly relies on some meta information of patents, such as number of forward/backward citations and number of claims. In this paper, we propose to identify *patent technological trends*, which carries information about technology evolution and trajectories among patents, to enable more effective and precise patent evaluation. We explore features to capture both the value of trends and the quality of patents within a trend, and perform patent evaluation to validate the extracted trends and features using patents in the United States Patent and Trademark Office (*USPTO*) dataset. Experimental results demonstrate that the identified technological trends are able to capture patent value precisely. With the proposed trend related features extracted from our identified trends, we can improve patent evaluation performance significantly over the baseline using conventional features.**

I. INTRODUCTION

The knowledge economy and the resulting increase in demand for innovation have led to more and more patents issued each year. It is of growing importance for firms to develop effective patent valuation methods to distinguish valuable patents from those less valuable, both among patents in their own patent portfolios and among patents in their fields.

In this paper, we argue that *patent technological trends* are important for patent evaluation because a patent technological trend carries important information about knowledge transmission and technology evolution in patents. We propose to extract and analyze patent trends to facilitate patent evaluation. Different from the conventional understanding of a trend, for the first time in the literature, we consider a trend as a set of patents that share similar attributes over a period of time. The attributes shared by the set of patents can either be explicit, e.g., U.S. class, assignee and inventor, or something implicit, such as technological ideas or business themes of patents.

In particular, we focus on extracting patent technological trends (hereafter *patent trends*). We define a patent trend as a set of patents that share certain technological *keywords* or *topics*, i.e., they are related to a common or similar technology. We hypothesize that patent trends could be an important factor for patent evaluation, because a patent trend carries useful information about the importance of the technology in general and the importance of a particular patent within the trajectory of the technology. For example, if patents in a technological trend are generally very valuable, it suggests the attractiveness and market value of the technology represented by the trend. Thus, any individual patent belonging to the trend could be more valuable than patents in other trends (e.g., patents related to smart phones). Moreover, the position of a patent in a patent trend might indicate the novelty and importance of the focal patent in the evolution and development of the technology. A patent that emerges early in the technological trajectory of a trend, i.e., this invention does not have many related prior works, suggests that the patent might be one of the founding and crucial inventions within the trend and thus could be very valuable. By contrast, a patent that lies on the tail of a technological trend and has many prior related inventions is likely to be incremental and thus less significant and valuable.

An immediate challenge is how to identify and extract patent trends among patents. To address the issue, we investigate two methods, *keyword-based* and *topic-based*, to identify important technologies represented by patents and subsequently extract technological patent trends based on these technologies. Basically, the keyword-based method assumes that each patent contains a set of discriminative keywords that represent the patented innovation. And we aims to identify those keywords considered as important (or popular), based on which we extract patent trends. On the other hand, the topic-based method assumes that each patent will be represented as a distribution of technological topics, each of which is a distribution of words. Therefore, topic-based patent trends consist of patents which are topically similar to each other. In this paper, we use these two methods to identify patent trends for comparison.

In order to demonstrate the usefulness of the extracted patent (technological) trends, we apply the trends to patent evaluation. As mentioned above, patent trends carry information that can reveal patent value from two aspects: (i) the importance, value, and attractiveness of the trend, and (ii) the importance and the role of a focal patent within the trend. Accordingly, we construct two sets of features from extracted patent trends: one includes features of the trend (hereafter *trend features*), and the other involves features about the patent within the trend (hereafter *within-trend patent features*).

To investigate whether our proposed methods can improve patent evaluation, we use the records of *patent renewals* as a proxy to assess patent values, due to the lack of data on patent monetary value. Accordingly, we formulate patent evaluation as a patent renew classification problem. In the United States, a patent needs to be renewed every four years after it is granted, in order to keep the patent right alive. In other words, assignees need to make renewal decisions at the 4^{th}, the 8^{th}, and the 12^{th} year after patent grant, with increasing renewal fees. Since patent owners decide whether their patents are to be maintained or abandoned based on their assessment of the value of the patents, the maintenance status reflects patent value.

Our results show that using information contained in patent trends significantly improves patent evaluation over conventional approach that use number of patent claims, number of patent inventors, U.S. class, and so on, as features. The extracted trend features are shown to be more important in predicting patent value, relative to within-trend patent features (that characterize the novelty and importance of a focal patent within the trend), though both sets of features are very helpful in patent valuation. This suggests that identifying the technological trends that a patent belongs to is important to understand the value of the patent.

To the best of the authors' knowledge, this work represents the first attempt to consider a patent trend as a set of similar patents. Moreover, we identify different features based on patent trends and apply them to patent evaluation. In summary, our work has made the following major contributions:

- This study first considers a patent trend as a set of patents and extracts patent technological trends using different methods (topic-based method and keyword-based methods).
- We identify important features from trend aspect (trend features) and patent aspect (within-trend patent features) which capture both the quality of the trend and the value of the patents with respect to the trend.
- Using patent renewals as an indicator of patent value, our experiments show that considering both the trend features and within-trend patent features can significantly improve patent evaluation, compared to the current practice which only involves the meta information such as number of claims and number of authors in patent evaluation.

The rest of the paper is organized as follows. We first describe the patent evaluation problem as well as our research goal and introduce related concepts in Section II. Then we propose different methodologies to extract patent technological trend and calculate defined features in Section III. Next, we present our experimental results and show some insights learned from the results in Section IV. Finally, we review the related work in Section V and draw conclusions in Section VI.

II. Patent Evaluation Using Patent Trends

In this section, we first describe the problem of patent evaluation. Next, we describe a trend-based framework proposed for patent evaluation. Finally, we define the notion of patent trends and discuss the features of trends (including both trend features and with-the-trend patent features) for patent evaluation.

A. Patent Evaluation Problem and Research Goal

In order to identify valuable patents in a field or a firm, a patent evaluation task aims to predict the value of a focal patent using a corpus of patents P which contains all the previously issued patents available at the time of prediction.

As discussed earlier, we argue that patent trends, which a patent takes part in, carry useful information about the market and technological potentials of the technology, and thus are useful for patent evaluation of the focal patent. In this work, we aims to demonstrate the information extracted from patent trends can complement the conventional patent information (such as number of claims, patent inventor/assignee information, number of citations, etc) to improve patent evaluation.

To proceed, we formulate patent evaluation as a classification problem, which predicts whether the given patent is worth maintaining after it is granted for n years. As there is no public available information on the monetary value of a patent explicitly, we use patent maintenance status as a proxy indicator of patent values.

Notice that, after December 12, 1980, US patent holders have to pay maintenance fees for granted patents. Thus, the patent holders need to decide whether to pay the maintenance fees to renew and maintain their patent rights at the 4^{th}, 8^{th} and 12^{th} years after patent grant. As the maintenance fees increase for later renewals, a patent that is renewed at both the 4^{th} and the 8^{th} years is more valuable than one that has only a 4^{th} year renewal. We assume that the patent owners have insights about the true value of their patents and therefore that the patent maintenance status reflects patent values. Formally, we define the time period from a patent's grant date to its abandon/expiration date as its *Active Period (AP)*. Given a patent p, $AP(p)$ is either 4 years, 8 years, 12 years or more, depending on how many times it gets renewed by the patent owner. Therefore, for a given patent p, we aims to learn three binary classifiers, C_4, C_8 and C_{12}, as follows to predict patent values.

$$C_n(p) = \begin{cases} 1 & if\ AP(p)\ > n \\ 0 & otherwise \end{cases} \tag{1}$$

where $n = 4, 8$ or 12.

Among them, C_4 distinguishes the least valuable patents (those not renewed at the fourth year), C_8 distinguishes patents of above-average value from those of below-average value, and C_{12} identifies the most valuable patents (those being maintained always).

B. Trend Based Framework for Patent Evaluation

In this work, we propose a new framework that exploits the trends of a patent, complementing the inherent patent features explored in previous studies (e.g., the number of claims, the number of inventors, etc), to implement the classifiers for patent evaluation. Fig. 1 shows an overview of the trend-based framework, which consists of three modules: (i) trend identification, (ii) feature extraction; and (iii) patent evaluation.

As shown at the right-hand side of Fig. 1, this framework first trains our classifiers using a large-scale *USPTO* patent data set. Then, as shown at the left-hand side of Fig. 1, given

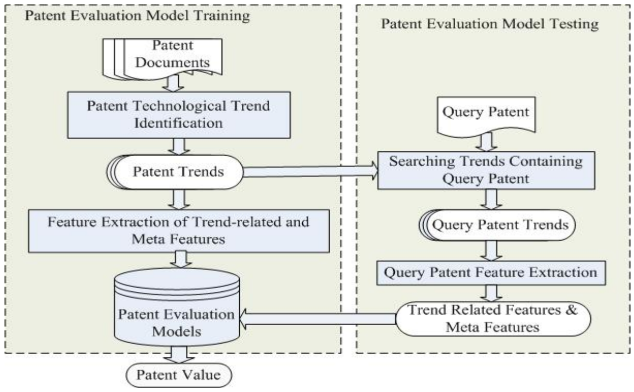


Fig. 1: An overview of trend-based framework

a query patent, the framework extracts its features for the classifiers to make prediction.

Specifically, the trend identification module identifies a set of technologically important trends based on the patent content. Next, for each patent in the data set, the feature extraction module extracts (a) conventional meta features directly from the patent, and (b) trend related features, i.e., trend features and within-trend patent features, from the identified important trends. Finally, the patent evaluation module use the extracted features (a) and (b) of each patent to train the classifiers for patent evaluation.

During evaluation time, upon arrival of a query patent Q, the framework first searches the trends of patent Q from the previously identified trends, then extracts the features (a) and (b) of the query patent Q using the feature extraction module. Finally, feeding these features to the classifiers to decide whether Q is a least valuable patent, or a most valuable patent, or whether it is above or below average.

C. Technological Patent Trends and Trend-Related Features

Here we formally define the notion of a *patent technological trend*. Additionally, we define a set of trend features and a set of within-trend patent features, based on the notion of patent technological trends.

We consider a technological trend as a set of patents that share a common or similar attribute, e.g., a specific technological topic or the same assignee. We formally define a trend as a series of patents of an attribute value as follows.

Definition 1. Patent Technological Trend. *Given a patent corpus P, The trend associated with the attribute α is a sequence of patents $p_t \in P$ in ascending order of their filing date t.*

$$Trend(\alpha) = \{p_{t_1}, p_{t_2}, ..., p_{t_n} | p_{t_i} \in P \wedge p_{t_i} \textit{has the attribute } \alpha\} \tag{2}$$

Generally speaking, patent technological trends are related to patent value in two aspects: (i) the popularity of the trend which a patent is in. It is probable for a patent to have higher value if it is in a popular trend, i.e., the idea or technique it is using is very hot and lots of opportunities emerge in this field. Therefore, we can extract *trend features* to represent the popularity of trend. (ii) The novel and impact of the patent within this trend. Patent novelty and impact in a trend can reveal the value of this patent compared to other patents in the same trend. A patent of high novelty and impact usually has higher value. Therefore, we extract *within-trend patent features* to reveal patent value.

Trend Features. We first extract some trend features, which characterize a patent trend as a whole instead of focusing on an individual patent, as follows.

- **Number of i^{th} year expiration.** We define four features , which are the number of patents in trend T that expire at the i^{th} year (where $i = 4, 8, 12$ and 16), respectively. These features aim to measure the number of patents with varied values in the trend. For example, the number of 4^{th} *year expiration* represents the number of patents that are least valuable in a trend and the number of 16^{th} *year expiration* represents the number of patents that have the highest value in a trend.
- **Ratio of i^{th} year expiration.** In addition to the absolute number of patents with varied values, we are also interested in measuring their respective percentage in the trend. Such distributions of patents with varied values reveal the market value and attractiveness of the trend overall. Formally, we define three features, which are the ratios of the number of patents in trend T that expire at the i^{th} year (where $i = 4, 8, 12$ and 16), respectively, to the total number of patents in the trend.
- **Trend size.** We construct a feature of Trend size, i.e., the number of patent in the trend, to measure the popularity and attractiveness of the trend. Given a patent technological trend T, the *size* of T is the number of patents in T as defined below.

$$Size(T) = |T| \tag{3}$$

- **Weighted trend value.** To better capture the business (or financial) interests and potential in the trend, we weigh the patents in a trend based on their patent renewals that reflect the patent value. We define *Weighted trend value* for a trend T as the number of patents in the trend weighted based on their number of renewals they have (including application and renewals). Notice that patents can be classified into four categories based on its expiration status: (i) expired at 4^{th} year; (ii) expired at 8^{th} year; (iii) expired at 12^{th} year; and (iv) expires at 16^{th} year, with the weights being one, two, three and four times renewals, respectively. Formally, we define Weighted trend value (Weighted) for a trend T as follows.

$$Weighted(T) = \sum \frac{n}{4} * |\{p_t : p_t \; expires \; at \; n^{th} \; year\}| \tag{4}$$

where $n = 4, 8, 12$ and 16.
The trend size and weighted trend value features intuitively represent the *popularity* and *attractiveness* of a patent trend, indicating the interests of firm and market in the underlying technology of the trend.
- **Average patent value of trend.** In addition to the weighted trend value, we also define the *Average patent*

value of trend as the (overall) weighted trend value divided by the number of patents in the trend T. This feature reflects the average market value of the technology underlying the trend. Formally, we have Average value trend (Average) for a trend T as follows.

$$Average(T) = Weighted(T)/|T| \tag{5}$$

Within-trend patent features. While the trend features try to characterize a trend, the *within-trend patent features* try to characterize a focal patent in the context of the trend. We are particularly interested in the novelty and the position of a query patent in its trends.

- **Novelty in Trend.** Given a patent technological trend T, the novelty (Nov) of a patent p_t in the trend T is measured by the number of patents in T with filing dates earlier than that of p_t. Formally, we have:
$$Nov(p_t, T) = |\{p|p \in T \wedge fdate(p) < fdate(p_t)\}| \tag{6}$$
where *fdate(·)* returns the filing date of a patent.
- **Position in Trend.** We further define the *position* of the patent relative to other patents in the trend. Given a patent technological trend T, the position (Pos) of a patent $p_t \in T$ is measured by the ratio of the number of prior patents of p_t to the total number of patents in T. Formally, we have:
$$Pos(p_t, T) = \frac{Nov\{p_t, T\}}{|T|} \tag{7}$$

As discussed, a novel patent, which does not have many prior inventions, could be one of the pioneering patents in the technology and thus very valuable. Moreover, the position of a patent in a trend could reflect the novelty and the impact of the patent in the technology. A patent with an *early position* in a *large-size* trend suggests that it has a lot of followup patents, and thus has a lot of impact on the trend. In other words, the *impact* of a patent on a trend is implicitly captured by our features (i.e., with position of the patent and trend size).

III. Extraction of Technological Trends

In this section, we describe our approaches to extract patent technological trends. Basically, we define a set of patents over a certain period of time that share some similar technological attribute as a patent technological trend. An immediate challenge is how to identify the attributes which are used to gather a set of similar or relevant patents as a trend. An intuitive idea is to aggregate patents that describe the same or similar technologies in their technical description/content. Therefore, we propose to investigate two different text mining methods to extract patent trends: (1) keyword-based method, and (2) topic-based method. Once the primary (important) trends are identified from the corpus of patents, we further extract trend features and within-trend patent features from the primary trends of a focal patent.

A. Keyword-Based Patent Trends

Patents with similar technical content (i.e., with high technological relevancy or in the same field) often use the same technological vocabulary (keywords) in their technological description. For example, the word "amino acid" is seen as a distinct keyword and all the patents that contain this keyword "amino acid" imply strongly that they are highly relevant to each other. Therefore, we aim to identify a set of important keywords W and use them to build the corresponding trends by grouping all patents based on the words in W. In the previous example, a trend corresponding to "amino acid" is formed by a set of patents p which contains the word.

While the idea looks simple, an issue is how to extract the set of important keywords. In the keyword-based method, we adopt *term frequency-inverse document frequency (tf-idf)* to reflect the importance of a word to a document in a corpus. As words with high *tf-idf* value in a patent p are considered as discriminative/important for representing the patent, we select *top-k* keywords based on the *tf-idf* score as the *discriminative keywords set* $D(p) = \{w_1, w_2, ..., w_k\}$ where *tf-idf*$(w_1) \geq$ *tf-idf*$(w_2) \geq ... \geq$ *tf-idf*(w_k) such that $\forall w \notin D(p)$, *tf-idf*$(w_k) \geq$*tf-idf*(w). In order to identify the *top-n* most interested keywords among all patents, we then count, for each keyword appearing in the union of discriminative word sets, its *document frequency*, i.e., the frequency it appears as one of the *top-k* keywords in patents. We then rank all the keywords based on their document frequency and select the *top-n* key words as the *trend keyword set*. For example, we first select *top-50* keywords of each patent as its discriminative words set and then rank all the keywords based on their frequency of being in the *top-50* keywords among all patents. For example, if the keyword "device" appears in the *top-k* keywords of 1000 patents, then it has a document frequency of 1000. On the other hand, the keyword "communication" may appear in the *top-k* keywords of 2000 patents. In this case, "communication" ranks higher than "device". Accordingly, we can identify the *top-n*, e.g., 100, keywords to form the trend keyword set.

After identifying the *top-n* trend keyword set, we then extract trends corresponding to each trend keyword. Each trend, i.e., corresponding to a keyword in this case, contains patents which have this keyword as one of its *top-k* technological keywords. As a result, we presume that each patent belongs to at most k technological trends. In summary, each technological trend T consists of a sequence of patents containing the same trend keyword (as the trend attribute), which are ordered by their filing dates.

Since one patent can belong to multiple trends in this keyword-based method, we can obtain a set of trend features and within-trend patent features for each patent based on each trend. In our experiment, for simplicity, use the average of the features obtained from multiple trends. For example, patent A belongs to 20 technological trend and its weighted trend value is the average of the weighted trend value obtained from these 20 trends.

B. Topic-Based Patent Technological Trends

The keyword based approach is straightforward and easy to understand. However, simply using keyword as a technological trend may have some problems. As technical documents with legal significance that can earn potential profits, patents tend to have complex structures and special nomenclature. Moreover, some technical terms may have more than one aliases that are semantically identical to each other. In order to capture ideas in a patent document more precisely, we use topics, which are mixtures of words, to represent patent technological trends.

To better extract patent topics, similar to the keyword-based method, we perform document stemming and remove stop words. We then use *tf-idf* as a measurement to obtain a set of words for each patent document. Next, we apply *Latent Dirichlet Allocation (LDA)* to discover the latent topics hidden in the patent content. Let W be the vocabulary of m unique words in our patent data set P. Suppose there are l topics $z_1, z_2, ...z_l$ discovered from P using LDA. A topic z_i $(i = 1..l)$ is a probabilistic distribution over words in W, i.e., $\{p(w|z_i)\}$ where $w \in W$. A patent $p \in P$ therefore is represented as a probabilistic distribution of the latent topics, i.e., $\{p(z_i|p)\}$ where $i = 1..l$. As such, we store each topic as a distribution over words in the vocabulary and each patent document as a mixture of topics.

After performing topic discovery on patent documents, we then gather similar patents in order to extract technological trend. To proceed, we use a clustering algorithm to partition content-similar patents in our patent data set P into clusters. Based on the topic distribution of each patent p, we choose the well-known *k-means* algorithm to cluster patents based on the distribution weights over topics as the coordinate to place patents in a multi-dimensional space. The distance between two patent documents is measured by the cosine similarity of their topic distributions. As such, patents in the same cluster C are considered to be in the same technological trend T (with the centroid of the cluster as the trend attribute). Based on this topic-based definition of a trend, a cluster of topic-similar patents is prepared as a trend by sorting in ascending order based on their filing dates.

IV. Evaluation

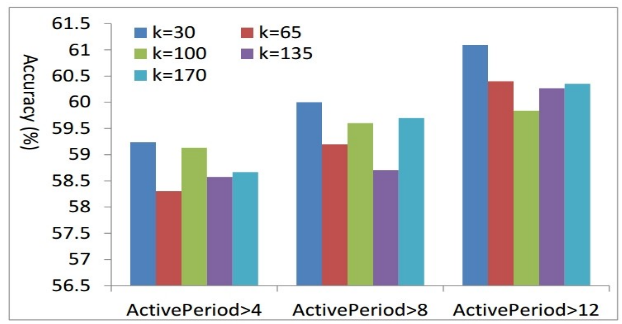
Our research goal in this work is to explore the usefulness of patent technological trends in patent evaluation. To validate our idea and evaluate the effectiveness of our proposal, we investigate whether the identified trend features and within-trend patent features can improve the performance of the patent value classifiers (as shown in Equation (1) in Section 2) over the conventional features extracted from patent meta information.[1] In this section, we introduce the experiment setup and then tune parameters for trend extraction algorithm. Finally, we present the prediction performance and discuss the feature importance.

A. Experiment Setup

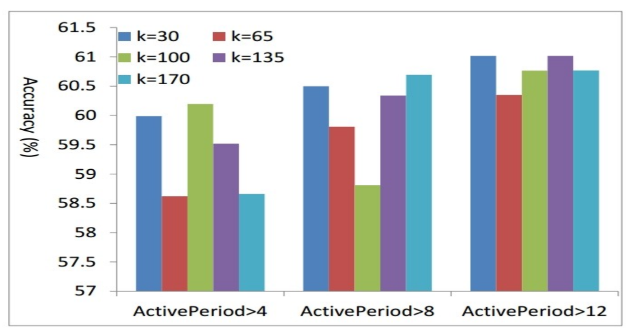
We extract and identify patent technological trends from a dataset consists of all patents in the field of Medical and Drug, issued by USPTO between 1981 and 2006. Based on the identified patent trends, we derive the proposed trend features and the within-trend patent features for each patent. We also use as the benchmark a number of baseline features, including number of patent claims, number of patent independent claims, number of patent inventors and patent U.S. class number, which have been shown to be effective for patent evaluation. We use these various features and do training and testing on the patents issued in year 2000 for the three classifiers (which predict whether the patent is renewed at 4^{th}, 8^{th} and 12^{th} year, respectively). We use *Support Vector Machine (SVM)* to build up our classifiers and perform a 5-fold cross validation on the data. The prediction performance is based on the average accuracy of the binary classifiers for testing data in the 5-fold cross validation.

B. Parameter Tuning

We first empirically tune the parameter configuration for both keyword-based and topic-based methods. As discussed in Section III, the number of trends (hereafter k) in the patent corpus is a key factor for identifying the trends (in the keyword-based method, it is the number of keywords of patents and in the topic-based method, it is the number of clusters in patents). Thus, in this section, we test different values of k by varying it from 30 to 175 with step size 35. Basically, we perform parameter tuning for keyword-based and topic-based feature sets, which include both trend features and within-trend patent features, respectively. Figure 2 shows the experimental results for keyword-based and topic-based approaches. It can be seen that for the keyword-based method, identifying 30 keyword-based trends (see Figure 2(a)) achieves the best performance for patent evaluation, while for the topic-based method, the optimal setting for k is also 30 (see Figure 2(b)). Accordingly, for the rest of experiments, we set the number of trends for the keyword-based method and topic-based method both to be 30, correspondingly.



(a) Keyword-based feature set



(b) Topic-based feature set

Fig. 2: Prediction accuracy (%) under varied numbers of trends (k)

[1]Again, due to the lack of ground truth, the classifiers are based on patent renewal status, a proxy for patent value.

C. Prediction Results

As mentioned above, we build up three classifiers, C_4, C_8 and C_{12}, to predict whether a patent remains active for more than 4 years, 8 years and 12 years, respectively. Figure 3 shows the prediction performance of the classifiers. In Figure IV-C, 'Keyword(Within)', 'Keyword(Trend)' and 'Keyword(Within&Trend)' refer to within-trend patent features, trend features and the combined features extracted from the keyword approach, respectively. Similar labels are also used for features extracted from the topic-based approach. These trend-related features are added to the baseline feature set for experimentation to observe their improvement over the baseline features.

As shown in the figure, incorporating our proposed features into classifier does achieve significantly higher accuracy in predicting patent value than using only the baseline feature set. These results suggest that: (i) Patent value is highly correlated with the popularity and market value of the trend it is in. A patent from a more popular and valuable trend tends to be of higher value and thus more likely to be maintained by its patent holder. (ii) Patent novelty and relative position in the technological trajectory of a patent trend are also useful for patent evaluation. These within-trend patent features suggest the novelty and impact of a patent in the trend and thus can reveal the inherent value of a patent. Basically, our features capture these implicit useful information that is imbedded in patent trends and thus improve the prediction performance.

Moreover, for the prediction performance of the three binary classifiers, we also observe in Figure 3 that the proposed "trend features" achieve a better overall performance than the "within-trend patent features" do. While the relative importance and value of a patent within a trend (which consists of the patents peers/competitors) indicate the novelty and impact of the patent, which is useful for patent evaluation, the popularity and the value of the trend itself is even more critical.

Furthermore, the topic-based feature set generally performs better than the keyword-based feature set (which can be observed by comparing 'Baseline+Keyword(X)' and the corresponding 'Baseline+Topic(X)', where X is 'Trend', 'Within', or 'Trend&Within'). This may be because that a topic is a better representative for patent technology than keyword and thus topic-based method achieves a better performance. Not surprisingly, we achieve the best performance when both the trend features and the within-trend patent features from both keyword and topic-based methods (i.e., 'Baseline+All') are used.

For the performance of the three classifiers, i.e., C_4, C_8 and C_{12}, which predict whether a patent remains active for more than 4 years, 8 years and 12 years, respectively, basically, C_4 predicts patents of the lowest value while C_{12} predicts patents of the highest value, we observe that C_{12} has a prediction accuracy generally better than C_8 which in turn is better than C_4.

D. Analysis of Feature Importance

To better understand the importance of various features used for our experiments, we measure their F-score [1], which is widely used in feature selection. Each feature is divided into groups based on the predictive label and F-score for this feature is the ratio of the variability between groups and the variability within each group. We report the *top-5* features in Table I for three categories of experiments: (i) Features from keyword-based trends+Baseline, (ii) Features from topic-based trends+Baseline, and (iii) All features – including keyword-based trends, topic-based trends and baseline. The features from trends under examination include all the trend features and within-trend patent features. It's worthy noting that the *top-5* features in all categories are all features from trends. Again, we observe that trend features are of more importance than within-trend patent features, for either the keyword-based approach or the topic-based approach. Moreover, among trend features, those related to the distribution in the value of the patents in a trend (i.e., those ratio-based features such as ratio/number of i^{th} year expiration) generally appear to be more important in predicting patent value than number based trend features. This makes sense as ratio-based trend features are more likely to reflect the overall market value and attractiveness of the trend, relative to number-based trend features. For example, suppose a patent A is in a trend of 10000 patents while another patent B is in a trend of only size 100. Also assume that for the trend of patent A, 1000 patents get renewed three times, while for the trend of patent B, 50 patents get renewed three times. If we only consider the numbers of patents that got renewed, the trend of patent A seems to be of higher quality than the trend of patent B. However, the trend of patent B has one half getting renewed at the 12^{th} year while patent A's trend only has 10%. It is thus more likely that patents in the former trend of patent B would be more valuable.

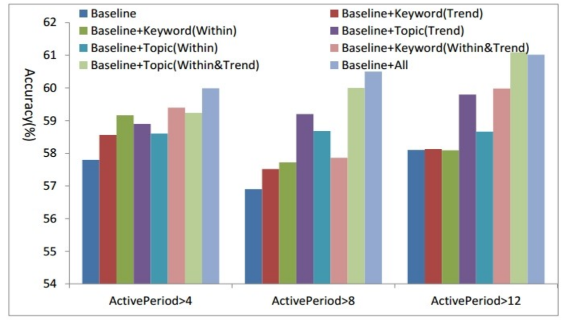


Fig. 3: Prediction accuracy of patent evaluation classifiers with various features.

Moreover, we observe that features based on trends using the topic-based approach dominate those extracted using the keyword-based approach, in terms of the top five most important features (see the *top-5* features in 'All Features' in Table I). As we argue in Section III, a topic is related to a concept represented as a mixture of words. Thus, it may be able to capture rich semantics in patents. Consequently, a topic is likely to represent the ideas/technology in a technological trend more precisely than a keyword does. This result is also consistent with the results in Figure 3, i.e., features based on trends extracted using the topic approach lead to a better performance than features based on trends extracted using the keyword approach.

Category	*ActivePeriod* > 4	*ActivePeriod* > 8	*ActivePeriod* > 12
Keyword-based Trends + Baseline Features	position ratio of 4^{th} year exp. number of 16^{th} year exp. weighted trend value number of 8^{th} year exp.	ratio of 4^{th} year exp. average patent value of trend position ratio of 12^{th} year exp. weighted trend value	ratio of 4^{th} year exp. average patent value of trend ratio of 8^{th} year exp. ratio of 16^{th} year exp. number of 12^{th} year exp.
Topic-based Trends + Baseline Features	ratio of 4^{th} year exp. ratio of 16^{th} year exp. ratio of 8^{th} year exp. average patent value of trend number of 12^{th} year exp.	ratio of 8^{th} year exp. average patent value of trend number of 16^{th} year exp. ratio of 12^{th} year exp. ratio of 4^{th} year exp.	ratio of 12^{th} year exp. average patent value of trend ratio of 16^{th} year exp. number of 4^{th} year exp. ratio of 4^{th} year exp.
All Features	ratio of 4^{th} year exp. (topic) ratio of 16^{th} year exp. (topic) position (keyword) ratio of 4^{th} year exp. (keyword) average patent value of trend (topic)	ratio of 8^{th} year exp. (topic) ratio of 4^{th} year exp. (keyword) average patent value of trend (keyword) average patent value of trend (topic) number of 16^{th} year exp. (topic)	ratio of 12^{th} year exp. (topic) average patent value of trend (topic) number of 16^{th} year exp. (topic) ratio of 4^{th} year exp.(keyword) number of 4^{th} year exp. (topic)

TABLE I: Top-5 important features based on feature F-score

V. Related Work

In this section, we review existing work on patent analysis and management in three relevant directions.

Patent Quality Assessment and Ranking. In this line of related work, word age and syntactic complexity are used to assess the quality of patent applications to compute a score called patentability, which indicates how likely an application will be approved by the patent office [8]. A latent graphical model is proposed to infer patent quality, based on clarity of claims, originality, and importance of cited works [5]. A set of patent lexical features are extraced for automatic recommendation on a patent maintenance decision [4]. Instead of focusing on assessing patent quality or ranking patents, our work uses patent evaluation to validate the identified technological trends and related features.

Patent Novelty Detection. In this line of research, automatic discovery of core patents with high novelty and influence in a domain is studied [3]. The novelty factor and the impact of important phrases in a patent are investigated [2]. Documents are clustered based on their word bags in order to identify patent novelty based on each cluster [6]. Different from these previous works, our work proposes a framework to capture patent technological trends and derive patent novelty using the identified trends.

Heterogenous Patent Network Analysis. A heterogeneous patent network, is represented by several types of objects (companies, inventors, and patent documents) jointly evolving over time. A dynamic probabilistic approach to model the topical evolution of different objects in the heterogeneous network is proposed [7]. Also, in order to enhance patent quality, inventors tend to collaborate with colleagues those are productive and have complementary skills within an enterprize network. A ranking factor graph model for suggesting co-invention relationships is proposed in [9].

VI. Conclusion and Future Work

In this paper, we aim to identify patent technological trends and apply them for patent evaluation. We define a patent trend as a set of patents that share the same or similar technological attribute, and develop two approaches, topic-based method and keyword-based method, to identify patent trends. Based on the extracted trends, we develop trend features and within-trend patent features to capture the popularity of patent trends and the inherent value (novelty and impact) of patents, respectively. Then we apply these trend-based features in patent evaluation, which is formulated as a renewal status prediction problem. The experimental results show that our trend-based features effectively capture the patent value, and that combining these features with conventional features extracted from patent meta data significantly improve the patent evaluation performance.

For the future work, we plan to extend our study by identifying patent trends using other attributes, e.g., patent assignees and inventors. Moreover, we plan to apply the identified trends to better understand the knowledge flow and transmission among patents, development of certain technologies, and patent prior arts search.

VII. Acknowledgements

This work is supported by U.S. National Science Foundation Grants SMA-1063806 and SMA-1360205.

References

[1] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[2] M. A. Hasan, W. S. Spangler, T. D. Griffin, and A. Alba. Coa: finding novel patents through text analysis. In *KDD '09"*, pages 1175–1184, 2009.

[3] P. Hu, M. Huang, P. Xu, W. Li, A. K. Usadi, and X. Zhu. Finding nuggets in ip portfolios: Core patent mining through textual temporal analysis. In *CIKM'12*, pages 1819–1823, 2012.

[4] X. Jin, S. Spangler, Y. Chen, K. Cai, R. Ma, L. Zhang, X. Wu, and J. Han. Patent maintenance recommendation with patent information network model. ICDM '11, pages 280–289, 2011.

[5] Y. Liu, P. yun Hseuh, R. Lawrence, S. Meliksetian, C. Perlich, and A. Veen. Latent graphical models for quantifying and predicting patent quality. In *KDD'11*, pages 1145–1153, 2011.

[6] B. Shaparenko, R. Caruana, J. Gehrke, and T. Joachims. Identifying temporal patterns and key players in document collections. In *IEEE ICDM Workshop on Temporal Data Mining: Algorithms, Theory and Applications (TDM '05)*, pages 165–174, 2005.

[7] J. Tang, B. Wang, Y. Yang, P. Hu, Y. Zhao, X. Yan, B. Gao, M. Huang, P. Xu, W. Li, and A. K. Usadi. Patentminer: Topic-driven patent analysis and mining. In *KDD '12*, pages 1366–1374. ACM, 2012.

[8] Y.-H. Tseng and Y.-J. Wu. A study of search tactics for patentability search: a case study on patent engineers. In *PaIR '08*, pages 33–36. ACM, 2008.

[9] S. Wu, J. Sun, and J. Tang. Patent partner recommendation in enterprise social networks. In *WSDM'13*, pages 43–52, 2013.