International Association for Management of Technology
IAMOT 2010 Proceedings

Predictive Modeling of Patent Quality by Using Text Mining

HISASHI KASHIMA

Department of Mathematical Informatics, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan,
kashima@mist.i.u-tokyo.ac.jp

SHOHEI HIDO, YUTA TSUBOI, AKIRA TAJIMA

IBM Research - Tokyo, 1623-14, Shimo-tsuruma, Yamato, Kanagawa, 242-8502, Japan, {hido, yutat, tajima}@jp.ibm.com

TAKESHI UENO

IBM Japan Ltd., 1623-14, Shimo-tsuruma, Yamato, Kanagawa, 242-8502, Japan, utakeshi@jp.ibm.com

NAOKI SHIBATA

Innovation Policy Research Center, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan,
shibata@ipr-ctr.t.u-tokyo.ac.jp

ICHIRO SAKATA

Policy Alternatives Research Institutes, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan,
isakata@ipr-ctr.t.u-tokyo.ac.jp

TOSHIYA WATANABE

CCR (Collaborative Research Center), The University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo, 153-8505, Japan,
watanabe@wlab.u-tokyo.ac.jp

Abstract

In this paper, aligned with Nagata *et al*. (2008)'s work which models Japanese patent quality as its legal validity, we investigated predictive modeling of patent quality with the same data set consisting of cases of patent invalidation trials judged by the IP High Court (formerly, Tokyo High Court). We improve the predictive performance by using three technologies, machine learning, text mining, and information integration. We use machine learning techniques such as regularization and biased sampling, which result in improvements in several predictive performance metrics such as the AUC and the breakeven point. Next, extending the feature set used by Nagata et al., we apply text mining techniques to find informative textual features. We also identify several interesting textual representations that appear in high-quality patents. Finally, we integrate Nagata et al.'s tailored features and the textual features to achieve better predictive performance. Our results show that these techniques work together for better quantitative modeling of patent quality.

Keywords: patent management, patent quality, predictive modeling, text mining

Introduction

It is quite an important problem, not only for intellectual property departments, but also for managers of intellectual property in competitive business environments to correctly evaluate the value of each patent to their own business. Companies make these evaluations by focusing mainly on a patent's (1) technical value (for example, whether or not the patent is a pioneering invention or an improvement), (2) legal value (whether or not the patent will be held patentable/valid), and most importantly (3) economic value (whether or not the patent will bear a cash flow in the future).

However, considering a patent's value only from the perspective of a particular company sometimes results in a misleading valuation. For example, patent applications that have too broad and vague claims with few embodiments are sometimes patented, resulting in future litigation over the scope and validity of the patent. Another problem is when rights holders who are not interested in using their own patents try to threaten permanent injunction while demanding excessive royalty payments, an abusive practice called patent trolling. Although such a patent has a "value" in some sense, the value does not conform to the original objectives of the patent system, increases social costs, and actually inhibits innovations in the society that created the patent system.

Recently, "quality of a patent" is attracting attention as a new concept which emphasizes the public nature of the patent system. The quality of a patent reflects the contribution of the patent not to a company, but to the entire society consisting of other applicants, users, the patent office, and so on. In addition to implementing government policies such as improving the examination standards in the patent office, it is important to cultivate high quality patents to build communities that share ideas about patent quality and related data. One of possible ways to accelerate the growth of such communities is to provide quantitative metrics of patent quality that can provide achievable targets and that can be shared within industries.

Nagata *et al*. (2008) assumed that the quality of a patent is its legal validity, and built a quantitative model relating the features of each patent and the corresponding legal decisions based on cases of patent invalidation decided by the IP High Court (formerly, Tokyo High Court), which is the appellate court to review the decisions made by JPO Panel regarding the validity and those by district courts regarding the patent infringement and the validity. Their model is a logistic regression model that predicts the decisions on patent quality based on patent features such as the coverage of the claims taken from the patent specifications and the numbers of rejections and position documents taken from the patenting process. However, its predictive performance is not high enough to assess patent quality.

In this paper, we extend the predictive modeling of patent quality in the direction of Nagata *et al*., and try to improve the model with three technologies, machine learning, text mining, and information integration. We use machine learning techniques such as regularization and biased sampling to improve predictions, which results in improvements in several predictive performance metrics such as the AUC and the breakeven point. Next, extending the feature set used by Nagata *et al*., we use text mining techniques to find additional textual features. We were able to identify several interesting textual features that appear in high-quality patents. Finally, we integrated Nagata *et al*.'s features and the textual features to achieve better predictive performance. Our results show that these two kinds of features work complementarily for better quantitative modeling of patent quality.

Background

In this section, we describe the background of our study in this paper. First, we review the Japanese patent granting system that we investigate in this work by using our modeling approach. Next, we introduce three definitions of the patent quality, which are patentability, safety, and stability. Then we give the summary of the data we use in this work, and the formal formulation of our task.

Japanese patent granting system

We first briefly review the Japanese patent granting system. Figure 1 gives a simplified flowchart of the process that an invention might be involved with. The process start with a patent application filed with the Japan Patent Office (JPO) by the applicant. The patent application is published 18 months after the filing date (or after the priority date claimed), and within three years from the filing date, an applicant must request for substantive examinations. At the stage of substantive examinations, the JPO patent examiner assesses whether an invention disclosed in the patent application satisfies criteria for patentability. After several round-trip communications of rejection notices by the JPO and responses by the applicant, the invention is finally granted as a patent, otherwise not (See the decision marked “1” in Figure 1).

Even if once the patent is granted, some other parties may challenge the validity of the patent by filing trials for invalidation with the JPO (See the decision marked “2” in Figure 1), and decisions are made by the JPO panel. Anyone can file a trial for invalidation any time after issuance, even anonymously. In response to the decision by the JPO panel, the parties can appeal to the IP High Court (See the decision marked “3” in Figure 1). Although not shown in the Figure 1, also in a patent infringement lawsuit, the defendant may challenge the validity of the patent.

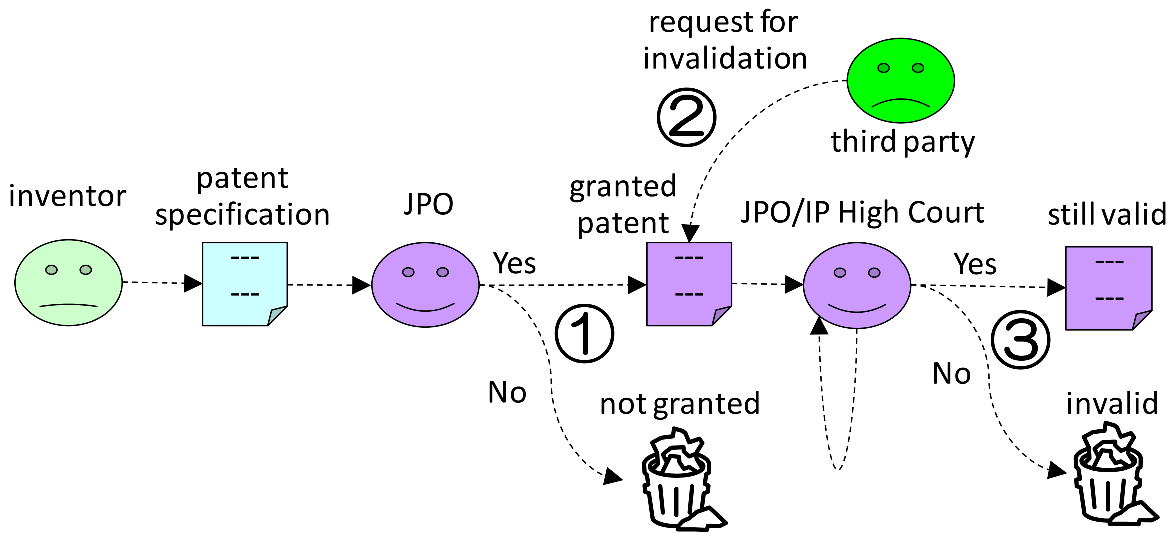


Figure 1. A simplified flowchart of the Japanese patent granting system.

Patent quality

When we view the patent granting system shown in Figure 1 with the purpose of modeling the quality of a patent, three visible targets for modeling patent quality are identified as follows.

(i) Patentability. Patentability of a patent indicates the possibility of the patent granted, since valid claims should be granted.

(ii) Safety. Safety of a patent indicates the possibility of the patent requested for invalidation by third parties. Patents given invalidation requests might have too broad claims (or have high economic value.)

(iii) Stability. Stability of a patent indicates the possibility of the patent surviving the court. Legally valid claims should have high possibility of survival.

Each of the three definitions reflects its own aspect of patent quality and is of interest. For example, Guellec and van Pottelsberghe (2000, 2002) took the patentability as the variable to be explained, and investigated the relationship between the patentability and various explanatory variables such as the internationality of the team invented the patent.

In this work, aligned with Nagata *et al*. (2008), we take the third definition of the patent quality.

Data Summary

Now we review the IP High Court decision data set used by Nagata *et al*. (2008) which we also use in this study. The data set has 710 cases, where 20% (142 cases) of the judgments were held valid (Note that judgments of partial invalidation were classified as invalid).

As the explanatory variable for describing the patent quality, Nagata *et al.* defined 60 tailored features including number of words in claims, number of "effect" words (for example, "can", "superior", and "advantageous"), number of domestic priorities, and number of references. See Figure 2 for the definition for some of them. Among the 60 features, and 24 of them were selected by using a statistical test based on simple linear regression models.

Action	Parameters	Definition
Applicant	Domestic_P	Number of Domestic Priorities
	Paris_P	Number of Priorities under the Paris Convention
	App_JPR	Number of Japan Patent References disclosed in a patent application by applicants
	App_FPR	Number of Foreign Patent References disclosed in a patent application by applicants
	Inventors	Number of Inventors
	Applicants	Number of Applicants
Agent	Claims	Number of Claims
	Claims_I	Number of Independent Claims
	Claims_D	Number of Dependent Claims
	Claims_C	Number of kinds of Claims Categories (e.g. "method claim" "product claim" "system claim")
	Words_AC	Number of words in All Claims (0.1 times)
	Words_TC	Number of words in Claim 1 (Top Claim)(0.1 times)
	Words_DE	Number of words in the item Detailed Explanation of invention (0.01 times)
	Words_DEEBA	Number of words in the rest except Background Art from "Words_DE" (0.01 times)
	Effects	Number of words describing "effect" in "Words_DEEBA" (e.g. "may/can" "could" "superior" "useful" "advantageous")
	Arguments	Number of filing arguments
	Amendments	Number of filing amendments
	Request_OI	Number of requests for oral Interview with examiner
Examiner	Exa_JPR	Number of Japan Patent References cited by the examiner
	Exa_FPR	Number of Foreign Patent References cited by the examiner
	Exa_NPR	Number of Non Patent References cited by the examiner
	Exa_AJPR	Number of Japan Patent References added by the examiner
	Exa_AFPR	Number of Foreign Patent References added by the examiner
	Rejection_W	Number of rejection notices for lack of written description
	Rejection_N	Number of rejection notices for lack of novelty/inventive step
	Rejection_O	Number of rejection notices for other reasons

Figure 2. Some of the features defined by Nagata *et al*. (2008).

Problem Formulation

Before describing our approach, we give a rather mathematical formulation of our task. Let denote by X:=$\{x_1, x_2, \ldots, x_N\}$ the N patents sent to the IP High Court for judgment, and by Y:=$\{y_1, y_2, \ldots, y_N\}$ the corresponding judgments given by the High Court, where each y_i takes a value either of +1 (which indicates "valid") or -1 (which indicates "invalid"). Our goal is to build a model f which relates the patent instances of X to their corresponding judgments Y, i.e., $y_i = f(x_i)$. Nagata *et al.* (2008) used the 60 features described in the previous subsection to represent each x_i as a 60-dimensional vector, and the logistic regression model as f.

Method

In this section, we describe our predictive modeling approach for patent quality. First, we review several machine learning techniques proposed for improving predictive performance, including the support vector machine, class-proportionate weighting, and L_1-regularization. Then we introduce text processing techniques since we handle patent specifications written in natural language, and information integration techniques for aggregating a tailored-feature-based model and a text-based model. Finally, we introduce two metrics for evaluating the predictive performance of a model, which are break-even point and AUC.

Predictive modeling of patent quality

The major interest by the existing work on modeling patent quality is in understanding the causal relationships between explanatory variable and the patent quality. In other words, their modeling approaches were descriptive. However, if we would like to use the model for decision making, for example, for prioritizing patents to be filed or maintained, or for evaluation of patents from the quality viewpoint, we expect the model has reasonable predictive power.

In this study, we focus on modeling approaches putting stress on the models' predictive performance. Our strategies for improving the predictive performance are three-fold: adopting recent techniques developed in the field of machine learning, text-based modeling using text mining techniques and information integration of textual features and tailored features.

Machine learning techniques for improving predictive performance

Various techniques for improving the predictive performance of models have been developed in the field of machine learning (Bishop, 2006). In this study, we adopt some of the techniques including the support vector machine, class-proportionate weighting scheme for the imbalanced data set, and the L_1-regularization method for preventing over-fitting to the high-dimensional data set.

Support vector machine

The support vector machine (SVM) is one of the state-of-the-art predictive models which was first introduced by Vapnik (2000). It is well-known for its high predictive performance, and has been applied in numerous application areas, for example, molecular biology, text and image processing, and financial engineering. The SVM assumes the following linear model:

$$y := \text{sign } f(x) := \text{sign}(w^{\mathrm{T}}x) := \text{sign}(w_1x_1+w_2x_2+\ldots+w_dx_d) \;,$$

where $x = (x_1, x_2, \ldots, x_d)$ is a d-dimensional feature vector and $w = (w_1, w_2, \ldots, w_d)$ is the parameter vector of the same dimension which specifies the model. A positive value of w_j indicates the j-th feature x_j positively contributes to the patent quality, while a negative value contributes to it negatively. The sign function returns +1 when its argument is positive, and returns -1 otherwise. Given the data set X and Y, the SVM learning algorithm find the optimal parameter w^* which minimizes the following objective function:

$$\Sigma_i \max \{ 1-y_i f(x_i), 0 \} + c \|w\|_2^2,$$

where the first term is the loss function which penalizes misclassifications, and the second term is the regularization term which avoids over-fitting to the given data set. c is a small constant which balances the two terms, which was set to 0.1 throughout our experiments. The regularization term $\|w\|_2^2 := w_1^2 + w_2^2 + \ldots + w_d^2$ penalizes the parameter vector being too large, and it is known to work well when we predict with data outside the given data set.

As for the parameter estimation of the SVM, we use an efficient sequential estimation method proposed by Duchi et al. (2008).

Class-proportionate weighting

Our data set is highly imbalanced, since only 20% of the 710 cases are valid. In such cases, it is known that the predictive performance is improved by appropriately weighting the valid and invalid data. The data in the minority class are given high weights, while those in the majority class are given low weights. Namely, we modify the objective function of the SVM as

$$\Sigma_i \beta_i \max \{ 1-y_i f(x_i), 0 \} + c \|w\|_2^2,$$

where we have the weight β_i of the i-th data. The β_i is usually set to the inverse of the class proportion, that is, we set $\beta_i = 1/0.2 = 5$ for the valid patent, and $\beta_i = 1/0.8 = 1.25$ for the invalid patent in our case.

L_1-regularization

When the dimensionality of the feature vector is extremely large compared to the number of data, the problem called over-fitting arises, which is a phenomenon that the predictive performance for the new data other than the data used for fitting the model severely degrades. As we saw in the formulation of the SVM, the regularization term plays a role for avoiding the over-fitting problem. However, for the extremely high-dimensional cases such as in the text-based modeling that we will see in the next subsection, the regularization term $\|w\|_2^2$ (called L_2-regularization) still suffers from the over-fitting. In such cases, use of a more aggressive regularization term called L_1-regularization term instead of the L_2-regularization term is known to be effective. The L_1-regularization term defined as $|w|_1 := |w_1| + |w_2| + \ldots + |w_d|$ tends to suppress many of the parameters to zero, which results in a small amount of parameters selected.

Text-based modeling using text mining techniques

In the work by Nagata *et al*. (2008), it is reported that the numbers of "effect" words such as "can", "superior", and "advantageous" are useful for predicting the judgments by the IP High Court. It is natural to imagine that other textual representations in the patent specification also might help. In this study, in contrast with the Nagata *et al*'s work where they prepared textual representations that seem to be effective in advance, we use textual representations exhaustively by using text mining techniques.

Morphological analysis

Since Japanese language is usually not segmented, we first have to extract "words" from the specifications. For this purpose, we use a morphological analyzer to segment each sentence into words with part-of-speech tags. In this study, we used a Japanese morphological analyzer called "MeCab" (http://mecab.sourceforge.net/). Figure 3 shows how a morphological analyzer works. The Japanese sentence (which means "we can significantly shorten the time for brewing" in English) is segmented into words where each of partitions indicates a word boundary. The tag associated with each segmented word such as "noun" and "particle" indicates the part-of -speech for the word. Using the segmented words, we construct the feature vector x for a patent specification, where each x_i takes 1 if the i-th word in the resultant dictionary (for example, the Japanese word corresponding to "significantly") appears in the text, and takes 0 otherwise.

We used only words appearing more than 20 times in the data set, and obtained about 2,400 words in total, which means each patent specification is a 2,400-dimensional feature vector.

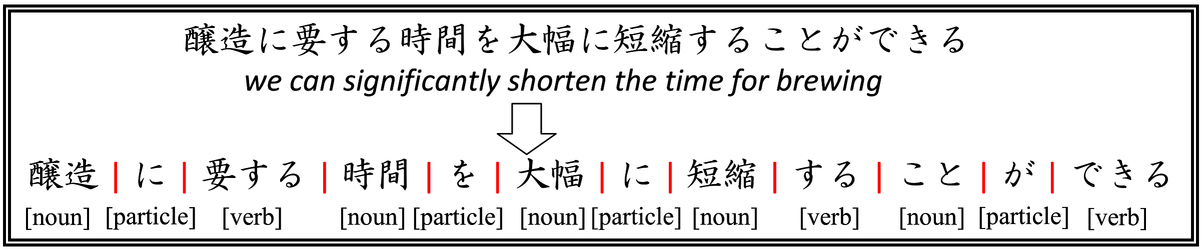


Figure 3. How a morphological analyzer works.

Extracting textual representations

Although the words themselves indeed have semantic information, textual patterns consisting of several consecutive words are more informative. Therefore, we also use 2 or 3 consecutive words as features. In our data set, we obtained 13,000 patterns.

Since the number of dimensionality is higher than the number of the data, we use the L_1-regularization, which works for extracting small number of effective words and textual representations and also makes it easier to interpret the resultant model.

Model integration of the tailored-feature-based model and the text-based model

In the previous subsection, in addition to the tailored-features such as the number of words in the claims of a patent, we proposed to use textual features extracted from the raw-texts of the patent specifications, which might work complementarily with the tailored features. It is quite natural to combine these two kinds of features in our model.

In this study, we investigate two types of integration. The first one is to take the sum (or average) of the outputs of two models as

$$y := \mathrm{sign}(\, f^{\,tailored}(x) + f^{\,text}(x)\,),$$

where $f^{tailored}$ is the model estimated by using the tailored features, and $f^{\,text}$ is the one by using the textual features. This type of integration assumes that the two kinds of information work cooperatively. Another type of integration we consider is to take the maximum of the two outputs, that is,

$$y := \mathrm{sign}(\ \max\{\ f^{\,tailored}(x),\quad f^{\,text}(x)\ \}\ ),$$

which assumes that the two kinds of information work complementarily. Note that the output ranges of the two modes are different. We normalize the scale of the output of each mode l by subtracting the average and dividing by the standard deviation.

Evaluation metrics

From the predictive point of view, what is important is not how well a model fits to a given data set, but how well it performs for future data. However, most of the existing works only discuss fitness of the model to the given data set. (In such sense, they are considered descriptive.) In this study, we investigate the model's predictive performance for the future data. Since the actual future data are not available, we simulate such situation by using cross validation scheme, which is a standard statistical procedure to estimate predictive performance. Also, we introduce two metrics of predictive performance, the break-even point and the AUC.

Cross validation

To simulate such situation where a model is trained on a given data set and tested on a future data set, we keep some part of the given data set that is not used for model estimation and used for performance evaluation. In *k*-fold cross validation scheme, the data set is divided in to *k* almost-equally-sized bins, one of which is used for performance evolution. By changing the bin for evaluation, we can evaluate the model performance *k* times. The *k* resultant measurements are averaged to obtain the final score. In our experiments, we used 5-fold cross validation.

Evaluation metrics of predictive performance: break-even point and AUC

To evaluate the predictive performance of a model on a test set, we use two evaluation measures: the break-even point for evaluating predictive accuracy with thresholding, and the AUC for evaluating quality of ranking without thresholding.

Let us assume that we would like to prioritize a pool of patents according to their quality to select a part of them to be maintained. One way to do so is to take some threshold value of the patent quality score, and decide to maintain those above the threshold. Another way might be choosing top 10 patents. If we know the ratio of the valid patents in the evaluation set, the asymptotically optimal threshold should be set to the score with which the number of patents predicted to be valid (i.e. the number of patents above the threshold) is equal to that of the actually valid patents. With the optimal threshold, the model achieves 100% prediction accuracy if it works perfectly. The prediction accuracy using the optimal threshold is called the break-even point. (See also Figure 4.)

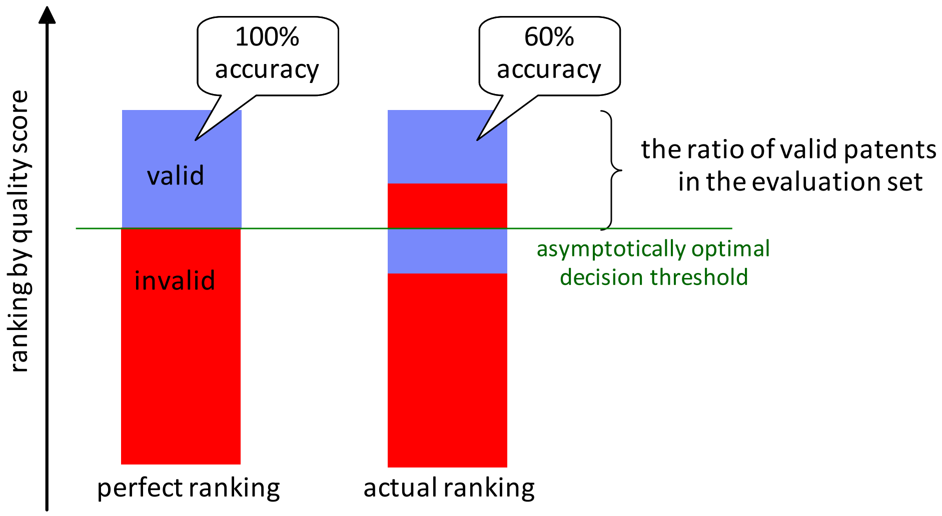


Figure 4. The break-even point.

AUC (which is an acronym of the area under the curve) is another metric of predictive performance. The AUC evaluates the goodness of ranking of the instances in the evaluation set given by a model, and defined as the probability of a randomly-picked valid patent ranked higher than a randomly-picked invalid patent. The AUC cares only about the relative order of the scores, and is not affected by the decision threshold. For this reason, the AUC is widely used as a standard performance measure of predictive models.

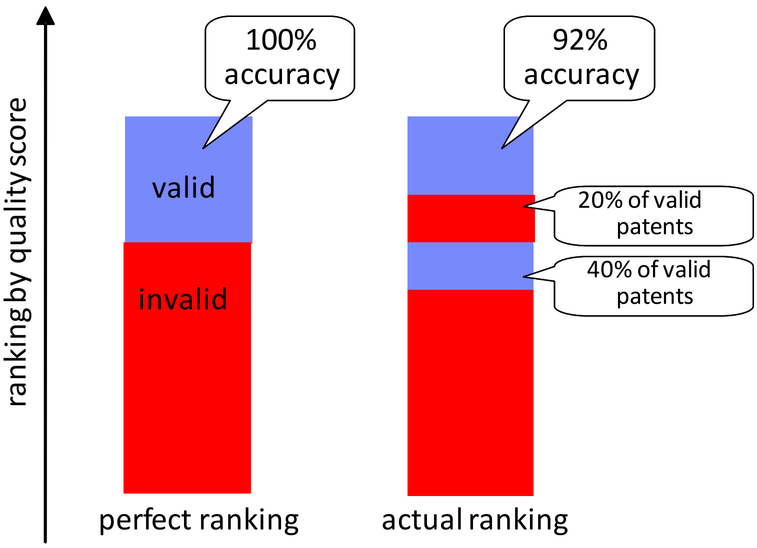


Figure 5. The AUC.

Results

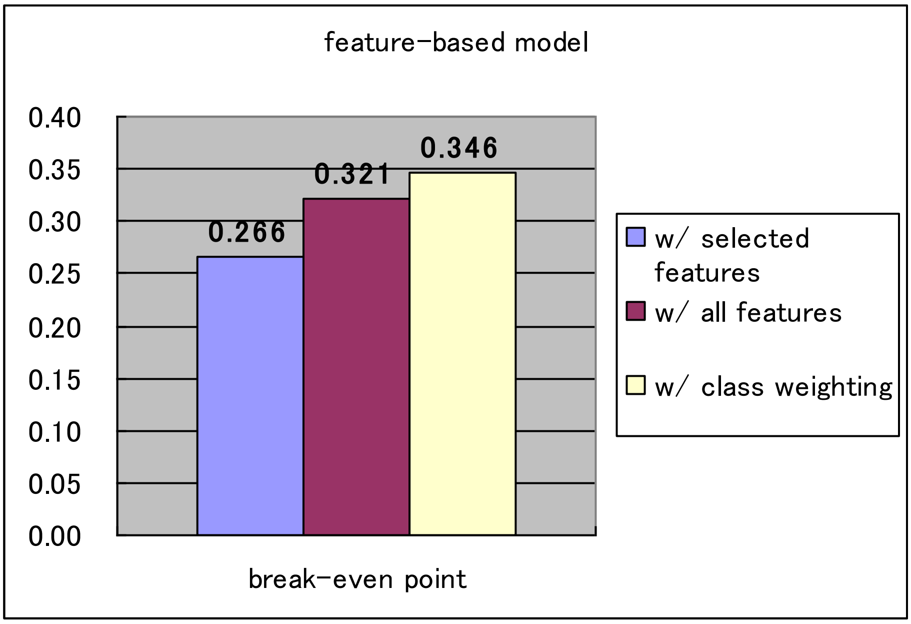
In this section, we report the experimental results obtained by applying the methodologies to the data set. The results are three-fold, that are, (i) various machine learning techniques actually improves the predictive performance, (ii) the textual-feature-based model performs well, and (iii) combination of the feature-based model and the text-based model work complementarily for improving the performance.

Machine learning techniques improve the predictive performance

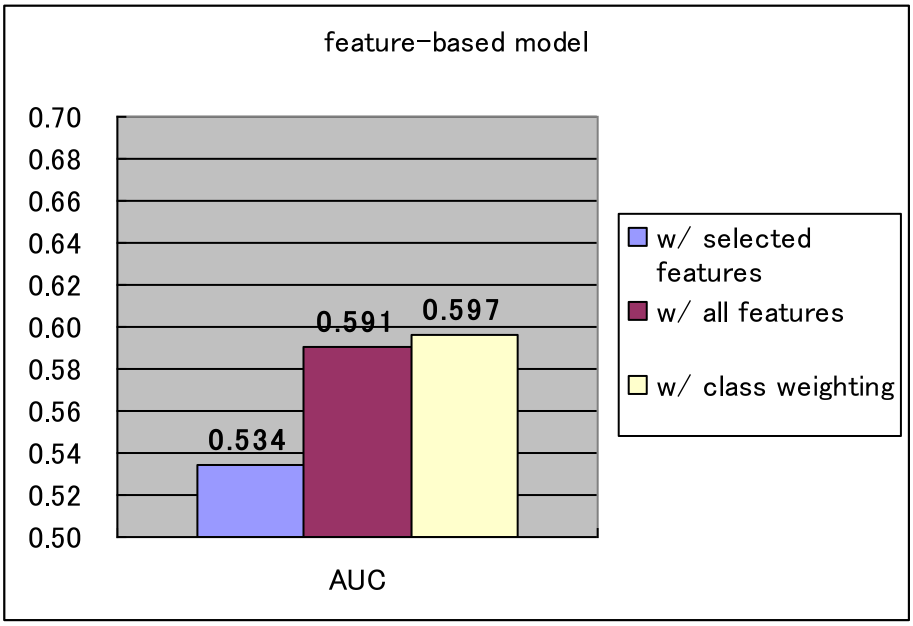
We compared predictive performance of the support vector machine with the 24 selected features used by the logistic regression model of Nagata *et al.*, one with all of the 60 features, and one with the class-proportionate weights.

The objective of this experiment is to show that, from predictive viewpoint, it is better to use all of the features, and that the class-proportionate weighting technique improves the predictive performance since our data set (where only 20% of the cases are labeled "valid") is highly skewed.

Figure 6 summarizes the results. Figure 6(a) shows the break-even points by the three models, and Figure 6(b) show the AUC values. We can observe that the model with the 60 features outperforms the model with the 24 features, and also, the class-proportionate weighting scheme further improves the performance.



(a)



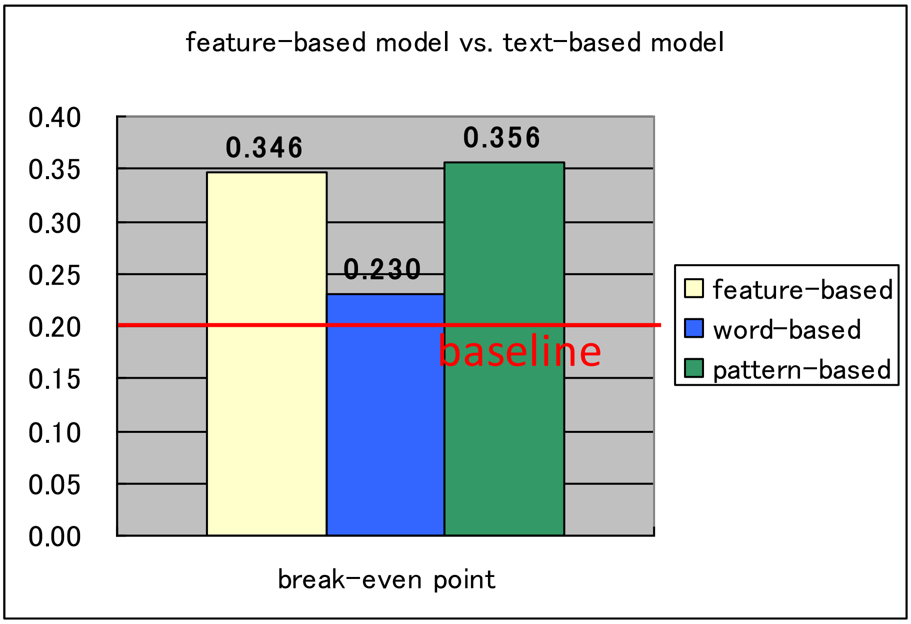
(b)

Figure 6. Comparison of the model with the 24 features, the model with the 60 features, and the model with the class-proportionate weighting.

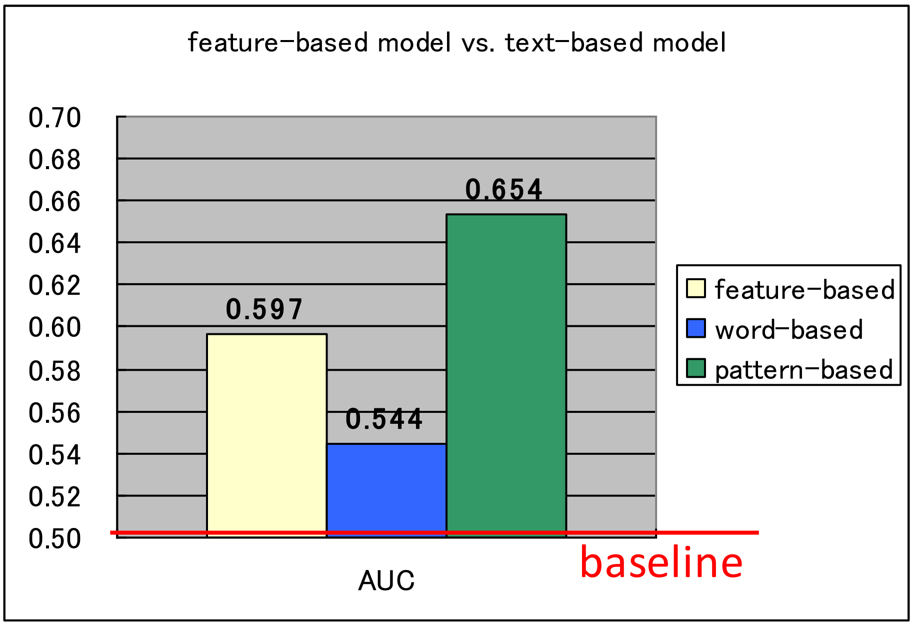
Text-based modeling

Using the textual features extracted by using the morphological analysis, we compared two types of text features. The first model is the one that uses only the segmented 2,400 words, and the second model is the one with the 13,000 patterns consisting of two or three contiguous words (bi-/tri-gram patterns). Since the numbers of features are large, we used the L_1-regularization scheme.

Figure 7 summarizes the results. Figure 7(a) shows the break-even points by the two models as well as the model based on the 60 tailored features and the class proportionate weighting, and Figure 7(b) show the AUC values. The word-based model has some predictive power and supports the observation that effect words are informative as mentioned by Nagata *et al.*, but its predictive performance is not so high when compared with that of the feature-based model. This is probably because single words are not sufficient to capture the semantics and representation in patent specifications. The pattern-based model rather overcomes this weakness, and surprisingly, it achieves higher performance than the feature-based model. This fact shows the possibility of extensive use of text mining technologies in patent quality modeling is a quite promising future direction.



(a)



(a)

Figure 7. Comparison of the models with the textual features and the model with the tailored features.

Next, we investigate the patterns selected by the model. By using L1-regularization, about 100 features were selected. Among them, we found three interesting clusters of patterns. The first two clusters are expressions that clarify or limit coverage of claims were found. The first cluster seems to be about parameter specification, which often makes the range of claims clear and sometimes conservative, which finally results in legal stability when the patent are sent to the IP High Court. The second one is about extensions of existing patents. Since they are based on existing patents, once they are granted for their novelty and progressivity, they become stable. In the last cluster, we can find effect representations. Again, this fact supports Nagata *et al.*'s observation since they were automatically found without preparing effect words in advance.

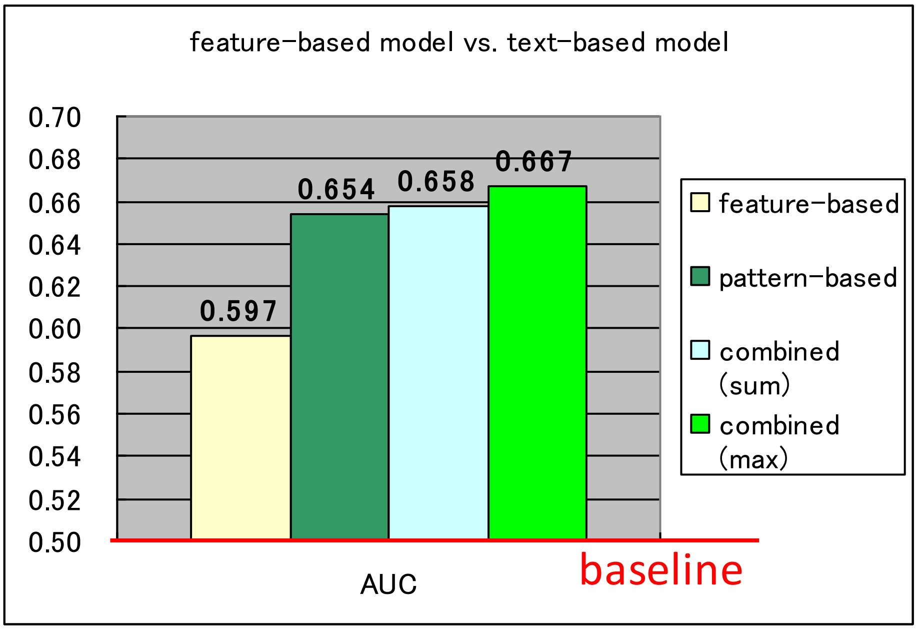
interpretations	patterns （in Japanese）	meanings of the patterns
parameters	度合い[noun]−を[particle]	degree of …
	確率[noun]−の[particle]	probability of …
	の[particle]−設定[noun]	setting of …
extension of existing patents	(実施)形態[noun]−による[particle], で[particle]−用い[verb]−て[particle]	executed in the condition of …
	に[particle]−置き換え[verb]	substitute … with …
	薄型[noun]−化[noun]	reduce the thickness of …
effect representation	を[particle]−良く[adjective]	well
	正しい[adjective]	correct
	可撓性[noun]	flexibility
	利点[noun], 利点[noun]−を[particle]	advantage
	調整[noun]−可能[noun]	adjustable

Figure 8. Examples of the textual patterns used in the model.

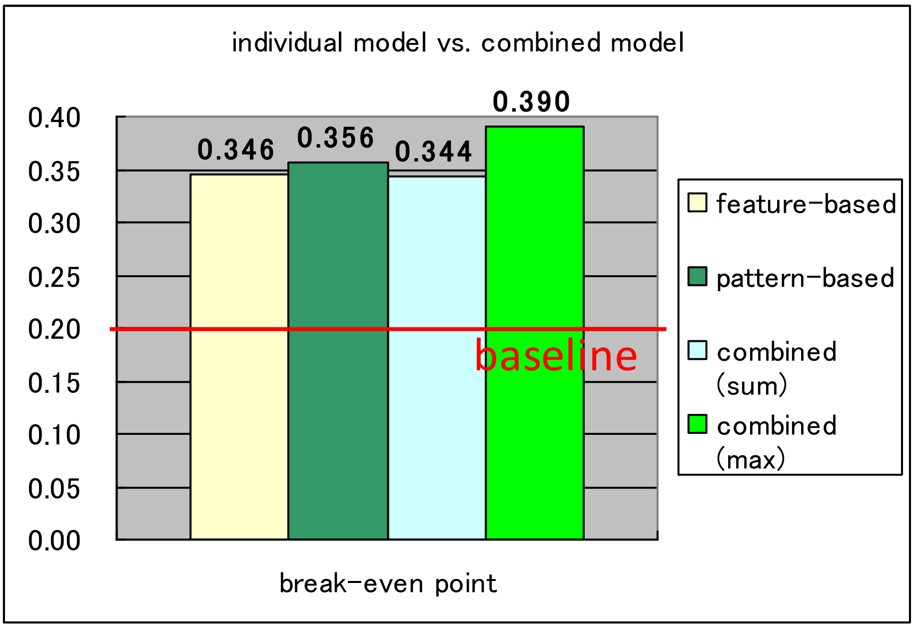
Combination of the tailored-feature-based model and the text-based model

Finally, we investigate combination of the feature-based model and the text-based model. We compared two aggregation methods, taking their (normalized) sum and by taking their (normalized) maximum.

Figure 9 summarizes the results. Figure 9(a) shows the break-even points by the two aggregation methods with those of individual models, and Figure 9(b) show the AUC values. We can observe the aggregation by using the maximum of the two models works well, which implies that the tailored features and the text features complementarily contribute to prediction. This is good news because two parallel efforts to define good tailored features based on domain knowledge and to promote data-driven exhaustive text pressing complementarily drive improvement of patent quality modeling.



(a)



(b)

Figure 8. Comparison of the sum aggregation and the maximum aggregation methods.

Conclusions

In this paper, aligned with Nagata *et al*.'s work which assumes that the quality of a patent is its legal validity, we investigated predictive modeling of patent quality with the same data set of cases of patent invalidation decided by the IP High Court (formerly, Tokyo High Court). We improved the predictive performance with three technologies, machine learning, text mining, and information integration. We used several machine learning techniques such as support vector machines, regularization and biased sampling to improve predictions, which resulted in improvements in several predictive performance metrics such as the AUC and the breakeven point. Next, extending the feature set used by Nagata *et al*., we applied text mining techniques to find additional textual features. We were able to identify several interesting textual features that appear in high-quality patents. Finally, we integrated Nagata *et al*.'s tailored features and the textual features to achieve better predictive performance. Our results show that these two kinds of features work together for better quantitative modeling of patent quality.

Acknowledgements

We would like to thank Rinju Yohda and Yusuke Kanehira of IBM Japan and Rikiya Takahashi of IBM Research - Tokyo for valuable discussion. We also thank Tetsuji Kuboyama of Gakushuin University and Kentaro Nagata for data preparation.

References

Bishop, CM (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Duchi, J, S Shalev-Shwartz, Y Singer and T Chandra (2008). Efficient Projections onto the l_1-Ball for Learning in High Dimensions, In *Proc. 25th international conference on Machine learning*.

Guellec, D and B van Pottelsberghe (2000). Applications, grants and the value of patent, *Economics Letters*, 69(1), 109-114.

Guellec, D, and B van Pottelsberghe (2002). The value of patents and patenting strategies: countries and technology areas patterns, Economics of Innovation and New technology, 11(2), 133-148.

Nagata, K, M Shima, N Ono, T Kuboyama and T Watanabe (2008). Empirical Analysis of Japan Patent Quality, In *Proc. 17th international conference on Management of Technology, the International Association for Management of Technology (IAMOT)*.

Vapnik, VN (2000). *The Nature of Statistical Learning Theory.* New York: Springer.