Predictive Modeling of Patent Quality by Using Text Mining

Hisashi Kashima, Naoki Shibata, Ichiro Sakata, Toshiya Watanabe
University of Tokyo
Shohei Hido, Yuta Tsuboi, Akira Tajima
IBM Research - Tokyo
Takeshi Ueno
IBM Japan

Data and text mining techniques improve predictive modeling of patent quality

- We model “patent quality” which is a goodness measure of a patent for entire society from the predictive viewpoint
- We show data mining and text-mining techniques improve prediction
- Combining both, we further improve prediction
Background: Patent value is important for companies ... but this is not always true for entire society

- It is important to evaluate the value of each patent to one’s own business:
  - Technical value for R&D (whether it is a pioneering invention or an improvement)
  - Legal value for IP departments (whether it will be held patentable/valid)
  - Economic value for business units (whether it will bear a cash flow in the future)
- There are several attempts to model and evaluate the patent value
- However, considering a patent’s value only for a particular company sometimes results in increasing social costs, and inhibiting innovations ...
  - Granted patents with too broad and vague claims with few embodiments result in future litigations
  - Patent trolling: abusive practice by rights holders trying to demand excessive royalty payments to other companies

Background: Patent quality is goodness of a patent for society ... but its quantitative modeling is the key

- We focus on “quality of a patent”, a new concept which emphasizes the public nature of the patent system (contrast with the patent value)
- The quality of a patent is the contribution of the patent not to a company, but to the entire society
- By sharing ideas about patent quality and related data, we expect to improve the quality of patent applications and examinations
- One of the ways is to provide quantitative metrics of patent quality that can provide achievable targets shared within industries.
- But how?
Prior work: Nagata et al. modeled patent quality as legal validity

- Nagata et al. considered legal validity as a proxy of patent quality
  - Patents with appropriate descriptions, claims, and examinations are robust to litigations, which will reduce social cost
- They built a regression model to explain 710 legal decisions (valid/invalid) by the IP High Court in Japan for cases of patent invalidation requests

Nagata, K, M Shima, N Ono, T Kuboyama and T Watanabe
Empirical Analysis of Japan Patent Quality
In Proc. 17th IAMOT, 2008

Model of patent quality score:
A more valid patent \( x \) will get a higher score \( f(x) \)

- A patent specification \( x \) is represented as a set of features \( (x_1, ..., x_d) \)
- Each parameter corresponds to contribution of each feature to the “patent quality score”, which is estimated from data

\[
f(x) = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d
\]

A patent
features of a patent \( x \)
patent quality score
model parameters
Tailored features used by Nagata et al.: They defined 60 hand-made features

\[ f(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d \]

- **Parameters**
  - **Dom_P**: Number of Domestic Priorities
  - **Fam_P**: Number of Priorities under the Pari-Mutuel 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
  - **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 times)
  - **Words_DE**: Number of words in the Item Detailed Explanation of Invention (0.01 times)
  - **Words_DEB**: Number of words in the rest except Background Art from “Words_DE” (0.01 times)
  - **Effects**: Number of words describing “effect” in “Words_DEB” (e.g., “may,” “can,” “could,” “superior,” “useful,” “advantageous”)
  - **Arguments**: Number of filing arguments
  - **Assessments**: Number of filing assessments
  - **Request_OI**: Number of requests for oral interview with 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_APR**: Number of Japan Patent References added by the examiner
  - **Exa_FPR**: Number of Foreign Patent References added by the examiner

Our goal: **Predictive** modeling of patent quality

- **Nagata et al.** focused on **descriptive** modeling
  - “Which feature is responsible for explaining court decisions (= patent quality) ? ”

- To be used as a reliable quality measure, the model should have high predictive power
  - Also useful for selecting patents to file or hold
- Our goal is to improve the predictive power of the patent quality model
Results: We improved the patent quality prediction model by using data mining and text mining techniques

- Data mining techniques for prediction:
  - Support vector machines (SVMs) for accurate predictive modeling
  - Class-proportionate weighting for addressing biased data

- Text mining techniques for exhaustive text feature construction from patent specifications
  - Morphological analysis for natural language processing
  - L1-regularization for addressing high-dimensional data

- Furthermore, combination of both boosts the predictive power

Key for improvement 1: Use all features

- Nagata et al. selected 24 promising features out of 60 features, but can we improve the predictive accuracy by using all of them?

- In data mining, it is common to use all features by using the framework called regularization
  - Regularization prevents model parameters \((w_1, w_2, \ldots, w_d)\) from being too large or too small

\[
    f(x) = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d
\]

- by penalizing \(\|w\|_2^2 := w_1^2 + w_2^2 + \ldots + w_d^2\)

- We use support vector machine, which is a state-of-the-art prediction model used in data mining
Key for improvement 2: Address the *bias* in the data

- Valid patents make up only 20% of the whole data
  - Invalid cases are 80%/20% = 4 times as many as valid cases
- Can we use this bias information to improve estimation?
- Intuitively, it sounds nice to put more importance on valid cases (=minorities)
  - Class proportionate weighting: Estimates the model by giving 4 times as large weights to valid cases as those to invalid cases
    - known to improve predictive performance

Result 1&2: Data mining techniques improve prediction!

- Using support vector machine, the predictive performance improves
  - when we use all 60 features
  - when we use class-proportionate weighting

![Graph showing AUC and BEP with different feature sets and weighting options](image)
Key for improvement 3: Use text information

- In patent specifications, we have rich text information
- We use text mining techniques to exhaustively construct features from texts
  - **Morphological analyzer** to segment Japanese language into words

  窒造に要する時間は大幅に短縮することができる
  (we can significantly shorten the time for brewing)

  魚造 | 要する | 時間 | を | 大幅 | に | 短縮 | する | こと | が | できる
  (noun | particle | verb | noun | particle | noun | particle | noun | verb | noun | particle | verb)

- Combining words to extract 13,000 patterns consisting of 2 or 3 words
- **L1-regularization** for addressing high-dimensional data (#features >> #data)
  - L1-regularization dramatically and automatically reduces the number of features used in the model (then we got about 100 selected features)
    - by penalizing $|w_1| := |w_1| + |w_2| + \ldots + |w_d|

---

Result 3.1: Text information improves prediction!

- We built two models:
  - The model with 2,400 words
  - The model with 13,000 patterns consisting of 2- or 3-consecutive-word patterns
- The model with word patterns improves the predictive performance
Result 3.2:
We found textual patterns implying high patent quality

- Investigating the model, we found informative text representations:
  - Textual patterns clarifying or limiting coverage of claims
  - Textual patterns representing effects of patent executions
    - This is consistent with the mention by Nagata et al.

<table>
<thead>
<tr>
<th>clarifying or limiting coverage of claims</th>
<th>patterns (in Japanese)</th>
<th>meanings of the patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameters</td>
<td>意味合い [noun]-を [particle]</td>
<td>degree of ...</td>
</tr>
<tr>
<td></td>
<td>確率 [noun]-の [particle]</td>
<td>probability of ...</td>
</tr>
<tr>
<td></td>
<td>の [particle]-設定 [noun]</td>
<td>setting of ...</td>
</tr>
<tr>
<td>extension of existing patents</td>
<td>(実施)形態 [noun]-による [particle], で [particle]-用いる [verb]-で [particle]</td>
<td>executed in the condition of ...</td>
</tr>
<tr>
<td>effect representations</td>
<td>に [particle]-置き換え [verb]</td>
<td>substitute ... with ...</td>
</tr>
<tr>
<td></td>
<td>薄型 [noun]-化 [noun]</td>
<td>reduce the thickness of ...</td>
</tr>
<tr>
<td></td>
<td>良く [adjective]</td>
<td>well</td>
</tr>
<tr>
<td></td>
<td>正しい [adjective]</td>
<td>correct</td>
</tr>
<tr>
<td></td>
<td>可撓性 [noun]</td>
<td>flexibility</td>
</tr>
<tr>
<td></td>
<td>利点 [noun]</td>
<td>advantage</td>
</tr>
<tr>
<td></td>
<td>利点 [noun]-を [particle]</td>
<td>advantage</td>
</tr>
<tr>
<td></td>
<td>調整 [noun]-可能 [noun]</td>
<td>adjustable</td>
</tr>
</tbody>
</table>

Key for improvement 4:
Combine tailored-feature-based model and text-based model

- Can we further improve the prediction by combining the 1st (tailored-feature-based) model and the 2nd (text-based) model

- Two ways of combining two models:
  - Collaborative model: sums the outputs by two models
    \[
    f^{\text{tailored}}(x) + f^{\text{text}}(x)
    \]
  - Complementary model: takes the maximum of the two models
    \[
    \max \{ f^{\text{tailored}}(x), f^{\text{text}}(x) \}
    \]
Result 4:
Two models work complementarily to improve prediction

- Complementary model (taking the max.) works well
- This means that two models work complementarily
  - “Right model in the right place”

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>BEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>feature-based model</td>
<td>0.654</td>
<td>0.346</td>
</tr>
<tr>
<td>pattern-based model</td>
<td>0.658</td>
<td>0.356</td>
</tr>
<tr>
<td>combined model (sum)</td>
<td>0.667</td>
<td>0.390</td>
</tr>
<tr>
<td>combined model (max)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion: **Data and text mining techniques improve predictive modeling of patent quality**

- We modeled not “patent value” for a specific company, but “patent quality” for entire society, from the predictive viewpoint
- We showed data mining techniques improve prediction (1, 2)
- Using text mining techniques, we showed texts are informative for patent quality modeling (3)
- Hand-made features and text-based features work complementarily to improve prediction (4)
- Future work includes:
  - More precise modeling using large scale data
  - Modeling with other proxies of patent quality (e.g. patentability)
Acknowledgements

We would thank:

- Rinju Yohda (IBM Japan, IP department)
- Yusuke Kanehira (IBM Japan, IP department)
- Rikiya Takahashi (IBM Research - Tokyo)
- Tetsuji Kuboyama (Gakushuin University)
- Kentaro Nagata

for their help
Simplified flowchart of the patenting system in Japan

- Nagata et al. focused on modeling (3)

Evaluation method of predictive accuracy: 
**Cross validation** and two predictive performance metrics (**AUC & BEP**)

- Cross validation allows us to virtually evaluate predictive performance on future cases
  - Use 80% of the data for modeling
  - Use the remaining 20% for evaluation (with court decisions hidden)

- 2 widely-used predictive performance metrics: AUC and BEP
  - AUC (Area Under the ROC Curve):
    - Evaluates the quality of ordering of predictions
    - Equivalent to AR(Accuracy Rate)-value used as a performance metric for default prediction in financial engineering
  - BEP (Break-Even Point):
    - Evaluate accuracy rate with an optimal decision threshold
    - Used for evaluating quality of automatic text classification
AUC: a measure of ranking quality

- The patents in the evaluation set are ordered by using the model
- AUC is probability of a randomly-picked stable patent ranked higher than a randomly-picked instable patent
- AUC is a measure of quality of ranking

Break-even point: a measure of predictive accuracy with threshold

- The patents in the evaluation set are ordered by using the model
- Top $N$ instances are predicted as “stable”, where $N$ is the number of stable patents in the evaluation set
  - because this is the optimal decision threshold if the model is correct
- Break even point is predictive accuracy for the instances given “stable” labels by using the optimal threshold