Patent Maintenance Recommendation System

Mengke Hu

ASPITRG Group, ECE Department
Drexel University

mengke.hu@gmail.com

April 2, 2015

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

References

[1] 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," in *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, (Vancouver,BC), pp. 280–289, December 2011.

[2] M. A. Hasan, W. S. Spangler, T. Griffin, and A. Alba, "COA: Finding Novel Patents through Text Analysis," in *KDD '09 Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, (Paris, France), pp. 1175–1184, June-July 2009.

[3] R. Kosala and H. Blockeel, "Web Mining Research: A Survey," *ACM SIGKDD Explorations Newsletter*, vol. 2, pp. 1–15, June 2000.

[4] I. De, "Experiments in First Story Detection," in *Proceedings of the 2005 National Conference on Undergraduate Research (NCUR)*, 2005.

[5] N. A. Arnosti and A. P. Danyluk, "Feature Selection via Probabilistic Outputs," in *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, 2012.

Related Work

- Automatic patent classification
- Prior art search
- Patent quality prediction
- Novelty detection
- Traditional link search (page rank citation ranking)
- HITS (Hyperlink-induced Topic Search/ hubs and authorities): link analysis algorithm by Kleinberg

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Patent Maintenance Problem

Patent Maintenance:

- A patent can be held for up to 20 years.
- Large maintenance fees need to be paid to keep it enforceable.
- If a patent is not valuable, the owner may decide to abandon it by stopping paying the maintenance fees to reduce the cost.

Decision for large companies and organizations:

- Too many patents need to be investigated, which makes it difficult to make decision.

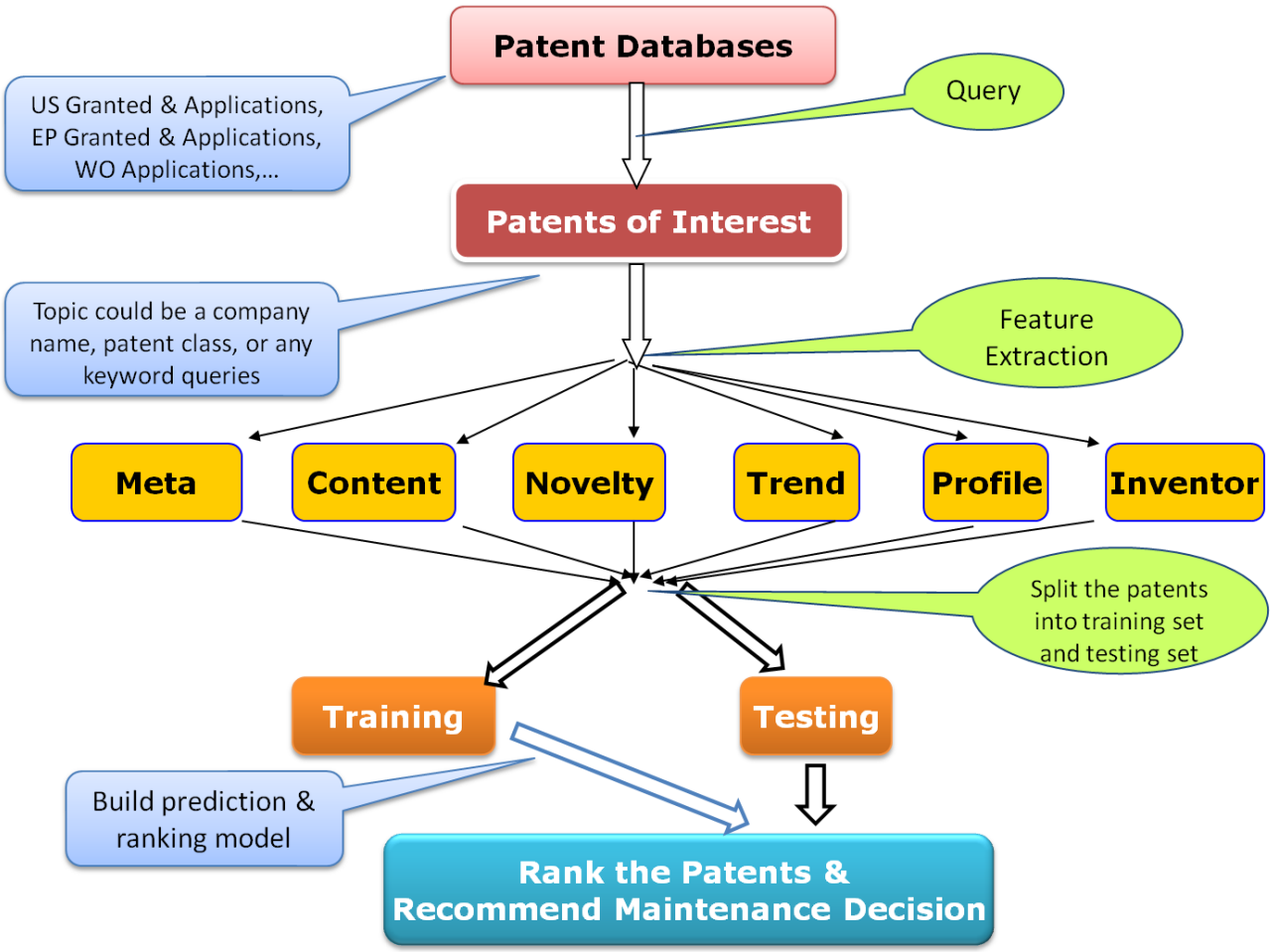
What do companies think to maintain a patent?

- Quality of the writing of the patent:
  If a patent is not well written, it is more likely to be rejected by the patent office or ruled as invalid in court.
- Novelty of the patent:
  If a granted patent is not novel anymore, the owner should abandon it to avoid possible lawsuit.
- Technique value:
  If there is another technique emerged as a better solution, then there is no need to maintain the patent.
- Technique relevancy:
  A company tends to keep patents that are related to its current core interests.
- Economic value:
  If a patent is making profit for the company, then the company will likely maintain it.

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Recommendation System [1]



The goal is to automatically extract patent properties from various aspects to help the company make a decision on maintenance.

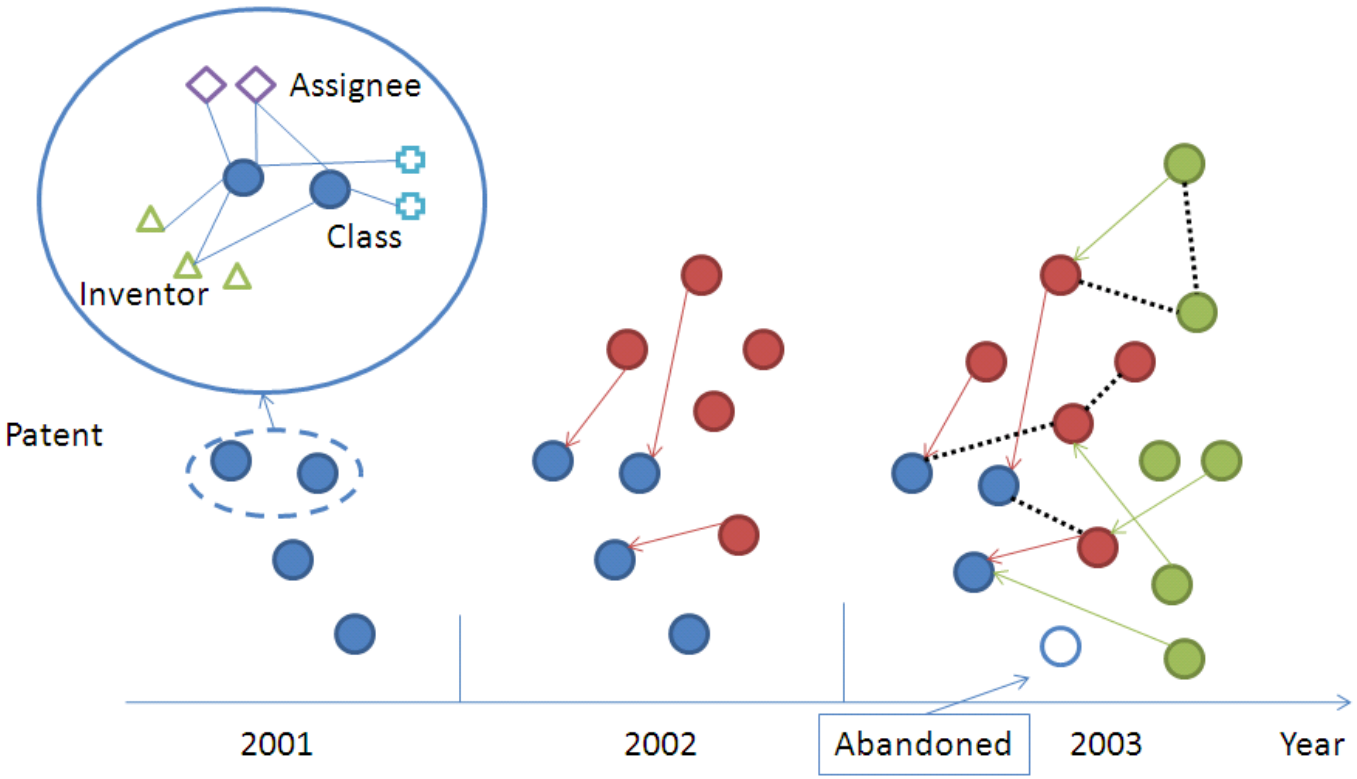
Recommendation System Description

1. Collection of patent database from USPTO
2. (optional) Given a query to select a subset of patents that are interesting to the user, such as a topic keyword, patent class etc.
3. Obtaining training data with historical maintenance decisions available
4. Building prediction model with network-based refinement
5. Computing the ranking score on the testing database and make recommendation for maintenance decision

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Patent Heterogeneous Time-evolving Information Network



$$G =< V, E, W, T >$$

- Node V: patents, inventors, assignees, USPC/IPC classes
- Edge E: patents are directly linked by citations (solid lines) and content similarity (dashed lines)
- W: weights for edges; T: time information

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Different types of features for patent information network

1. Meta Features
2. Novelty Features
3. Writing Quality Features
4. Trend Features
5. Inventor Features
6. Assignee Profile Features

Those features can capture the patent information in the network well to help companies to make maintenance decision.
The baseline is the traditional tf-idf weighting features.

Meta Features

Feature	Description
Pubyear	Publication year
Waityear	Pubyear minus filed year
NInv	Number of Inventors
NAss	Number of Assignees
NClaim	Number of Claims
NUSclass	Number of US classes
NIPC	Number of IPC codes
Kind	US patent kind
NFC	Number of Forward Citations
NBC	Number of Backward Citations
NOC	Number of Other Citations
NTC	Number of Total Citations

These informations are contained in each patent, and can be easily pulled out by software.

Novelty Feature

Observation and Novelty Feature Extraction

Observation:

- A novel patent tends to use more new or recent words/phrases.
- Novel patents have higher chance to be ruled as valid in court

How to extract?

- Extract top$-k$ key $n-$ grams
- Calculate the age of $n-$ grams (Phrase Age)
- Novelty features are computed as statistic summarization of the ages

Novelty Feature

Determine the Phrase Age [2] [3] [4]

Phrase age is defined with respect to a subject patent.

- Search for a given phrase in all patents that bear the same patent class
- If this search returns no earlier patent, then the age of the phrase is 0
- If an earlier patent is found, the difference between the publish dates is the age of the phrase.

Writing Quality Features

Patent Writing Quality in Two Aspects

- Complexity: a complex patent tends to use more words in order to describe the various aspects of the invention
- Completion: Based on US patent regulation, the patent claims need to be validated by complete description

Writing Quality Features

Feature	Description
TitleL	Title Length
TitleUW	Title Unique Words number
ClaimL	Claim Length
ClaimUW	Claim Unique Words number
FClaimL	First Claim Length
FClaimUW	First Claim Unique Words number
AbsL	Abstract Length
AbsUW	Abstract Unique Words
DesL	Description Length
DesUW	Description Unique Words
CohAC	Cohesion between Abstract and Claims
CohAD	Cohesion between Abstract and Description
CohCD	Cohesion between Claims and Description

Writing Quality Features

How to extract those features ?

- Extract the number of words and number of unique words in the claim, title, abstract and description
- Compute the pair-wise text similarities between abstract, claims and description as features to indicate the writing completion property
  - preprocessing: stop words removal, synonymy merging, stemming
  - tf*idf style weighting

Trend Features

Definition

Trend of a technique, indicated as a keyword or USPC/IP class

$$T =< f_1, \cdots, f_n >, f_i \in Z$$

- The frequency f_i is the number of related patents published in year i
- A trend T is a time ordered non-negative integer series

Trend Features

Definition

- Beginning of a trend

$$start(T) = \min\{i | f_i > 0\}$$

- Peak of a trend

$$\max(T) = max\{f_i\}$$

  The year that achieves $\max(T)$

$$maxyear(T) = \min\{i | f_i = \max(T)\}$$

- The overall volume

$$sum(T) = \sum_{i=1}^{n} f_i$$

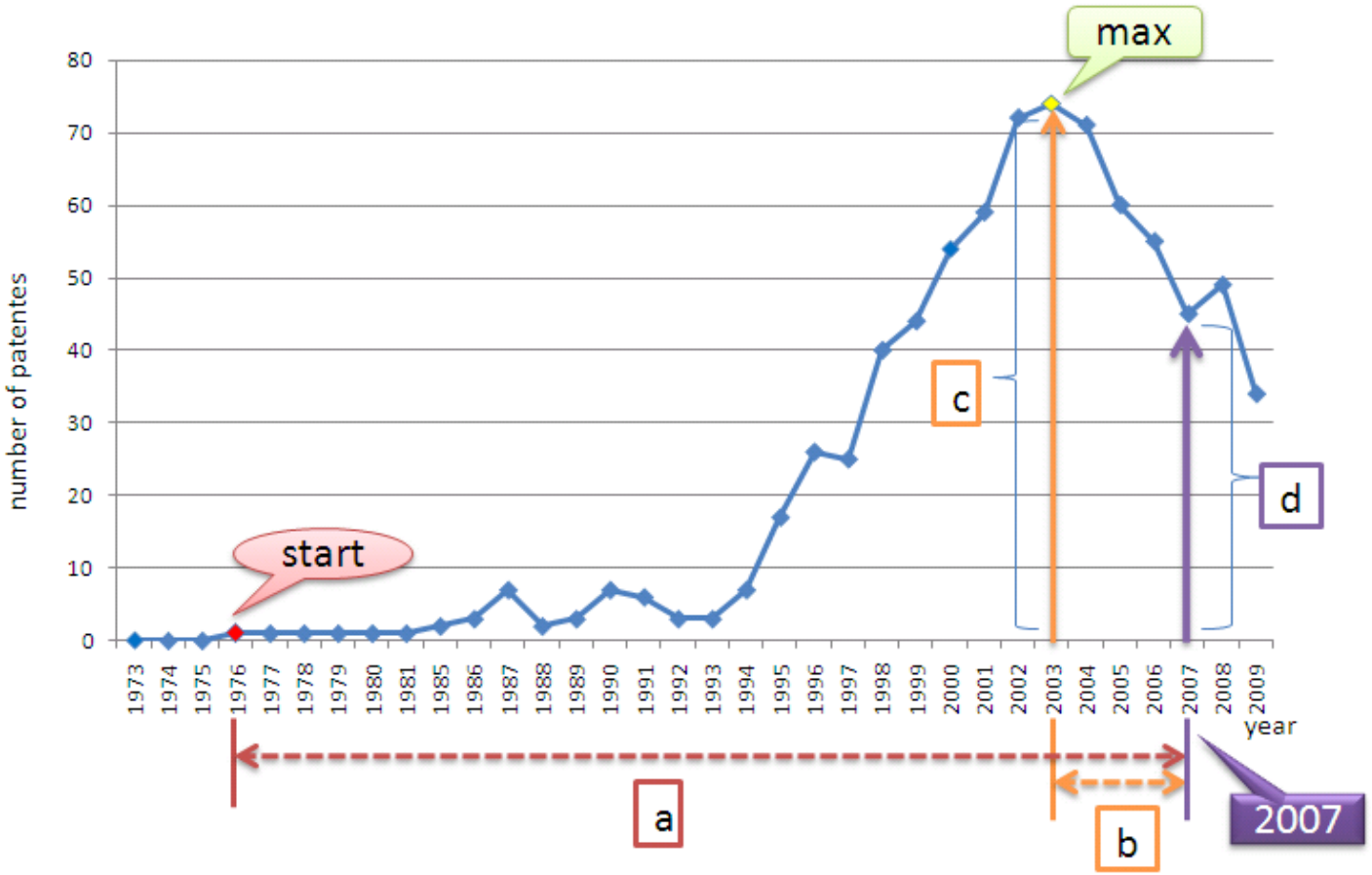
Trend Features

q is a time point (publication year)

1. $age(q, T) = q - start(T)$
2. $age2max(q, T) = q - maxyear(T)$
3. $thisSup(q, T) = f_q$
4. $maxSup(q, T) = \max(T)$
5. $sumSup(q, T) = sum(T)$
6. $thisMaxRatio(q, T) = \frac{f_q}{\max(T)}$
7. $thisAllRatio(q, T) = \frac{f_q}{sum(T)}$

Remark: Given a patent p and a particular technique c_i, those set of basic trend features can be obtained based on $T(c_i)$, the trend of c_i.

Trend Feature Example



Trend Features

Collective Trend Features

Given a set C of m techniques: $C = \{c_1, \cdots, c_m\}$.

$$MAage(p) = \max\{age(year(p), T(c_i))|c_i \in C\}$$

$$MIage(p) = \min\{age(year(p), T(c_i))|c_i \in C\}$$

$$SUage(p) = \sum_{i=1}^{m} age(year(p), T(c_i))$$

$$MEage(p) = \sum_{i=1}^{m} age(year(p), T(c_i))/m$$

Trend Features

Collective Trend Features

Two values for $year(p)$

- the year that the patent p was published
- the decision making year

Two sets of collective trend features:

- General set: T is generated from all related patents in the database. It reflects the overall trend of the technique
- Assignee-oriented set: T is formed by only considering the patents of a target company/organization
  Remark: It is possible that the general trend increasing while a company's trend is decreasing. The company has reduced its efforts on a particular technique and does not want to maintain related patents.

Inventor Feature

Observation: if an inventor's previous patents are all maintained, then it is more likely his/her recent patent will also be maintained

Given a patent p and its inventors $V = \{v_1 \cdots v_m\}$, where m is the number of inventors of p

1. $PCount(p, v_i)$: the number of patents of inventor v_i which were published before patent p
2. $ECount(p, v_i)$: the number of early expired patents of inventor v_i which were published before patent p
3. $EPratio(p, v_i) = \frac{ECount(p,v_i)}{PCount(p,v_i)}$
4. First Inventor Features: $iv1PCount$ $iv1ECount$ $iv1EPratio$
5. Collective Inventor Features: computing $\min$ $\max$ sum $mean$ value of $PCount(p, v_i)$ $ECount(p, v_i)$ $EPratio(p, v_i)$

Assignee Profile Features

Notation

- P_c: set of patents with c as its assignee
- P_c^y: set of patents published in year y

Assignee Profile $AP(c, y, w)$ is the collection of patents by assignee c which were published around year y within time window $w \in Z$.

$$AP(c, y, w) = \{p | p \in P_c, y - w \leq year(p) \leq y\}$$

Assignee Profile Features

1. Capture the overall similarity between patent p and assignee's profile

$$PAP(p, c, w)_{sum} = \sum_{p_i \in AP^p; p_i \neq p} Sim(p, p_i)$$

where $AP^p = AP(c, year(p), w)$ and $Sim(p, p_i)$ is the text similarity (cosine similarity) between p and p_i

2. Reduce the effect of profile size

$$PAP(p, c, w)_{mean} = \frac{1}{|AP^p|} \sum_{p_i \in AP^p; p_i \neq p} Sim(p, p_i)$$

3. Measure maximum relevance to patents in the profile

$$PAP(p, c, w)_{max} = \max\{Sim(p, p_i) | p_i \in AP^p\}$$

Special case features

Observe: It is possible that the assignee profile has a patent that is very similar to the query patent but all others are totally unrelated.

APK^p is the top k most similar patents to p

1. Capture the overall similarity between patent p and APK^p

$$PAPK(p, c, w)_{sum} = \sum_{p_i \in AP^p; p_i \neq p} Sim(p, p_i)$$

2. Reduce the effect of profile size

$$PAPK(p, c, w)_{mean} = \frac{1}{|APK^p|} \sum_{p_i \in AP^p; p_i \neq p} Sim(p, p_i)$$

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Initial Prediction

Remark: Each granted US patent will have at most three maintenance decisions.

Training instances generation

- For each patent, three instances will be generated.
- Each instance has a binary label: maintain or abandon.

Features (time sensitive) are using:

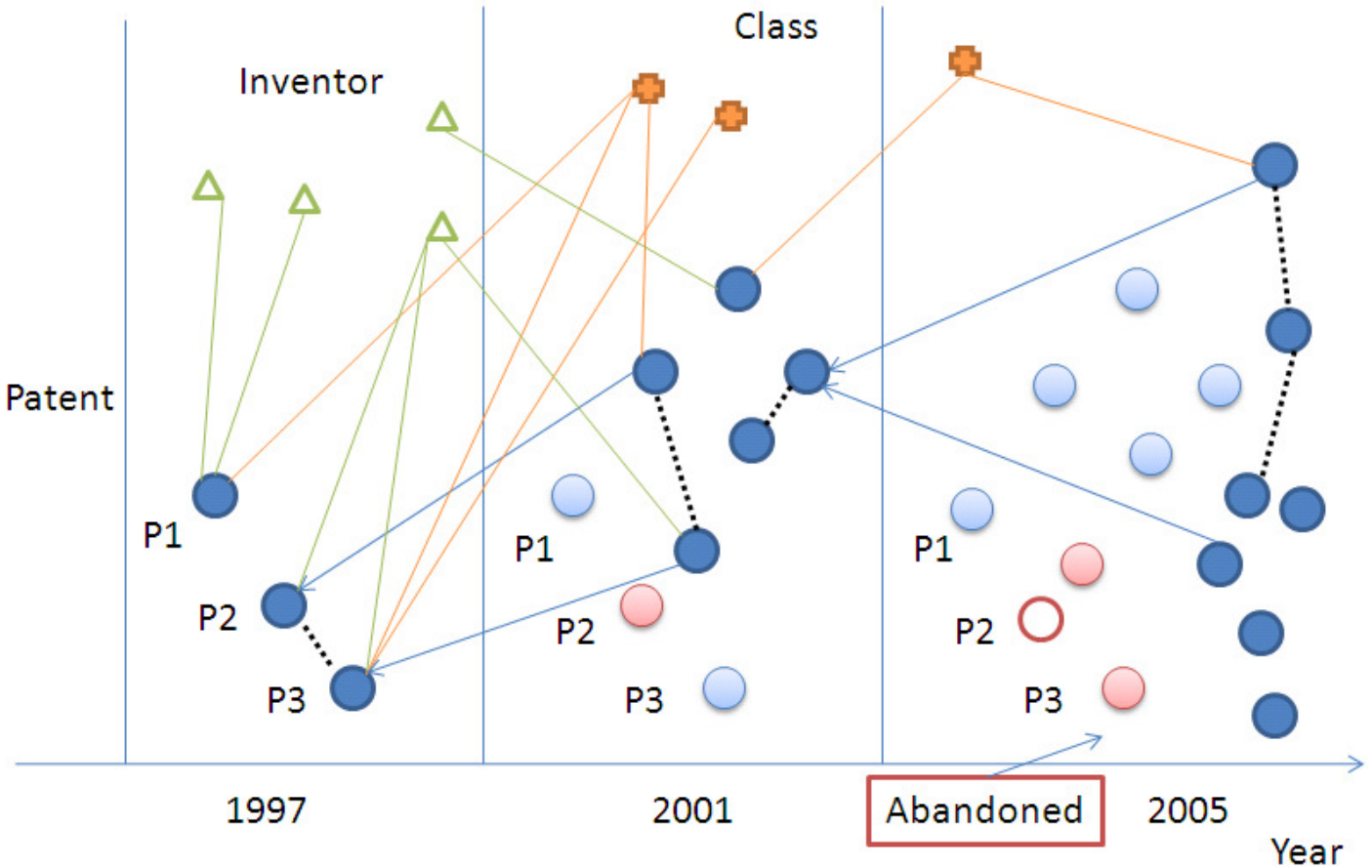
1. Trend Features
   - thisSup_min
   - age_min
   - thisMaxRatio_min
   - age2max_min
2. Profile Features
   - PAPsum
   - PAPKmean

Use SVM to learn a classification model.

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Network-based Refinement



Network-based Refinement

Notation

- V: Nodes
- V^A: inventor nodes
- V^C: USPC/IPC classification code nodes
- E: edges
- E_L: the set of edges formed by patent citations
- E_S: the set of edges modeling the similarity between patents
- E_A: the set of edges between patent and inventor
- E_C: the set of edges between patent and USPC/IPC class
- $w_{i,j}$: the patent text similarity score
- V^T: the set of training patent decision points
- V^P: the set of testing patent decision points
- Solid blue points: patents published in that particular year
- Light blue points: patents that are maintained
- Light red points: patents that are abandoned.
- Dashed edge: two patents are content similar

Optimization Framework

Prediction Scores $R = \{r_i\}(i = 1 \cdots |V|)$
The objective function:

$$
\begin{aligned}
O(R) = \alpha \sum_{i \in V^T} ||r_i - r_i^{\dagger}||^2 + \beta \sum_{i \in V^P} ||r_i - r_i^{\ddagger}||^2 \\
+ \gamma_L \sum_{<i,j> \in E_L} ||r_i - r_j||^2 + \gamma_S \sum_{<i,j> \in E_S} ||r_i - r_j||^2 \\
+ \delta_A \sum_{(i,a) \in E_A} \psi_{ai}^A ||r_i - r_a||^2 + \delta_C \sum_{(i,c) \in E_C} \psi_{ci}^C ||r_i - r_c||^2
\end{aligned}
$$

The goal is to find R to minimize $O(R)$.
r serves as the confidence score.

Optimization Framework

Notations and Parameter settings

Notations:

- ψ_{ai}^{A}: importance of inventor a to patent i; Maintenance percentage of the inventor's history.
- ψ_{ci}^{C}: importance of USPC/IPC class c to patent i; Maintenance percentage of class c.
- $r_i^{\dagger} \in \{0, 1\}$: decision for each training patent point $i \in V^T$
- $r_i^{\ddagger} \in [0, 1]$: initial prediction score for each testing patent
- $r_i = 1$: decision is to maintain; $r_i = 0$: decision is to abandon

Parameter setting:

$$\alpha = 10$$
$$\beta = 5$$
$$\gamma_L = \gamma_S = \delta_A = \delta_C = 1$$

Optimization Framework

Training Instances

$$\frac{\partial O(R)}{\partial t} = 2\alpha(r_t - r_t^\dagger) + 0 + 2\gamma_L \sum_{(t,j)\in E_L} (r_t - r_j) + 2\gamma_S \sum_{(t,j)\in E_S} w_{tj}(r_t - r_j) + 2\delta_A \sum_{(t,a)\in E_A} \psi_{at}^A (r_t - r_a) + 2\delta_C \sum_{(t,c)\in E_C} \psi_{ct}^C (r_t - r_c)$$

Let $\frac{\partial O(R)}{\partial t} = 0$,

$$r_t = \frac{\alpha r_t^\dagger + \gamma_L \sum_{(t,j)\in E_L} r_j + \gamma_S \sum_{(t,j)\in E_S} w_{ij} r_j}{Z_t} + \frac{\delta_A \sum_{(t,a)\in E_A} \psi_{at}^A r_a + \delta_C \sum_{(t,c)\in E_C} \psi_{ct}^C r_c}{Z_t}$$

Optimization Framework

Testing Instance

$$\frac{\partial O(R)}{\partial p} = 2\beta(r_p - r_p^{\ddagger}) + 2\gamma_L \sum_{(p,j)\in E_L} (r_p - r_j) + 2\gamma_S \sum_{(p,j)\in E_S} w_{pj}(r_p - r_j) + 2\delta_A \sum_{(p,a)\in E_A} \psi_{at}^{A}(r_p - r_a) + 2\delta_C \sum_{(p,c)\in E_C} \psi_{ct}^{C}(r_p - r_c)$$

Let $\frac{\partial O(R)}{\partial p} = 0$,

$$r_p = \frac{\beta r_p^{\ddagger} + \gamma_L \sum_{(p,j)\in E_L} r_j + \gamma_S \sum_{(p,j)\in E_S} w_{pj} r_j}{Z_p} + \frac{\delta_A \sum_{(p,a)\in E_A} \psi_{at}^{A} r_a + \delta_C \sum_{(p,c)\in E_C} \psi_{ct}^{C} r_c}{Z_p}$$

Optimization Framework

For inventor and classes

$$\frac{\partial O(R)}{\partial a} = -2\delta_A \sum_{(t,a)\in E_A} \psi_{at}^A (r_t - r_a) - 2\delta_A \sum_{(p,a)\in E_A} \psi_{ap}^A (r_p - r_a)$$

Let $\frac{\partial O(R)}{\partial a} = 0$

$$r_a = \frac{\sum_{(t,a)\in E_A} \psi_{at}^A r_t + \sum_{(p,a)\in E_A} \psi_{ap}^A r_p}{\sum_{(t,a)\in E_A} \psi_{at}^A + \sum_{(p,a)\in E_A} \psi_{ap}^A}$$

r_c is in the similar form for $c \in V_C$

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Two Experiment Achievement

1. Classification experiment
   - Feature Design
   - Feature Selection (Choose important features to use)
   - Binary Classification: abandon or maintain
   - Performance: classification accuracy for each company
2. Obtain Confidence Score
   - Form the patent information network
   - Solve optimization problem and obtain the confidence score r for each company
   - Performance: classification accuracy vs confidence score

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Dataset

- 4 millions granted patents from USPTO
- Major companies: BASF, Boeing, Kodak, General Motor, General Electric
- U.S. Department of Energy, NASA, Navy and Army

Outline

1. Patent Maintenance Motivation and System Description
   - References and Related Work
   - Patent Maintenance Motivation
   - System Description
2. Patent Information Network and Feature Design
   - Patent Information Network
   - Feature Design
3. Prediction Model
   - Initial Prediction
   - Patent Information Network Refinement
4. Experiments
   - Main Achievements in Experiments
   - Dataset
   - Feature Selection

Importance of Features

Probabilistic Significance Score [5]

Use Probabilistic Significance score to rank the features and choose the top features.

Probabilistic Significance Score for features comes from . For binary classification problem, the score for feature j is

$$S(j) = \int |\mathbb{P}(y = 1|\mathbf{x}) - \mathbb{P}(y = 1|\mathbf{x}^{-j})|p(\mathbf{x})d\mathbf{x}$$

- $y \in \mathcal{Y} = \{+1, -1\}$ (abandon or maintain)
- $\mathbf{x}$ Features
- $\mathbb{P}(y = 1|\mathbf{x})$ represents the true probability that the sample $\mathbf{x}$ belongs to class 1
- $\mathbb{P}(y = 1|\mathbf{x}^{-j})$ represents the true probability that the sample $\mathbf{x}^j$ (removing j^{th}) belongs to class 1.

Importance of Features

Probabilistic Significance Score Estimation [5]

$$\hat{S}_S(j) = \frac{1}{n} \sum_{i=1}^{n} |\hat{\mathbb{P}}(y = 1|\mathbf{x}_i) - \hat{\mathbb{P}}(y = 1|\mathbf{x}_i^{-j})|$$

Features that cause significant changes in the estimated class probabilities are ranked higher than those that do not.

Probabilistic Significance of Top Features

Type	Top Features	Probabilistic Significance
Inventor	ivEPratio_mean	0.490
	iv1EPratio_max	0.426
	iv1ECount_max	0.349
Trend	thisSup_min	0.319
	age_min	0.177
	thisMaxRatio_min	0.136
	age2max_min	0.125
Profile	PAPsum	0.166
	PAPKmean	0.138
Writing	CohAD	0.175
	AbsL	0.123
	ClaimUW	0.122
	CohCD	0.121
Meta	Kind	0.151
	NClaim	0.078

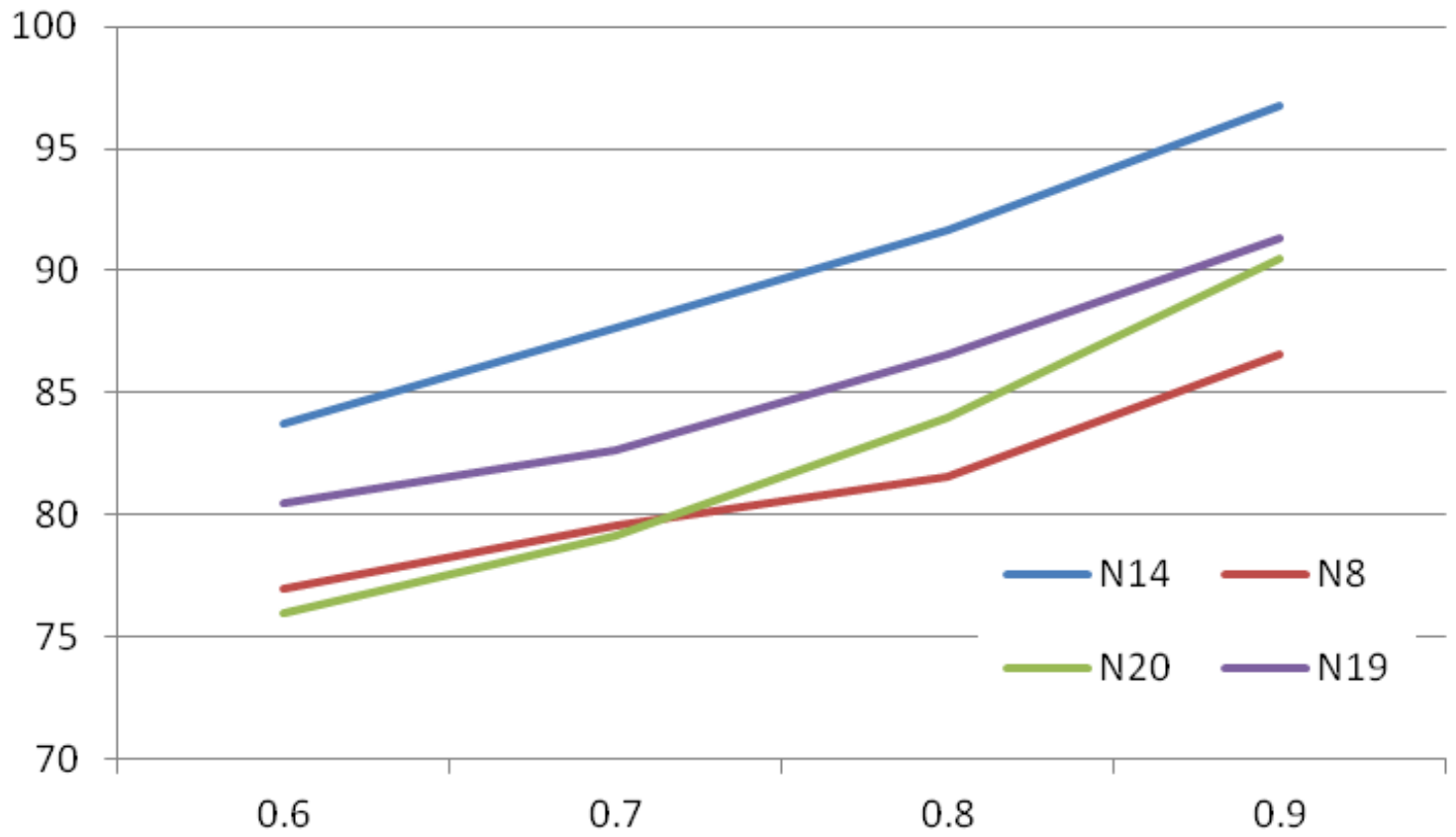
Accuracy Comparison in Classification Experiment

Baseline:

1. Probability Guess: uses the maintenance probability of the training data to guess the label of the testing data
2. TextC: Use term vector space features for patent representation and use K-Nearest Neighbor to classify

Authors's approach: Using the important features selected by Probability Significance method. Use KNN to classify

Classification Accuracy vs Confidence Score



- N14, N8, N20, N19 are companies's ID
- Y-axis denotes the classification accuracy
- X-axis denotes the confidence score.

References

[1] 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," in *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, (Vancouver,BC), pp. 280–289, December 2011.

[2] M. A. Hasan, W. S. Spangler, T. Griffin, and A. Alba, "COA: Finding Novel Patents through Text Analysis," in *KDD '09 Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, (Paris, France), pp. 1175–1184, June-July 2009.

[3] R. Kosala and H. Blockeel, "Web Mining Research: A Survey," *ACM SIGKDD Explorations Newsletter*, vol. 2, pp. 1–15, June 2000.

[4] I. De, "Experiments in First Story Detection," in *Proceedings of the 2005 National Conference on Undergraduate Research (NCUR)*, 2005.

[5] N. A. Arnosti and A. P. Danyluk, "Feature Selection via Probabilistic Outputs," in *Proceedings of the 29th International Conference on Machine Learning (ICML 2012)*, 2012.