기술경영경제학회 동계학술대회

Seek your fortune!: Technological impact-guided technology opportunity analysis using generative-predictive machine learning models

이규민[1], 이창용[2,*]

2024. 02. 02. (금)

1) 고려대학교 행정학과 정부학연구소 연구원(glee.optimizt@gmail.com)
2,*) 고려대학교 행정학과 교수, 교신저자(changyonglee@korea.ac.kr)

Contents

1. Introduction

Introduction

❖ Background

- Increased academic attention to technology opportunity analysis (TOA)
  - Enables firms to foster economic growth, sustain competitiveness, and mitigate uncertainties involved in new businesses
- Research progression of TOA: Expert-centric approaches → Data-driven approaches
  - Mainly based on patent analysis
  - Effective for identifying new technology ideas that have not yet been fully explored
- Lack of practicality of the previous approaches
  - Limited resources and complexities associated with technology development
  - Firms' preference for maximising the potential of their existing technologies, over developing new, potentially risky technologies

1. Introduction

Introduction

❖ Motivation

- The impact of a technology can vary depending upon the domains where it is applied
- A shift in technological domains could lead to promising technology opportunities
  - Many significant inventions derived from the domain shift
- Example: Autographer
  - Hands-free, wearable intelligence camera with sensors for deciding when to automatically take photos
  - Originally intended to be used in the healthcare domain to assist memory-impaired patients
  - Later repurposed for consumer use, as life-logging device for common people

Autographer

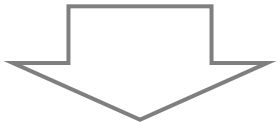
❖ Purpose

- To develop an analytical framework that identifies new technological domains where existing technologies may yield higher technological impacts using generative-predictive machine learning models

2. Background

Patent analysis-based TOA

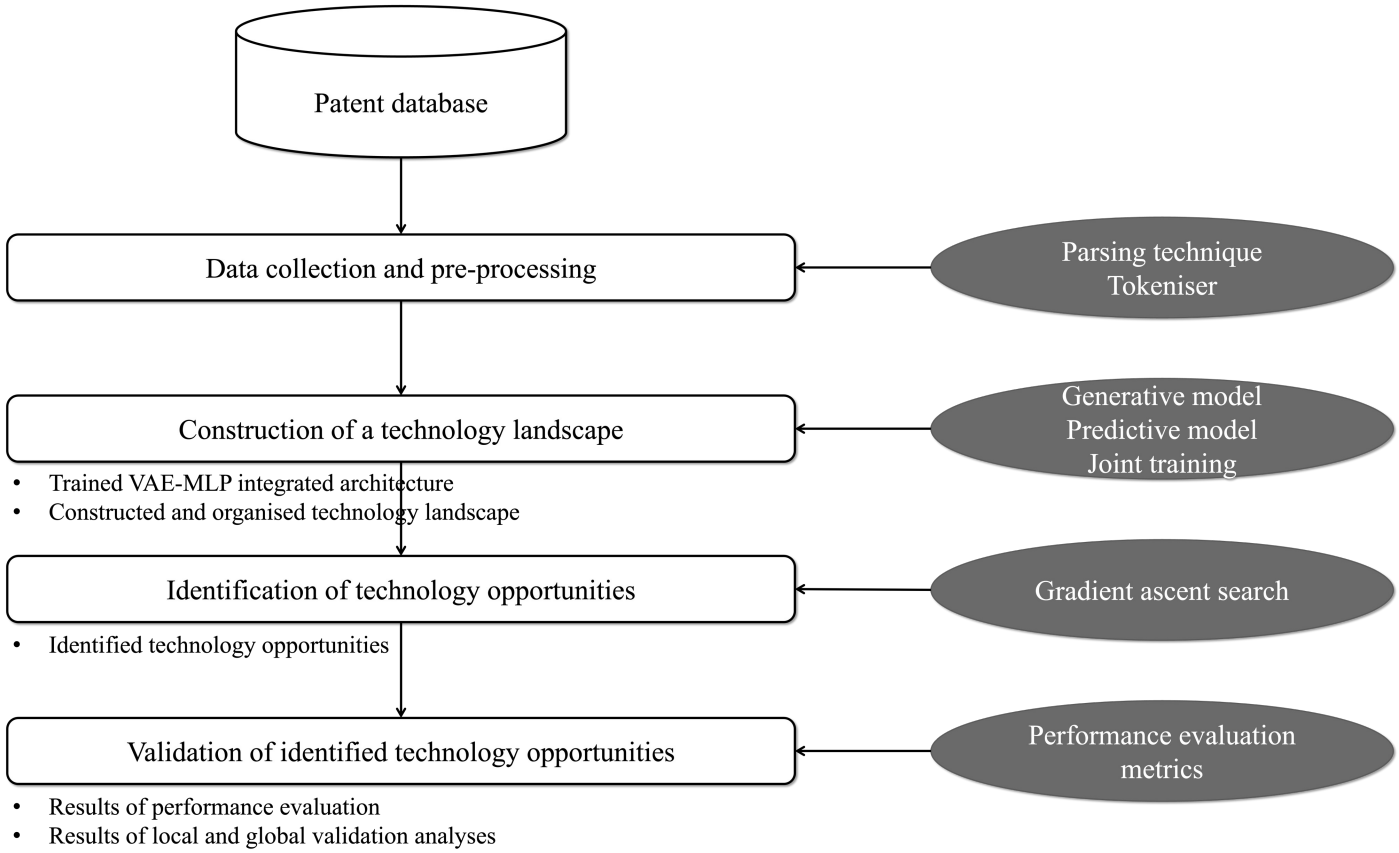
- **Patent mapping** (Lee et al., 2009)
  - Constructing two-dimensional patent map and discerning vacancies as emerging technology opportunities
- **Morphological analysis** (Yoon and Park, 2005)
  - Decomposing a complex technology system into separate dimensions and identifying unoccupied configurations as technology opportunities
- **Recombinant search** (Lee and Lee, 2019)
  - Building a patent landscape and evaluating novelty and value of areas in the landscape to identify technology opportunities



Cannot provide insights into new technological domains where existing technologies may yield higher technological impacts

3. Data and methodology

Research framework

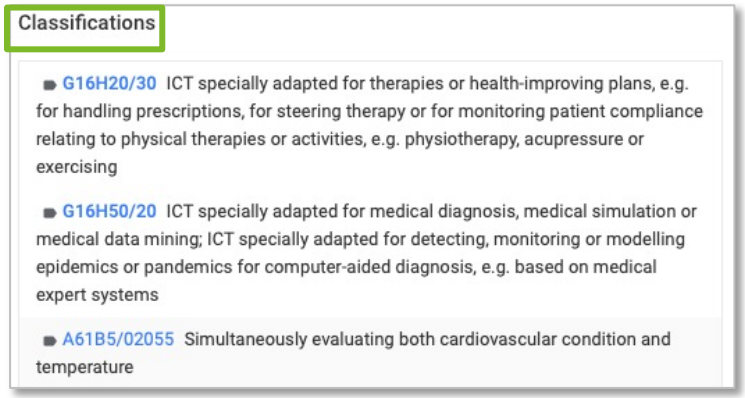


3. Data and methodology

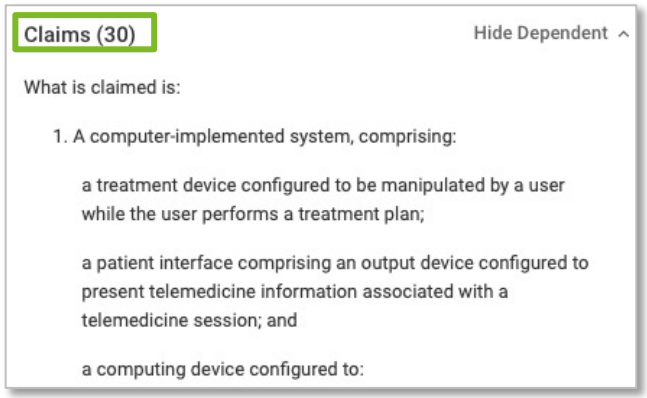
Data collection and pre-processing

❖ Patent data

- Technological domains
  - Patent class information → **Sequence of IPC (International Patent Classification)**
- Technological functions
  - Patent claim information → **Claim texts**
- Technological impact of patented inventions
  - Number of forward citations → **L1 (breakthrough) vs. L2 (common)**
    - L1 threshold: top 10% forward citation counts within a specific technology field



Technological domains



Technological functions

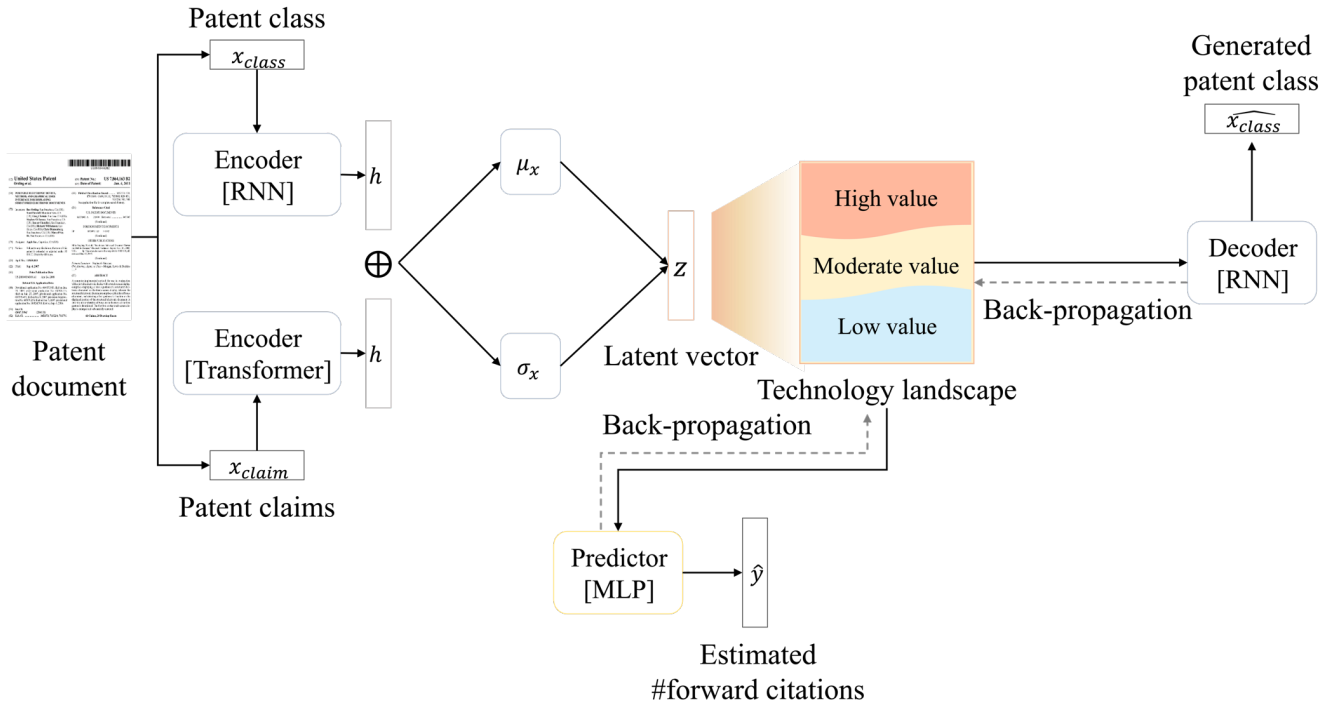
Technological impact

3. Data and methodology

Construction of a technology landscape

❖ VAE-MLP integrated architecture

- Joint training
  - $\boldsymbol{Total\ loss} = CrossEntropy(x_{class}, \widehat{x_{class}}) + D_{KL}\big(N(\mu_x, \sigma_x) \parallel N(\mathbf{0}, \mathbf{1})\big) + CrossEntropy(y, \hat{y})$
    [Reconstruction loss] [Regularisation loss] [Prediction loss]
- Model structure

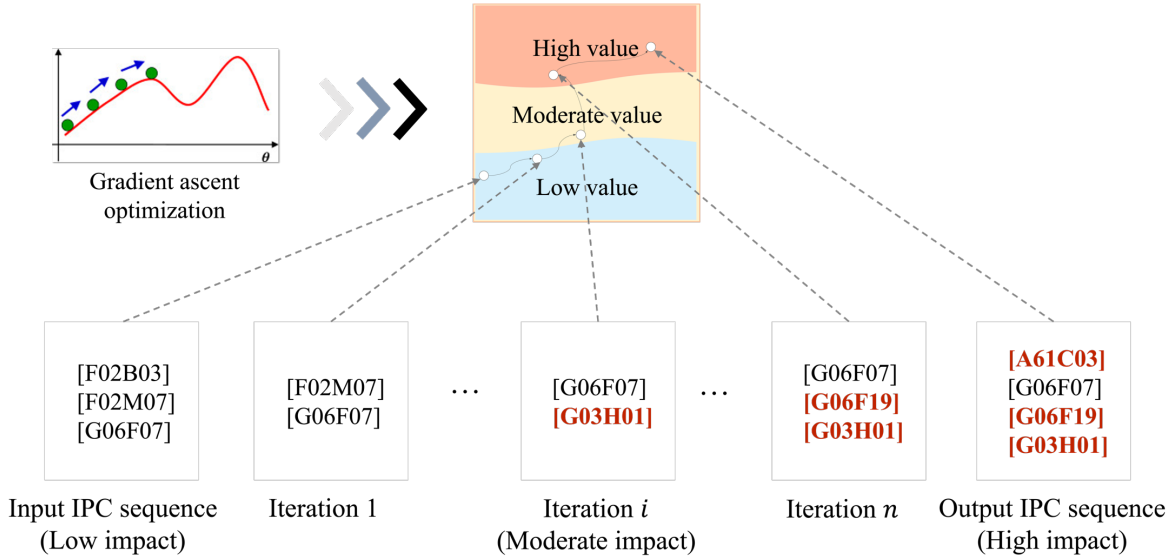


3. Data and methodology

Identification of technology opportunities

❖ Technology landscape exploration

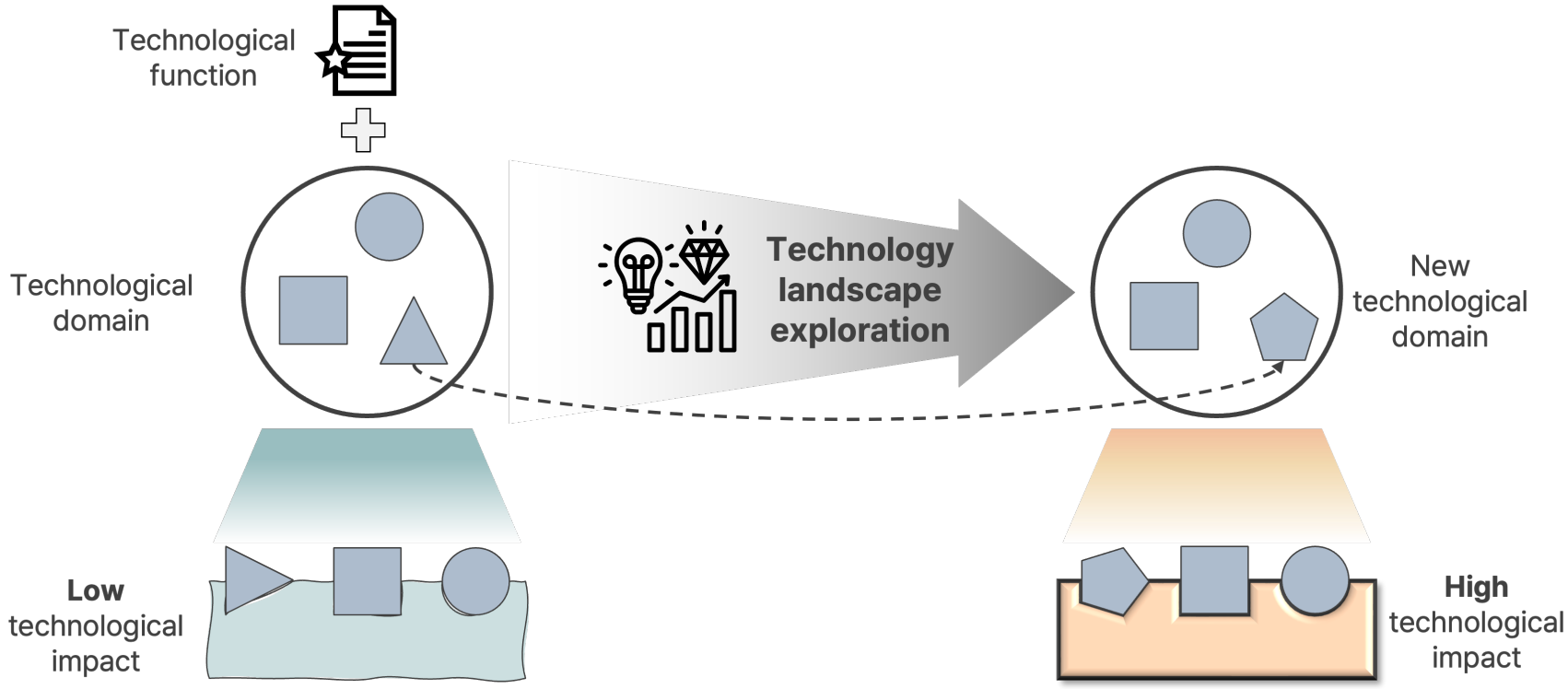
- Using gradient ascent search
  - To update the position of the focal latent vector in landscape
- MLP predictor as the target function
  - Estimating the probability of L1 technological impact label for the focal latent vector
- Iteration termination condition
  - The estimated L1 probability > The desired L1 probability



3. Data and methodology

Identification of technology opportunities

❖ Technology landscape exploration



3. Data and methodology

Validation of identified technology opportunities

❖ Reliability of the proposed analytical framework

- **Generative** performance evaluation
  - Comparison between the actual IPC sequence and the generated IPC sequence
  - Jaccard similarity
    - $Jaccard\ Similarity = \frac{|A \cap B|}{|A \cup B|}$
- **Predictive** performance evaluation
  - Comparison between the actual and predicted technological impact labels
  - Performance evaluation metrics for classification task
    - $Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$
    - $Precision = \frac{tp}{tp+fp}$
    - $Recall = \frac{tp}{tp+fn}$
    - $F1-score = 2 \cdot \frac{precision \cdot recall}{precision+recall} = \frac{2tp}{2tp+fp+fn}$

		Actual technological impact labels	
		Positive	Negative
Predicted technological impact labels	Positive	tp	fp
	Negative	fn	tn

Confusion matrix

3. Data and methodology

Validation of identified technology opportunities

❖ Feasibility of the proposed analytical framework

- **Local validation analysis** using patent citation relationships
  - Notations

Term	Definition	Term	Definition
$Patent^{i}_{original}$	A patented invention with patent number i	$IPC^{i}_{identified}$	IPC sequence of the **new technological do main** identified for $Patent^{i}_{original}$
$IPC^{i}_{original}$	IPC sequence of $Patent^{i}_{original}$	$Patent_{citing,shifted}$	The citing patents assigned to the IPC sequence that is **different from** $IPC^{i}_{original}$
$Patent^{i-j}_{citing}$	A specific citing patent with patent number j that cited $Patent^{i}_{original}$	$Patent_{citing,matched}$	The citing patents assigned to the IPC sequence that is **same** with $IPC^{i}_{identified}$
IPC^{i-j}_{citing}	IPC sequence of $Patent^{i-j}_{citing}$	$\{Patent_{citing,shifted,matched}\}^{i}$	The citing patents assigned to the IPC sequence that is **different from** $IPC^{i}_{original}$ and **same** with $IPC^{i}_{identified}$

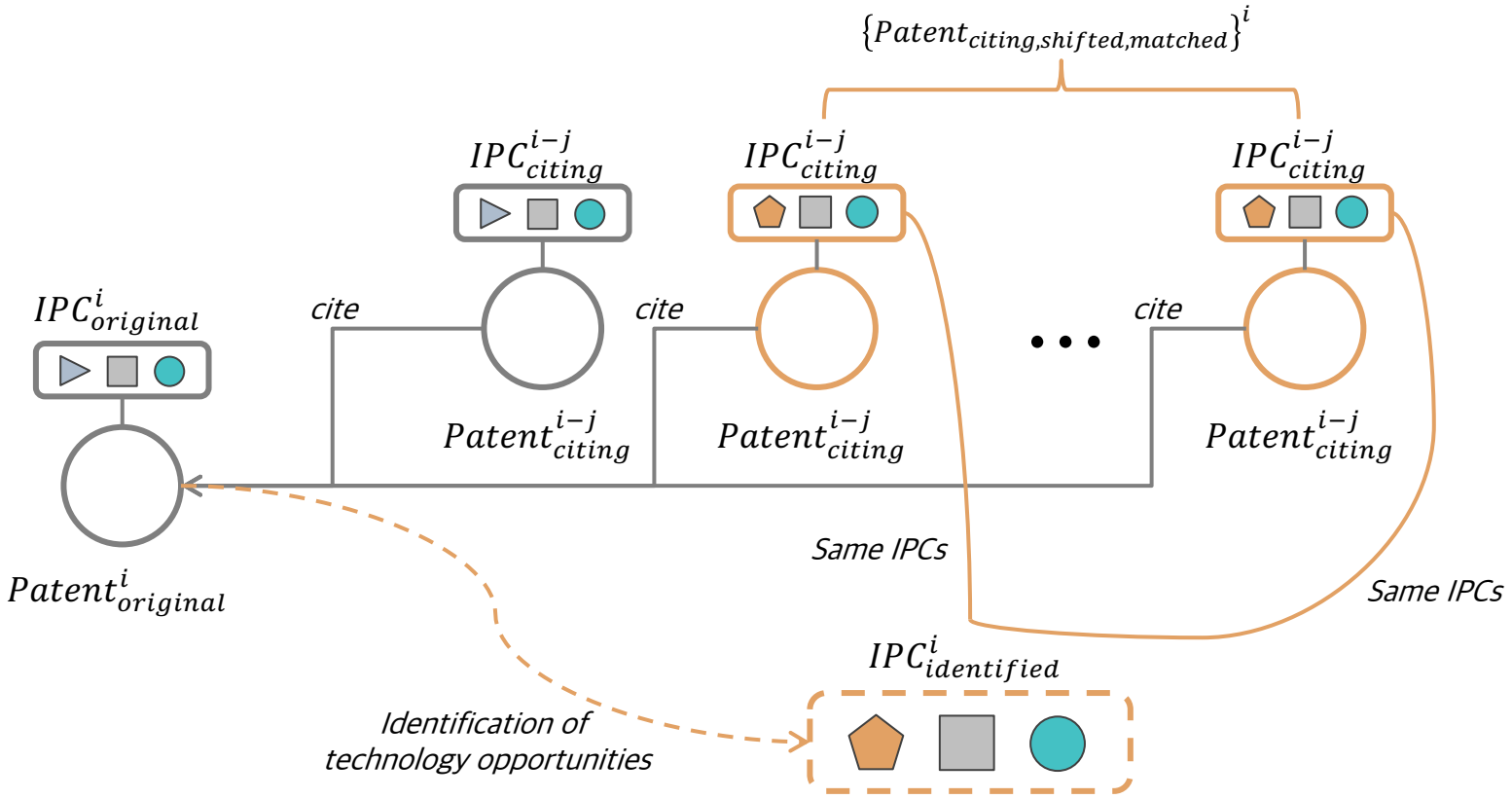
  - A patent cites a previous invention when it is developed by **improving or modifying** the technological functions of the original patent (i.e., prior art)
  - $\{Patent_{citing,shifted,matched}\}^{i}$ → Realisation of the technology opportunities based on **domain shift**
  - Validating the proposed analytical framework to identify **practical and valuable** technology opportunities

3. Data and methodology

Validation of identified technology opportunities

❖ Feasibility of the proposed analytical framework

- **Local validation analysis** using patent citation relationships



3. Data and methodology

Validation of identified technology opportunities

❖ Feasibility of the proposed analytical framework

- **Global validation analysis** using patent indicators
  - Comparing the technological impact, originality, and market coverage between the $IPC_{original}$ and $IPC_{identified}$
    - By calculating patent indicators for the patents assigned to the $IPC_{original}$ and $IPC_{identified}$
  - Patent indicators

Patent indicator	Operational definition	References
Technological impact (TI)	Number of forward citations of the patents over five years	Lerner, 1994; Narin et al., 1987
Technological originality (TO)	Herfindahl index on patent classes of cited patents	Trajtenberg et al., 1997
Market coverage (MC)	Number of patents registered in multiple countries with t he coverage of the same invention	Guellec and de la Potterie, 2000

4. Empirical analysis and results

Data collection and pre-processing

❖ Data collection - Patents associated with AI technology

- Using USPTO and Patentsview
- Search query
  - Period: **2007-2012**
  - Patent class: **G06F**
  - Keywords: machine, unsupervised, supervised, reinforcement, learning, data, neural network, ⋯
- A total of 189,915 patent documents related to AI technology

❖ Data pre-processing

- Patent class information
  - **Main group IPC codes** → IPC sequence
- Patent claims
  - **First claim** → Claim sequence
- The number of forward citations
  - **Forward citation counts within 5 years** after a patent is granted → L1/L2
  - L1 threshold: **14** (top 10% forward citation counts within the field of AI technology)
- A total of **69,050** data samples remained after pre-processing
  - Dataset split: Training 70%, Validation 20%, Test 10%

4. Empirical analysis and results

Construction of a technology landscape

❖ Results of generation and prediction

Patented inventions			Model outputs	
Patent number	**Patent classes (IPC sequence)**	**True label (#forward citations)**	**Generated patent classes (IPC sequence)**	**Predicted label (L1 probability)**
8068433	G08C17, H04L12, H04L29	L2 (10)	G08C17, H04L12, H04L29	L2 (0.3962)
8108543	G06F13, G06F15	L1 (44)	G06F13, G06F15	L1 (0.5221)
7739999	F02B03, F02M07, G06F07	L2 (5)	F02B03, F02M07, G06F07	L2 (0.1750)
…	…	…	…	…
7164981	F16H61, G06F07	L2 (4)	F16H61, G06F07	L2 (0.2358)
7392483	G06F03, G06F17	L1 (54)	G06F03, G06F17	L1 (0.5859)
8001433	G01R31, G06F17	L2 (2)	G01R31, G06F17	L2 (0.1971)

4. Empirical analysis and results

Identification of technology opportunities

❖ Technology landscape exploration

- L1 probability criterion: Desired probability for the L1 technological impact label → **0.5**

Patented inventions		**Identified technology opportunities**						
Patent number	**Patent classes (True label)**	**Generated IPC sequence (L1 probability)**						
		Iteration 1	Iteration 2	Iteration 3	…	Iteration (N-2)	Iteration (N-1)	Iteration N
8068433	G08C17, H04L12, H04L29 (L2)	G08C17, H04L29 (0.5280)	G08C17, H04L29, **H04W36** (0.6139)	G08C17, H04L29, **H04Q07** (0.6921)	…	**H04W36, H04W24** (0.9545)	**H04W36, H04W24** (0.9562)	**G01S03, H04W36, H04W24** (0.9578)
8108543	G06F13, G06F15 (L1)	G06F13, G06F15 (0.6133)	G06F13, G06F15 (0.6678)	G06F13, G06F15 (0.6921)	…	G06F13, G06F15, **H04M11** (0.9171)	G06F13, G06F15, **H04N05** (0.9199)	G06F13, **H04B01** (0.9226)
…	…	…	…	…	…	…	…	…
7392483	G06F03, G06F17 (L1)	G06F03, G06F17 (0.7054)	G06F03, G06F17 (0.7898)	G06F03, G06F17 (0.8288)	…	G06F03, G06F17 (0.9597)	G06F03, G06F17 (0.9611)	G06F03, G06F17 (0.9623)
8001433	G01R31, G06F17 (L2)	G01R31, G06F17 (0.3768)	G01R31, G06F17 (0.5846)	G06F17, **G01C21** (0.6355)	…	**G09G05, G06F03, G01C21** (0.9543)	**G09G05, G06F03, G01C21** (0.9560)	**G09G05, G06F03, G01C21** (0.9576)

4. Empirical analysis and results

Validation of identified technology opportunities

❖ Performance evaluation results

- For a total of **6,905** patented inventions in test set

TEST_SET	Support	Accuracy	Precision	Recall	F1-score	Jaccard similarity
L2	6232	-	0.9296	0.7667	0.8403	
L1	673	-	0.1762	0.4621	0.2551	
macro-averaged	6905		0.5529	0.6144	0.5477	**0.8323**
micro-averaged	6905	**0.7370**	0.7370	0.7370	0.7370	
weighted-averaged	6905		**0.8561**	**0.7370**	**0.7833**	

4. Empirical analysis and results

Validation of identified technology opportunities

❖ Local validation analysis results

- Descriptive statistics
  - $\{Patent_{citing,matched}\}^{i}$

#forward citations	Count	Average	Standard deviation	Minimum	Median	Maximum
Absolute value	1353	7.8765	16.2325	0	3	202
Percentile rank		0.6056	0.2845	0.0321	0.6026	1.0000

  - $\{Patent_{citing,shifted,matched}\}^{i}$

#forward citations	Count (over L1 threshold)	Average	Standard deviation	Minimum	Median	Maximum
Absolute value	49 **(12)** - 24%	15.8776	29.6231	0	5	139
Percentile rank		0.6689	0.3036	0.0758	0.7500	1.0000

*percentile rank is the rank of $Patent^{i-j}_{citing,matched}$ in $\{Patent_{citing}\}^{i}$

4. Empirical analysis and results

Validation of identified technology opportunities

❖ Global validation analysis results

- Patent indicator analysis

$Patent_{original}$	IPC sequence		Average TI	Average TO	Average MC
7922575	Original patented invention	G06F17, G06F19	4.2776	0.6156	7.1448
	Identified technology opportunity	G06F17	5.0802	0.4137	5.3354
8060267	Original patented invention	B60W10, G06F17	3.4500	0.6282	7.4783
	Identified technology opportunity	G06Q40, G06F17	3.1500	0.5738	6.6767
…	…		…	…	…
7673080	Original patented invention	G06F13	3.0627	0.3924	6.7798
	Identified technology opportunity	G06F12, G06F13	3.6773	0.3005	6.2209
Overall (averaged)	**Original patented inventions ($IPC_{original}$)**		**6.1409**	**0.5150**	**6.3733**
	Identified technology opportunities ($IPC_{identified}$)		**6.2577**	**0.5111**	**6.9632**

4. Empirical analysis and results

Validation of identified technology opportunities

❖ **Example of $Patent_{original}^{7600135}$-$Patent_{citing,shifted,matched}^{7600135-7844839}$ pair**

- The citing patent expanding the application scope of the existing patented invention

$Patent_{original}^{7600135}$

❖ Title: "Apparatus and method for software specified power management performance using low power virtual threads"

❖ Patent classes
- G06F01: "Details not covered by groups G06F03-G06F13 and G06F21"
- G06F09: "Arrangements for program control"

❖ #forward citation (over 5 years): **11**

forward citation

$Patent_{citing,shifted,matched}^{7600135-7844839}$

❖ Title: "Distribution of network communications based on server power consumption"

❖ Patent classes
- G06F01: "Details not covered by groups G06F03-G06F13 and G06F21"
- G06F09: "Arrangements for program control"
- **G06F15: "Digital computers in general"**

❖ #forward citation (over 5 years): **35**

5. Discussion and conclusion

Key contributions and limitations

❖ Key contributions

Theoretical perspective

- ❖ Incorporating evaluation of **domain-specific technological impact** into TOA research
- ❖ Use of **generative models**
  → Identifying realistic technology opportunities from a large search space

Practical perspective

- ❖ Offering wider range of **exploration** and reliable **generation**
- ❖ Providing an **automated software system** for the identification of technology opportunities
  → Enhancing decision-making for new technology development

❖ Limitations

- ❖ Need of employing a lower level of patent classes (e.g., subgroup) and greater number of patent claims
- ❖ Need of improving the predictive performance
- ❖ Need of utilising diverse non-technological factors associated with technology development and commercialisation
- ❖ Need of a further qualitative validation analysis by practitioners

References

Guellec, D., Potterie, B. van P. de la, 2000. Applications, grants and the value of patent. Economics Letters, 69, 109–114.

Lee, C., Lee, G., 2019. Technology opportunity analysis based on recombinant search: patent landscape analysis for idea generation. Scientometrics, 121, 603-632.

Lee, S., Yoon, B., Park, Y., 2009. An approach to discovering new technology opportunities: Keyword-based patent map approach. Technovation, 29, 481-497.

Lerner, J., 1994. The Importance of Patent Scope: An Empirical Analysis. The RAND Journal of Economics, 25, 319.

Narin, F., Noma, E., Perry, R., 1987. Patents as indicators of corporate technological strength. Research Policy, 16, 143–155.

Trajtenberg, M., Henderson, R., Jaffe, A., 1997. University Versus Corporate Patents: A Window On The Basicness Of Invention. Economics of Innovation and New Technology, 5, 19-50.

Yoon, B., Park, Y., 2005. A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. Technological Forecasting and Social Change, 72, 145-160.

Thank you

이규민, 이창용

2024. 02. 02. (금)

INSTITUTE of GOVERNMENTAL STUDIES
IGS
KOREA UNIVERSITY

KOSIME 기술경영경제학회