Identifying the latent technology opportunities based on a perspective of coupling publications with patents

Munan LI [1, 2] Wenshu Wang [1] Keyu Zhou [1]

1. School of Business Administration, South China University of Technology

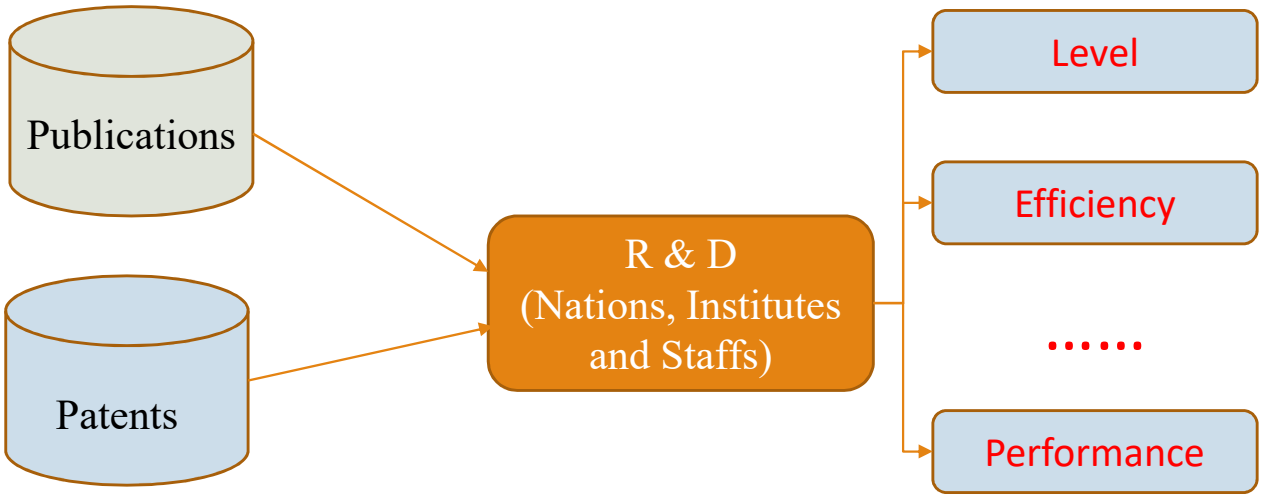
1. Guangdong Province Key Laboratory on Decision & Innovation Management

2019/10/20

Outline

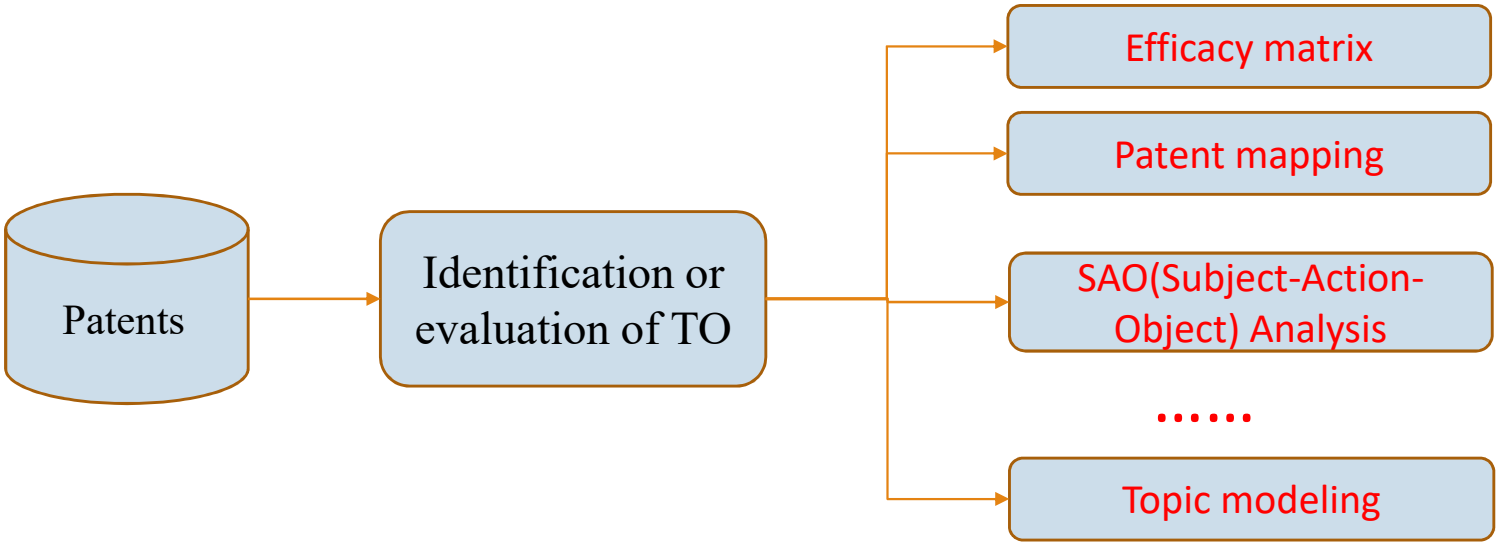
Introduction

- Within a specific topic, the publications and patents are the outputs of scientific research and technology transformation respectively, and also they are often utilized to evaluate the levels of R&D.



Introduction

- Under the traditional framework on identifying the TO (technology opportunity), the patents are the usually data source. And, the classic and emerging analytical methods on patents are the mainstream of current studies.



Introduction

• Research questions

- **RQ1:** Within some topics or under the specific situations, the coupling relations between publications and patents can be conducted to identify the latent technology opportunities?
- **RQ2:** And how can we use these coupling relations between publications and patents to identify TO?
- **RQ3:** The new TO based on the proposed coupling methods can be identified by the other methods or can be compared with the other methods including traditional or emerging?

Related work

- The driving force of technology progress has been generally discussed since 1950s , and it should be endogenous from the view of productivity development (Solow, 1957). In consequence, to forecast the evolution of technology, or technology foresight has become one of critical approaches for strategic planning and industrial-policy making (Martin, 1995; Hasan and Tucci, 2010; Miles, 2010; Li et al., 2018).

- In general, the data of publications is often utilized to evaluate the scientific output and academic performance, efficiency or collaboration between different institutes or scientists, and so forth (Guenter et al., 2007; Kyle et al., 2015).

- Some latest studies have integrated social media data, e.g. Twitter, Facebook or WeChat into the framework of technological forecasting and TOA (Chen, 2018).

Related work

- Regarding the relevant studies on TOA (technology opportunity analysis), technology forecasting, emerging technologies, TCE(technological capability evaluation) and tech mining etc., the patent data play the key role in those processing (Breitzman & Mogee, 2002; Tseng et al., 2007; Youtie et al, 2008; Lee et al, 2009; Porter & Newman, 2011; Zhang et al, 2016; Wang et al., 2017).

- Also, the technique on combing bibliometrics with patent analysis has been conducted to **forecast those emerging technologies** or **R&D evaluation** (Hullmann and Meyer, 2003; Daim et al., 2006).

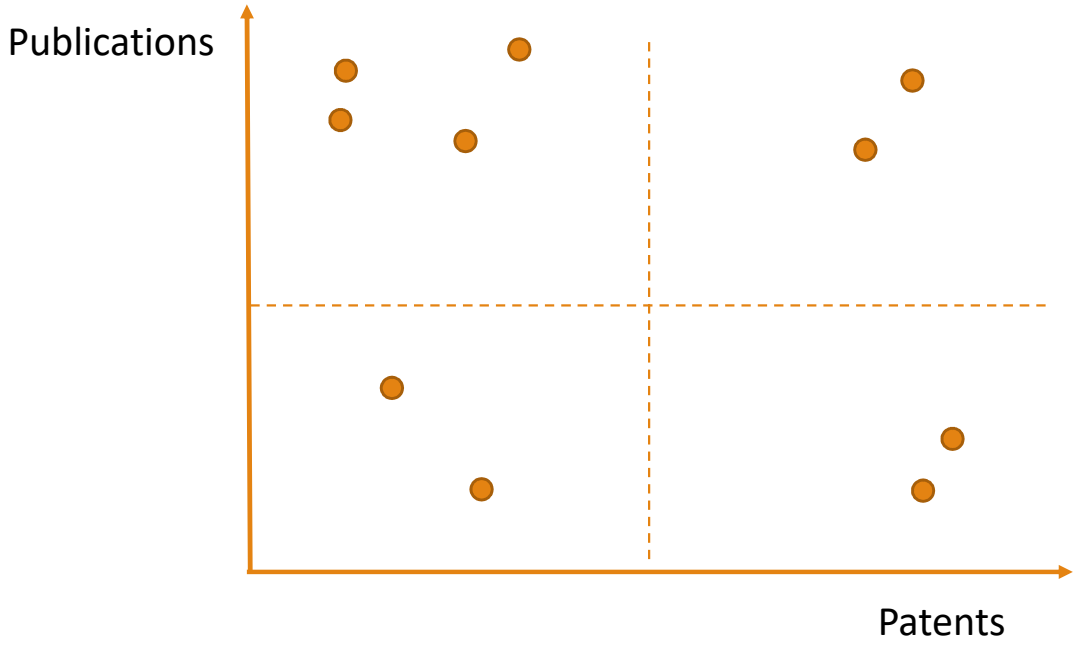
Related work

Bibliometrics & publications data	Patent data & Patent analysis	Bibliometrics & Patent analysis	Multiple data source analyses
Scientific evaluation on output, efficiency, performance and international collaborations and so forth.	TOA (Technology Opportunity Analysis), TCE (Technology Capability Evaluation), Technology Evolution Analysis and so on.	Forecasting emerging technologies, Evaluating the performance or efficiency of R&D	Social medias data(Twitter, Facebook, WeChat Internet data (homepages, search results and so on)

Basically, the general computing-framework based on the coupling relations between publications and patents is still insufficient, especially for those multi-disciplinary or trans-disciplinary topics, for example, artificial intelligence, smart cities and Industrial Internet etc.

Methodology

Here, an hypothesized framework is proposed to forecast the latent technology opportunities based on the coupling relations between scientific articles and corresponding patents within a specific topic.



Publications	Patents	Approximate classification
More	More	High R(Research) & High D(Development)
More	Less	High R(Research) & Low D(Development)
Less	More	Low R(Research) & High D(Development)
Less	Less	Low R(Research) & Low D(Development)

Methodology

Corresponding concepts and equations for computing:

Perplexity: *In natural language processing, it is used to measure the quality of the trained language model and evaluate the generalization ability of the model* (Blei et al., 2003).

$$\text{Perplexity} = \exp\left\{-\left(\sum_{m-1}^{M}\sum_{n-1}^{N_m}\log\left(\sum_{k-1}^{K}p(\omega_n|z_k)p(z_k|d_m)\right)\right)/\left(\sum_{m-1}^{M}N_m\right)\right\}$$

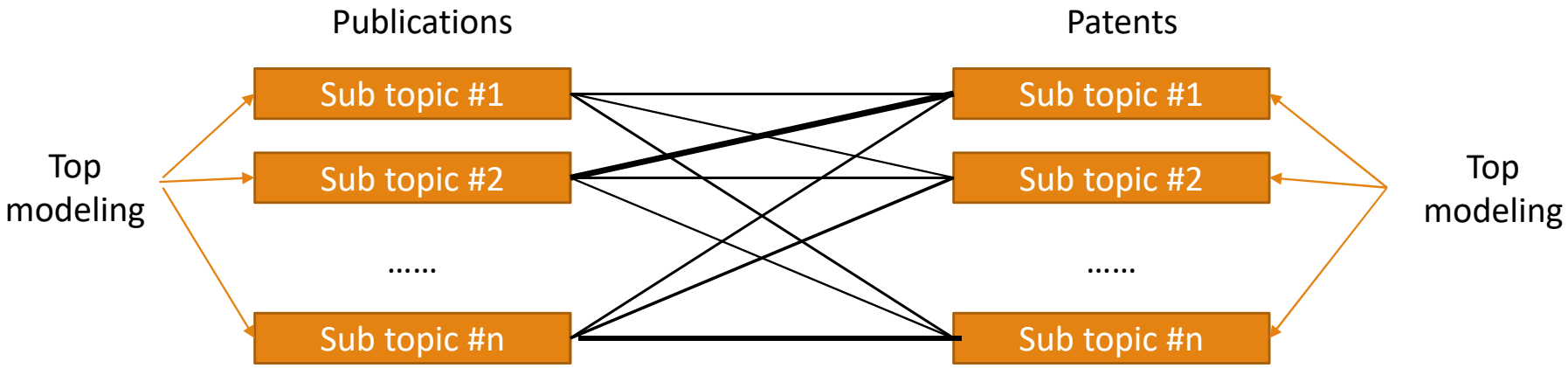
Coherence: *The coherence of topics can be measured by calculating the degree of correlation between the features with higher scores in the topics, which is helpful to classify the topics into understandable categories* (Newman et al., 2010; Stevens, 2012).

$$\text{Coherence} = \sum_{i<j} score(\omega_i, \omega_j, \epsilon)$$

Methodology

For a specific topic, the coupling relationships can be described as such two aspects:

- For a specific term, the amount of publications and the amount of patents can present some useful information.
- Based on the topic modeling, some sub-topics can be extracted from the retrieved publications or patents, and the similarities between these sub-topics can be computed and conducted to evaluate the strength of coupling relations.



Methodology

SDS (*Similarity between Different Sub-topics*): *the Euclidean Distance between two sub-topics, which also needs some techniques for computation.*

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

Two new concepts are proposed:

- ***TI (Transformation Intensity):*** *we just utilized two factors: (1) the probability of a term belonging to a specific sub-topic in publications; (2) this term has been transformed into how many sub-topics of the corresponding patents.*
- ***TE(Transforming Efficiency):*** *here, the concept of HLP(half-life period) is conducted to calculate the TE.*

$$\rho_{s_{z_j}} = \sum_{k=1}^{n} p_{\omega_k} \cdot m_k$$

$$\rho_{e_{z_j}} = \sum_{r=1}^{a} p_{\omega_r} \cdot m_a + \sum_{s=1}^{b} p_{\omega_s} \cdot m_b \cdot \left(\frac{1}{2}\right)^1 + \sum_{t=1}^{c} p_{\omega_t} \cdot m_c \cdot \left(\frac{1}{2}\right)^2$$

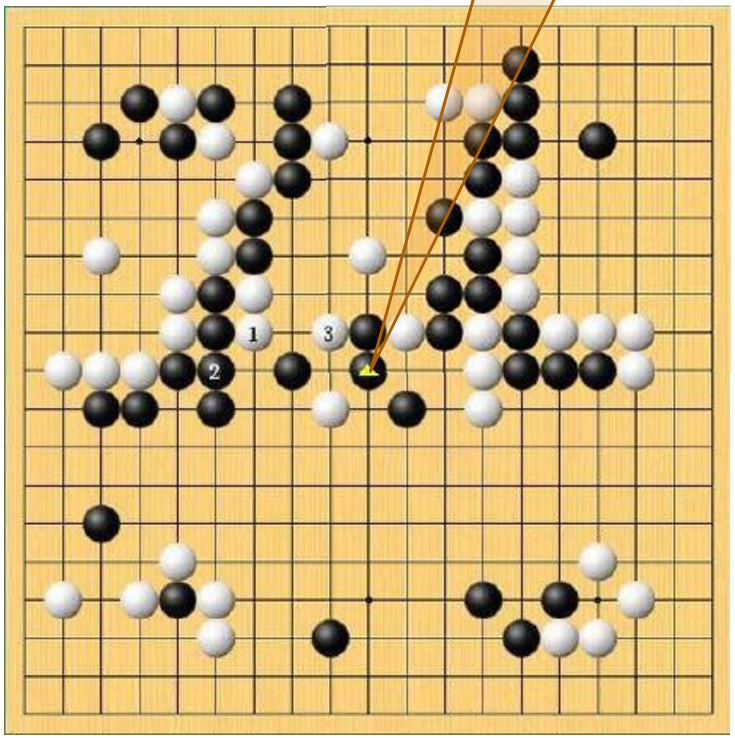
s1

Case study: Artificial intelligence

March, 2016, the milestone event about Go game,
AlphaGo vs Sedol Lee
(4:1)

AlphaGo

A mazing hand in the center from Sedol Lee

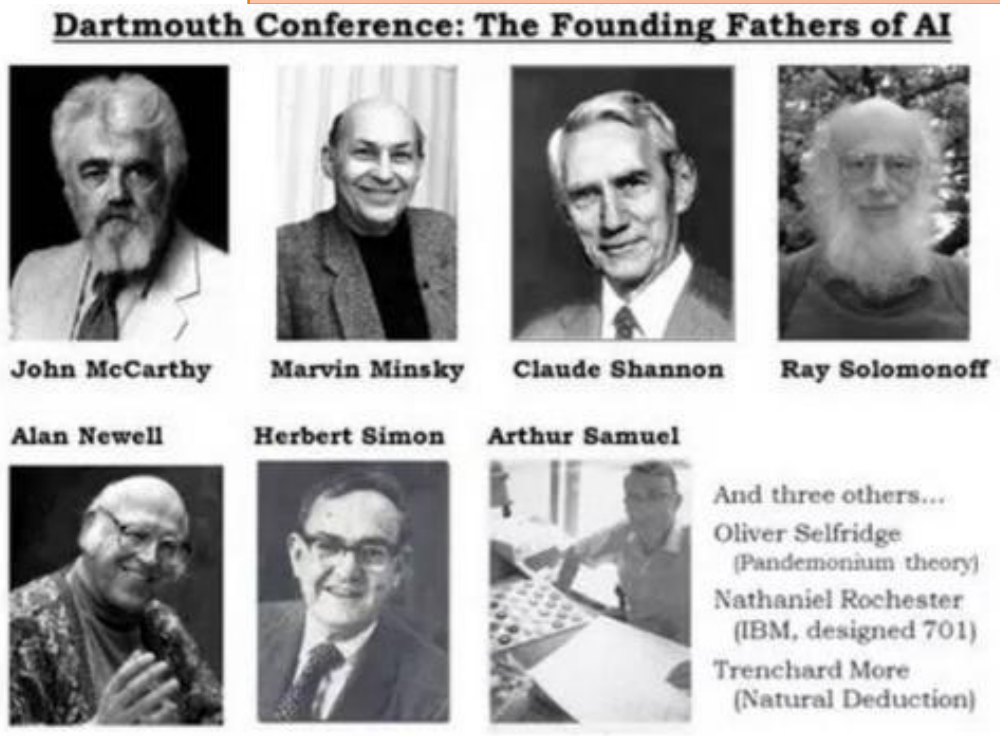


幻灯片 13

s1 scutlimn, 2019/7/23

Case study: Artificial intelligence

In 1956 , the Dartmouth conference is thought as the mark event about the origins of artificial intelligence.

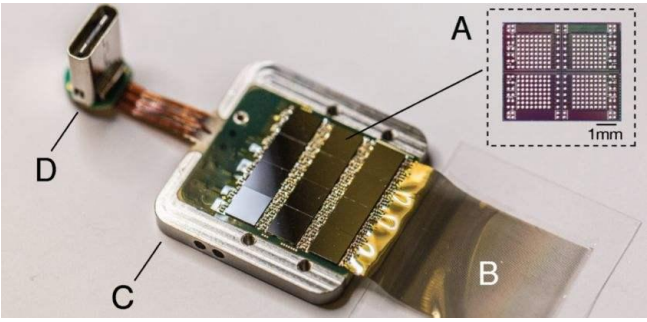


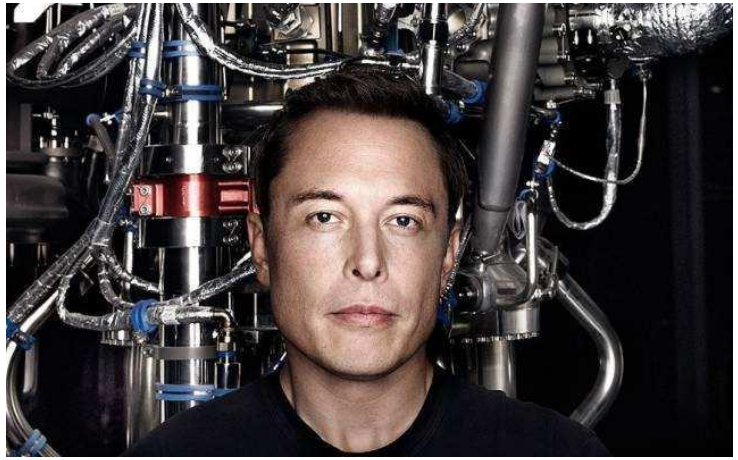
In August 1956, at Dartmouth college, John McCarthy , Marvin Minsky, Claude Shannon, Alan Newell , Herbert Simon and other scientists are together, talking about a complete out-of-touch theme: **Using machines to mimic human learning and other aspects of intelligence**.

The meeting lasted for two months, and although there was no general consensus, a name was given to the discussion: artificial intelligence. Thus, 1956 became the first year of artificial intelligence.

Case study: Artificial intelligence

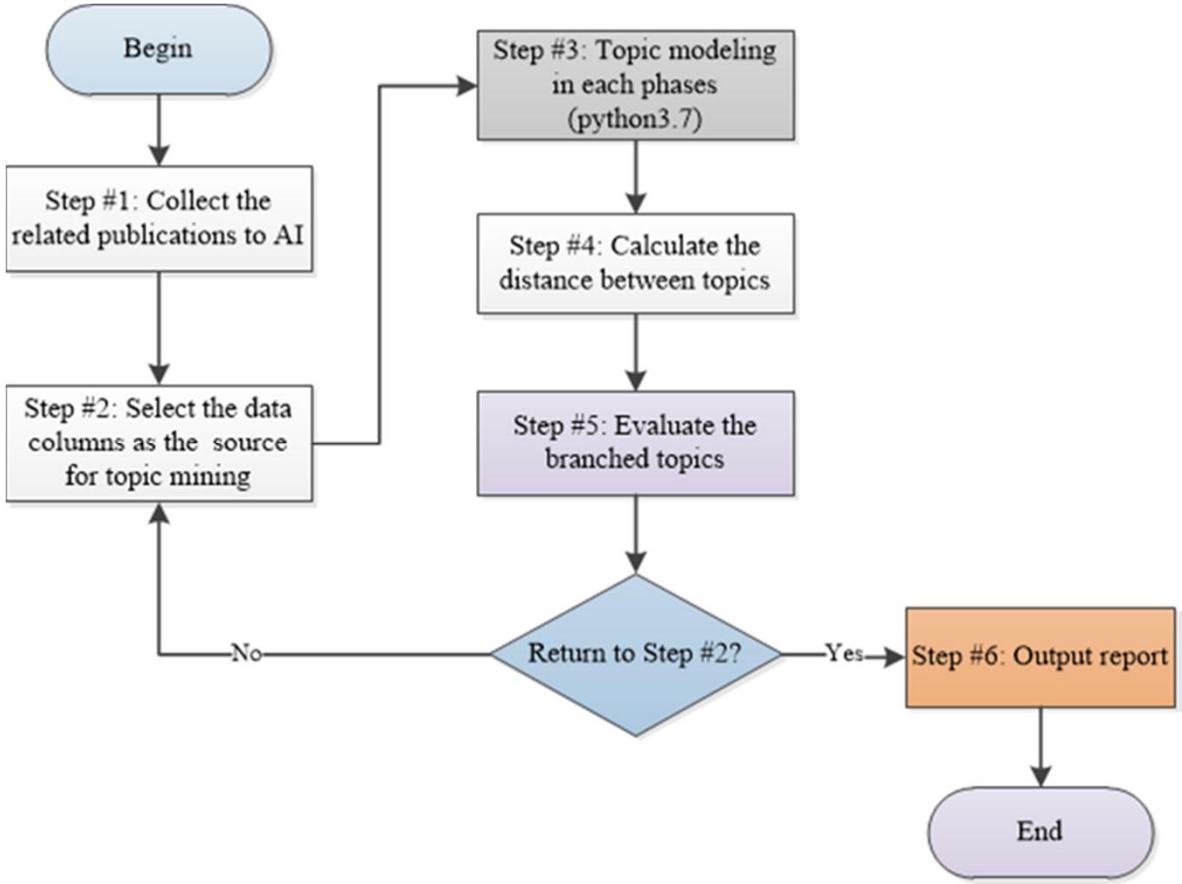
In the 1960s, artificial intelligence evolved different theories and branches such as computational complexity, cognitive computing and computational linguistics. With the promotion and application of relevant models and algorithms of artificial intelligence, research related to human labor union intelligence has spread to different fields and disciplines.



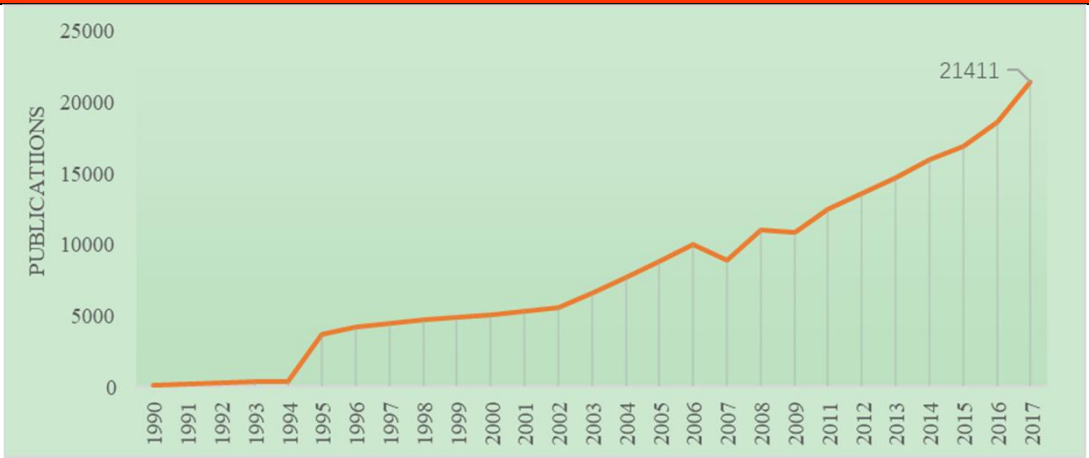


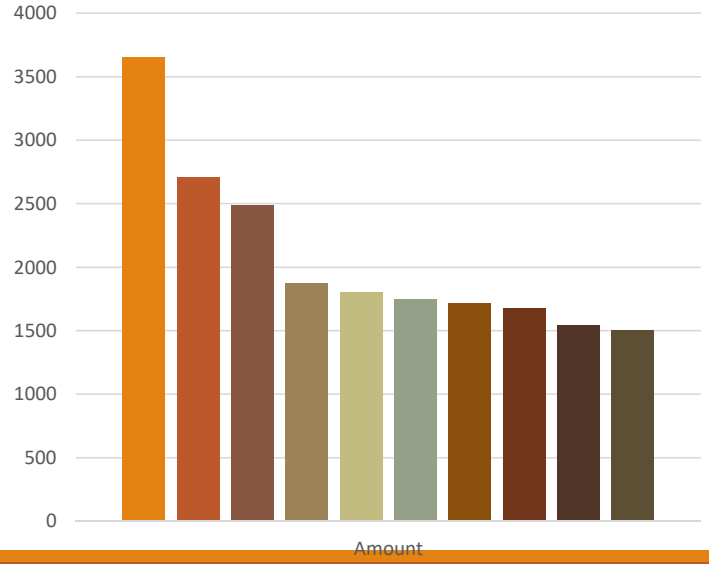
Considering that there are many researches on artificial intelligence and the scope of knowledge diffusion is wide, that is, it is more or less involved in most research fields, the author considers topic modeling to conduct more in-depth content mining and presentation. LDA（Latent Dirichlet Allocation）LDA is a popular and mature topic mining algorithm at present. In essence, LDA is a Bayesian model including subject, document and topic, which is completely based on bayesian reasoning mechanism and has good knowledge interpretation ability (Blei 2003; 2012).

Case study: Artificial intelligence



Case study: Artificial intelligence

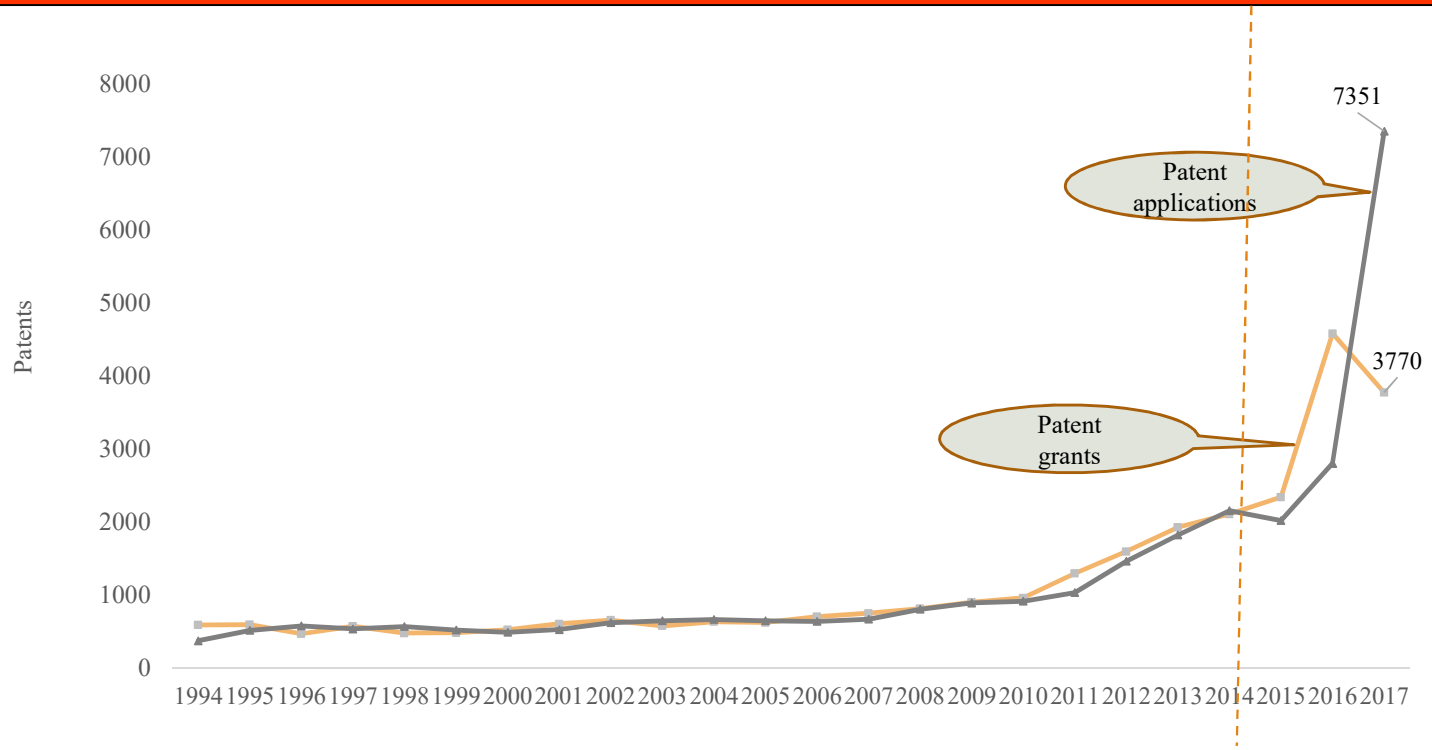




- China
- USA
- UK
- Iran
- India
- Taiwan
- Spain
- Germany
- Japan
- Canada

- CHINESE ACAD SCI
- ISLAMIC AZAD UNIV
- INDIAN INST TECHNOL
- ZHEJIANG UNIV
- UNIV TEHRAN INTELLIGENCE
- SHANGHAI JIAO TONG UNIV
- NANYANG TECHNOL UNIV
- NATL UNIV SINGAPORE

Case study: Artificial intelligence



Rank	Regions	Amout of Patents
1	CHINA	14022
2	USA	6889
3	JAPAN	3085
4	KOREAN	1775
5	GERMANY	724

Patent applications surged 7,351 in 2017, of which 5,655 were from China, accounting for 76.93 percent of the total growth in 2017.

Case study: Artificial intelligence

1990-1999			2000-2009			2010-2017		
Topic code	Publications	Topic tag	Topic code	Publications	Topic tag	Topic code	Publications	Topic tag
L_{1-6}	20957	Classification and classifier	L_{2-13}	57945	Artificial neural network prediction model	L_{3-7}	79750	Fuzzy model and fuzzy system
L_{1-7}	19345	Associative memory and pattern recognition	L_{2-12}	51028	Dynamic neural network stability	L_{3-18}	67900	Neural network model
L_{1-10}	13698	Nonlinear dynamic simulation	L_{2-8}	48661	Optimization design of genetic algorithm	L_{3-5}	64369	Parametric prediction model
L_{1-14}	13200	Clinical medicine	L_{2-2}	48181	Protein structure sequence prediction	L_{3-10}	63070	Support vector machine classification
L_{1-9}	12593	Adaptive fuzzy control	L_{2-3}	45952	Spectral analysis of cancer	L_{3-15}	60883	Optimization design of genetic algorithm

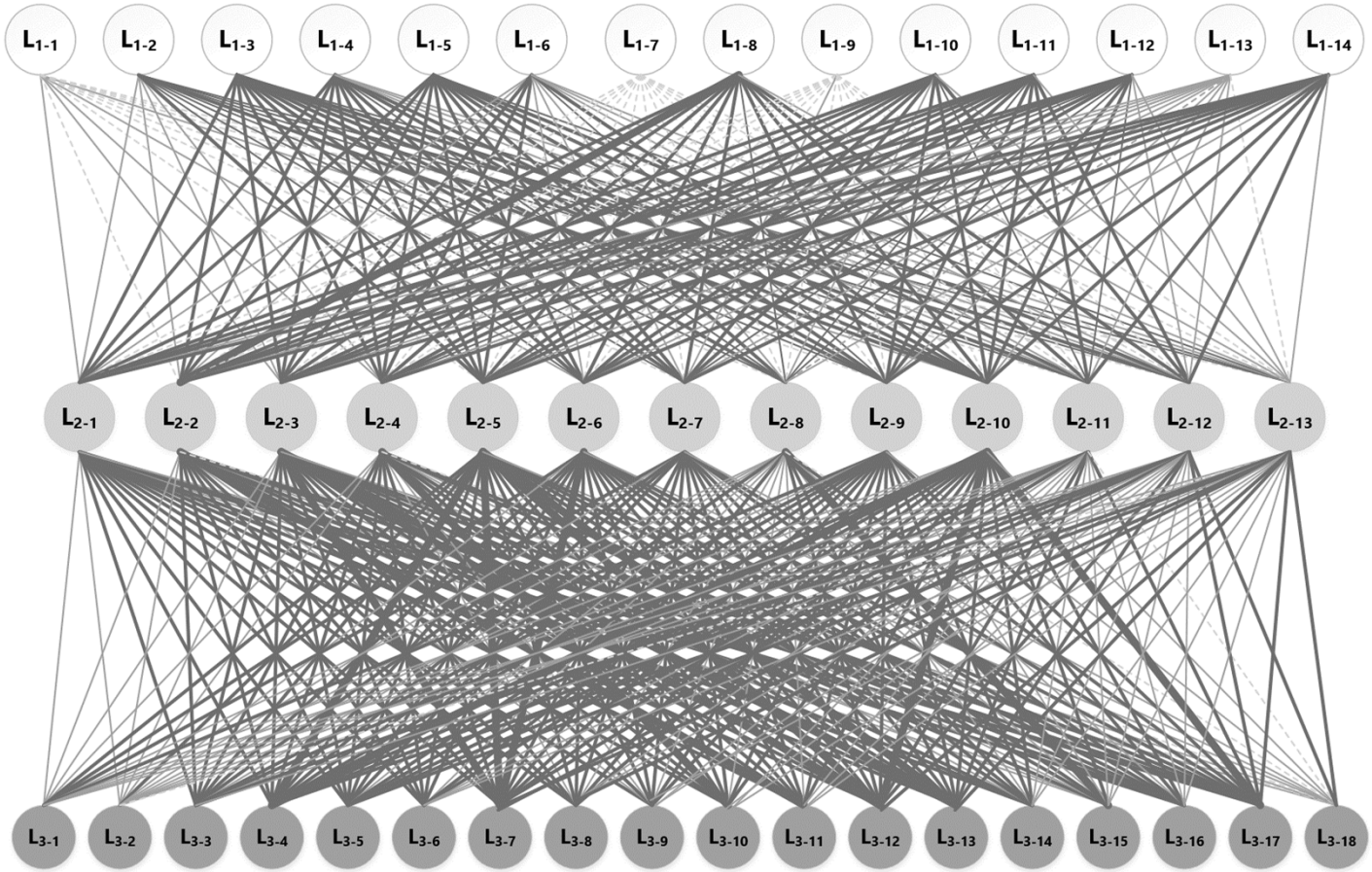
Case study: Artificial intelligence

1990-1999			2000-2009			2010-2017		
Topic code	Patents	Topic tag	Topic code	Patents	Topic tag	Topic code	Patents	Topic tag
P_{1-4}	2001	Neural network	P_{2-5}	5494	Drawing methods based on neural network	P_{3-8}	12276	Data transmission of neural network
P_{1-1}	1731	Image processing	P_{2-3}	4363	Expert system	P_{3-10}	11979	Drawing algorithm
P_{1-6}	1686	Signal sensor	P_{2-6}	3787	Pattern recognition	P_{3-5}	11503	Vector drawing method
P_{1-5}	1369	Computer system	P_{2-4}	3695	Image description	P_{3-7}	10980	Modular control system
P_{1-2}	1318	System and parameters of neural network	P_{2-1}	3589	Traffic sensoring	P_{3-9}	10724	Signaling system

Case study: Artificial intelligence

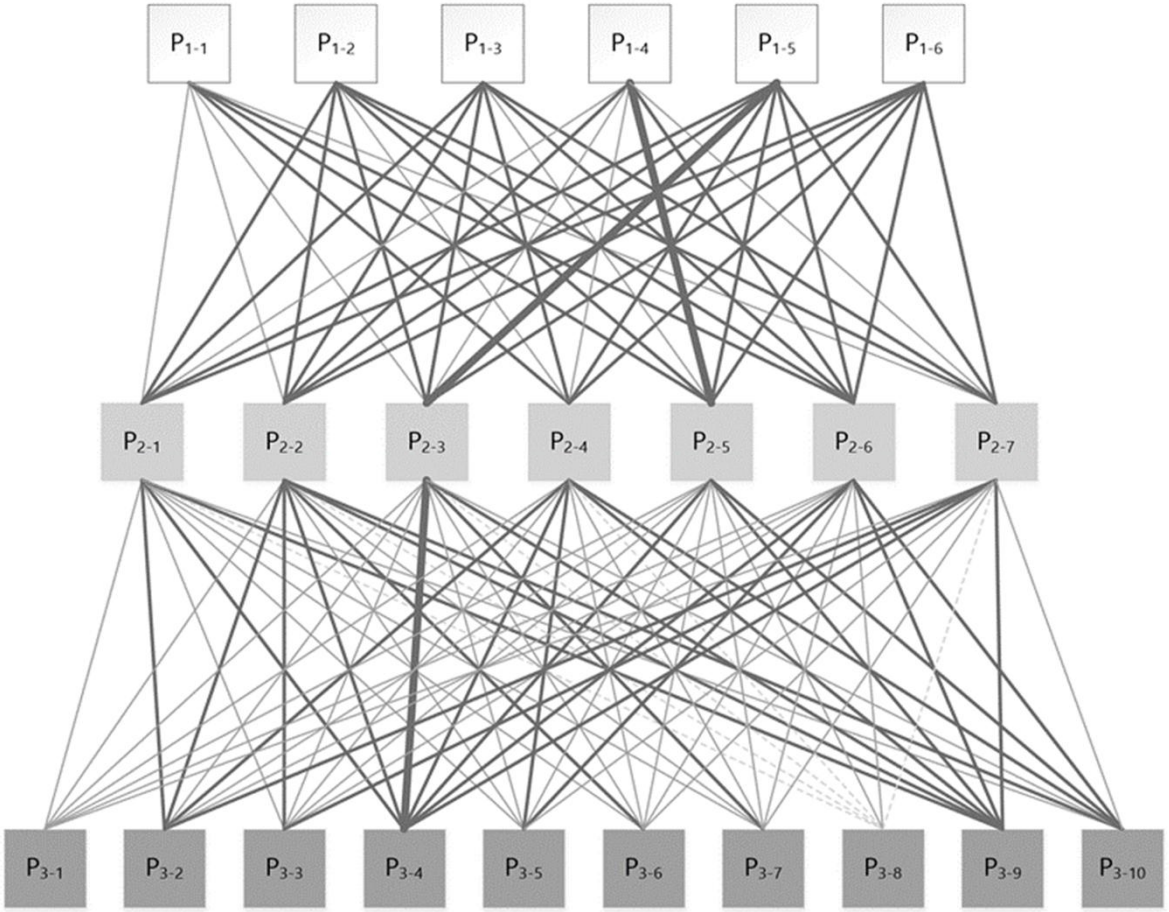
The sub-topics evolution of publications on artificial intelligence

The thickness of the line between two topics is expressed as follows: the smaller the distance (the smaller the difference) is, the thicker the line, indicating the closeness of the relationship and the degree of similarity.

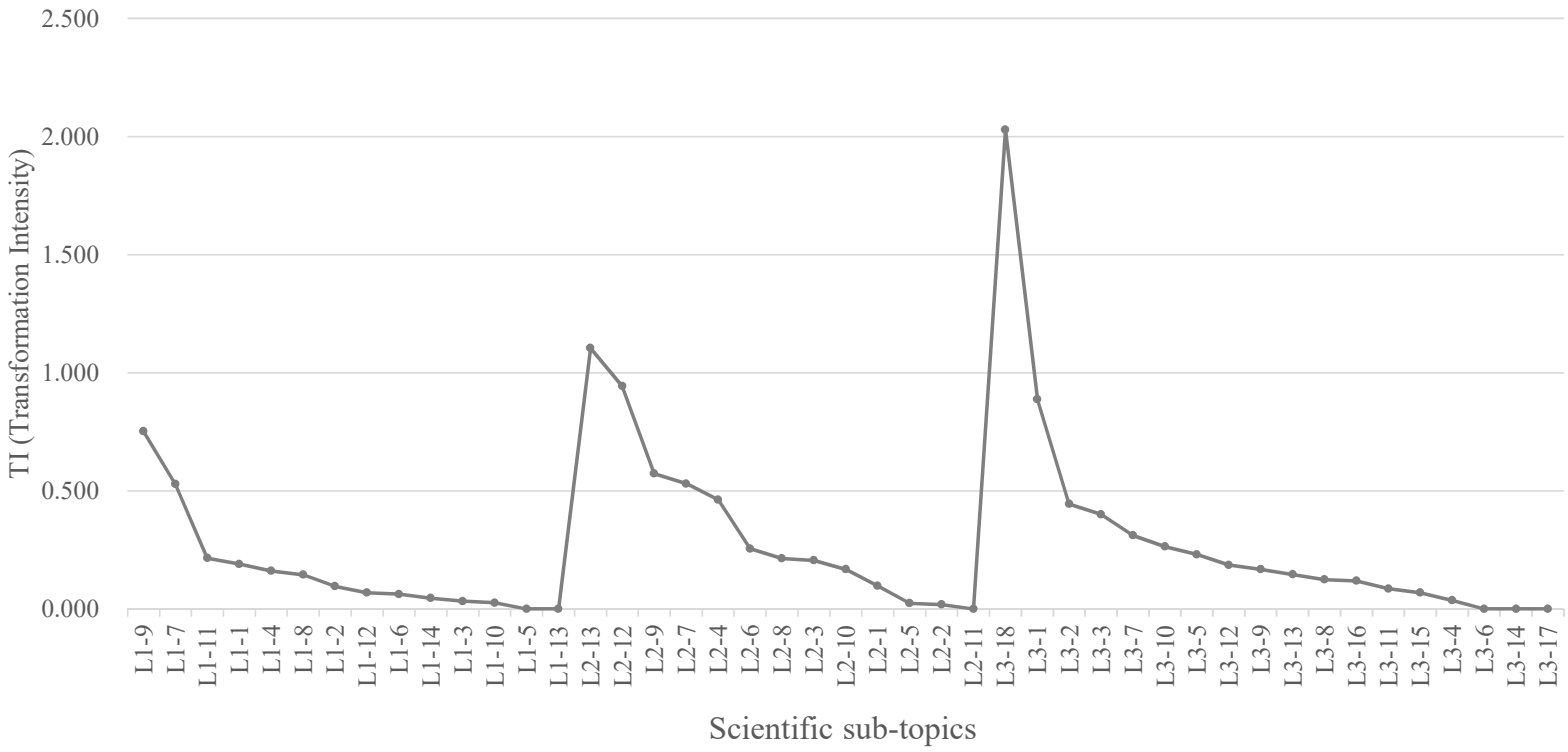


Case study: Artificial intelligence

The sub-topics evolution of patents on artificial intelligence

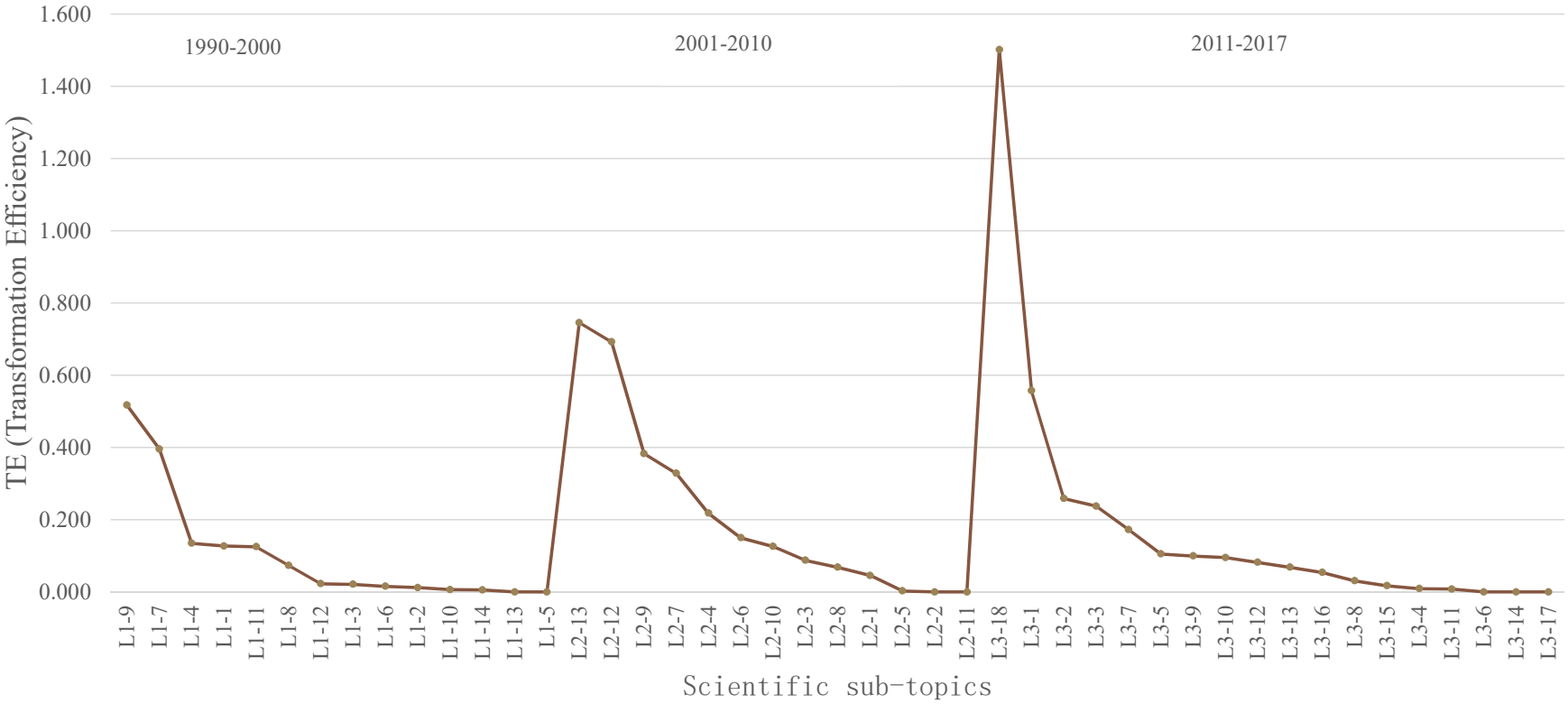


Case study: Artificial intelligence



The TI (Transformation Intensity) on those sub-topics on artificial intelligence

Case study: Artificial intelligence



The TE (Transformation Efficiency) on those sub-topics on artificial intelligence

Conclusion & Discussions

A framework was proposed to forecast the latent TO (technology opportunity) based on the coupling relations between publications and patents, and the case study on artificial intelligence was conducted.

In case study, this paper partially addresses some interesting phenomenon on artificial intelligence.

However, the generalization capability of this proposed framework still needs more cases and verifications in the following studies.

Thank you!