Identification of promising technology in Proptech using topic modeling and technology level map

2022. 09. 02.

Soeun Uhm ,Eunji Jeon, So young Sohn

ISL@YONSEI
Industrial Statistics Lab

Table Of Contents

1.

Introduction

YONSEI UNIVERSITY

1.1 Research Background

✓ **What is Proptech?**
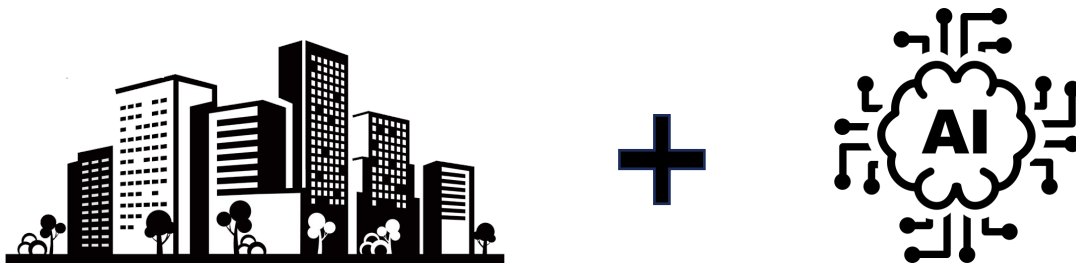
- real estate industry 와 기술의 결합 -> Proptech(Property + Technology)

- Artificial Intelligence, Big Data, Blockchain, Virtual Reality, Drones, Smart Contracts, BIM 와 같은 최신 기술을 real estate brokerage, management, development 에 적용한 것

- Proptech 는 기존 real estate industry 를 혁신하고 있음 (e.g. Zillow in US - 'Zestimate'라는 자체 가치평가 알고리즘을 이용하여 미국 real estate 시장을 선도함)

- Proptech 에 venture funding 은 21년도에 $32 billion 으로, 2020년도에 대비하여 28% 증가함

✓ **Why analyze Proptech from a technological point of view?**

- Real estate 시장을 가격, 거래 내역, 인구, 수입 등과 같이 structured data를 이용해 분석한 연구는 많았음

- Proptech는 기존 Real Estate 분야에 최신 기술을 도입한 새로운 시장이므로, 기술적인 측면에서 분석할 필요가 있음

- Proptech 와 관련된 특허 데이터를 이용하여 분석한다면, Proptech 시장의 기술적인 하위 분야를 정의하고, 미래 Proptech 시장의 promising technology 를 예측할 수 있음

- 기존 real estate 종사자들에게 새로운 proptech 시장에서 경쟁력을 유지할 수 있도록 정보를 제공할 수 있음

✓ **Research Question**

1) **Patent Collection related to Proptech**

2) **Identification of promising technology fields**

4

1.2 Literature Review

✓ Technology Innovations in Real estate Industry

Title	Objective
Siniak, N., Kauko, T., Shavrov, S., & Marina, N. (2020). The impact of proptech on real estate industry growth. *IOP Conference Series: Materials Science and Engineering*, 869(6), 062041.	Describes the **employment changes** resulting from the introduction of technology in the US real estate sector (in real estate brokerage, finance, appraisal, leasing and insurance).
Starr, C. W., Saginor, J., & Worzala, E. (2021). The rise of PropTech: Emerging industrial technologies and their impact on real estate. *Journal of Property Investment & Finance*, 39(2), 157–169.	The use of **real estate online platforms and online transactions** has increased rapidly since COVID-19
Lin, R. F.-Y., Ou, C., Tseng, K.-K., Bowen, D., Yung, K. L., & Ip, W. H. (2021). The Spatial neural network model with disruptive technology for property appraisal in real estate industry. *Technological Forecasting and Social Change*, 173, 121067	Using satellite images to **estimate house price in the area**
Ullah, F., Sepasgozar, S., & Wang, C. (2018). A Systematic Review of Smart Real Estate Technology: Drivers of, and Barriers to, the Use of Digital Disruptive Technologies and Online Platforms. *Sustainability*, 10(9), 3142.	Reviews the **published literature** on real estate and its technologies

✓ Methodology

Title	Objective
Joung, J., & Kim, K. (2017). Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. *Technological Forecasting and Social Change*, 114, 281–292	Framework for **detecting emerging technology** based on patent analysis
Kim, G. J., Park, S. S., & Jang, D. S. (2015). Technology Forecasting using Topic-Based Patent Analysis. *Journal of Scientific & Industrial Research*, 74, 6, 265-270	Suggest Framework for **analyzing vacant technology** and technology forecasting
Moehrle, M. G., & Caferoglu, H. (2019). Technological speciation as a source for emerging technologies. Using semantic patent analysis for the case of camera technology. *Technological Forecasting and Social Change*, 146, 776–784	**Suggest Framework for early and direct detection of emerging technologies** using semant **patent analysis and LDA**
Kwon, K., Jun, S., Lee, Y.-J., Choi, S., & Lee, C. (2022). Logistics Technology Forecasting Framework Using Patent Analysis for Technology Roadmap, *Sustainability,* 30.	Use Technology Level map to assess technology level in logistics

5

1.3 Research Gap

✓ **Technology Innovations in Real estate Industry**

1. Starr, C. W., Saginor, J., & Worzala, E. (2021). The rise of PropTech: Emerging industrial technologies and their impact on real estate. *Journal of Property Investment & Finance*, 39(2), 157–169.

-> Real Estate 분야에서 기술이 적용될 시 나타날 **employment 의 변화**를 중심적으로 서술

2. Ullah, F., Sepasgozar, S., & Wang, C. (2018). A Systematic Review of Smart Real Estate Technology: Drivers of, and Barriers to, the Use of Digital Disruptive Technologies and Online Platforms. *Sustainability*, 10(9), 3142.

-> **Published literature** 를 사용하여 Proptech 내에 어떤 분야가 있고, 어떠한 기술이 사용되었는지 리뷰

-> 기존 연구에서는 proptech의 기술적인 측면에서 접근한 연구보다 논문들에 대한 review 와 사회에 미친 영향을 중점적으로 서술함

-> 따라서 본 연구에서는 **Patent Data** 를 사용하여 Proptech 분야를 분석하는 방법을 제안

✓ **Methodology**

- Topic Modeling
  1. Kim, G. J., Park, S. S., & Jang, D. S. (2015). Technology Forecasting using Topic-Based Patent Analysis. *Journal of Scientific & Industrial Research*, 74, 6, 265-270
  2. Moehrle, M. G., & Caferoglu, H. (2019). Technological speciation as a source for emerging technologies. Using semantic patent analysis for the case of camera technology. *Technological Forecasting and Social Change*, 146, 776–784

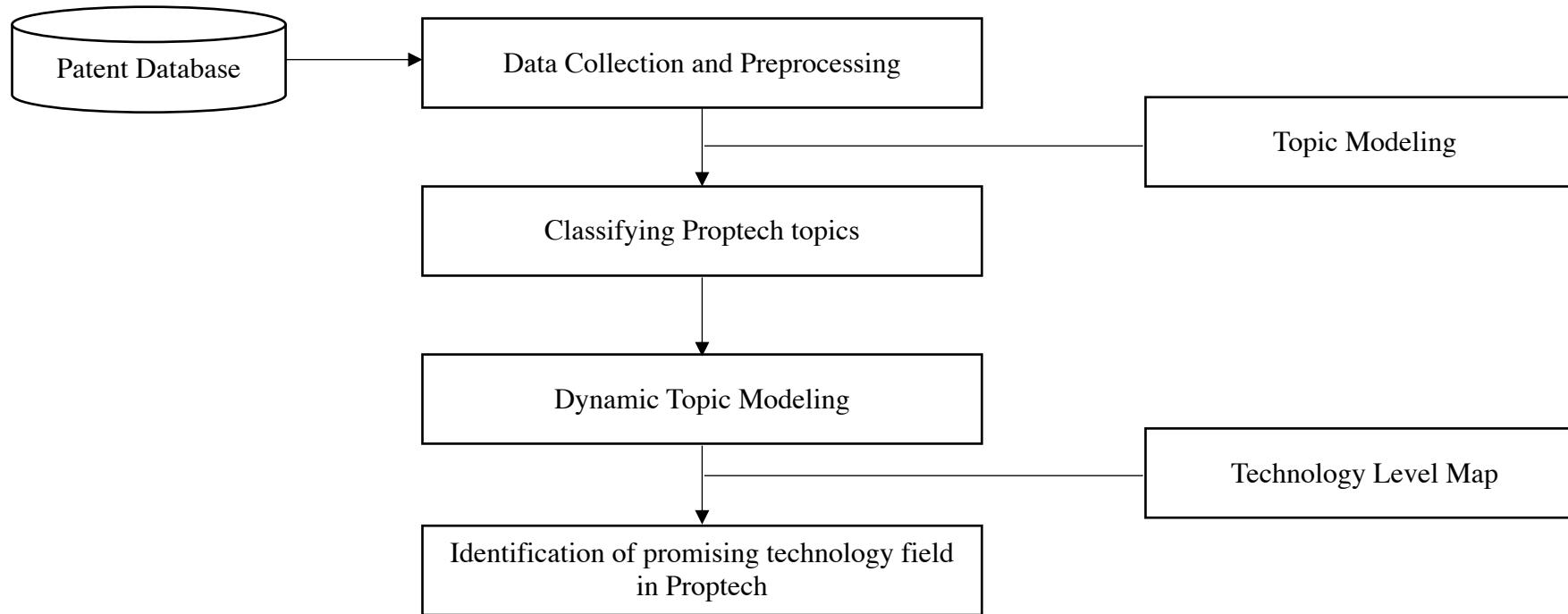   -> 기존 Technology cluster 는 **Delphi method**, 전문가의 의견을 토대로 진행되는 연구가 많았음

     -> Text-mining 분야가 발달하며 Topic Model 기법으로 **LDA** 가 등장하자 Technology 내 cluster를 정의하는 기법으로 쓰임

     -> 본 연구에서는 Topic Model 기법으로 TF-IDF기법으로 중요한 단어의 의미를 유지하면서 자동적으로 토픽의 수를 생성해주는 BERTopic 을 사용하고자 함

2.

Data & Methodology

2.1 Research Framework

Suggest patent collection keywords and identify future promising technology fields in proptech by Technology Level Map

YONSEI UNIVERSITY

2.2 Data Collection and Preprocessing

1. **Patent data collected by WIPSON**

*   Patents enrolled in all US, EP, Japan Patent Office
*   Keyword '**real estate**' -> 1676 patent data (a)

2. Implementing Keyword list

✓ **Co-word Analysis**
*   To implement keywords based on co-occurring frequency
*   Top Co-occurrence List on data(a)

-> property, display, location, network, image, user, application, communication, database, management, seller, sale, service, buyer, client, unit, time, agent, device, asset, loan, purchase, access, interface, transaction, price

✓Word2Vec
*   To implement keywords based on word similarity
*   e.g. Most similar word of 'transaction'

-> sellers, bid, list, sales, fee, agents, profile, listing, lender

```
model2.wv.most_similar("transaction")

[('sellers', 0.9931749701499939),
 ('potential', 0.9923407435417175),
 ('bid', 0.9907337427139282),
 ('list', 0.9900424480438232),
 ('sales', 0.9897407293319702),
 ('fee', 0.9897056818008423),
 ('agents', 0.9894698858261108),
 ('profile', 0.9889885187149048),
 ('listing', 0.9878023862838745),
 ('lender', 0.9877550601959229)]
```

```
model2.wv.most_similar("mortgage")

[('financial', 0.9979677200317383),
 ('loan', 0.9969314932823181),
 ('equity', 0.9953702688217163),
 ('investor', 0.9950581192970276),
 ('market', 0.9934873580932617),
 ('contract', 0.9931070804595947),
 ('purchase', 0.9930815696716309),
 ('owner', 0.9925937652587891),
 ('lender', 0.9908874034881592),
 ('sale', 0.9906339645385742)]
```

```
model2.wv.most_similar("database")

[('request', 0.9962267875671387),
 ('criteria', 0.994577944278717),
 ('service', 0.9935052394866943),
 ('listings', 0.9930511713027954),
 ('records', 0.9925845861434937),
 ('identifying', 0.9911995530128479),
 ('received', 0.9909391403198242),
 ('selection', 0.9907487630844116),
 ('attributes', 0.9907217621803284),
 ('selected', 0.9901400208473206)]
```

Word2vec

Co-word analysis

YONSEI UNIVERSITY
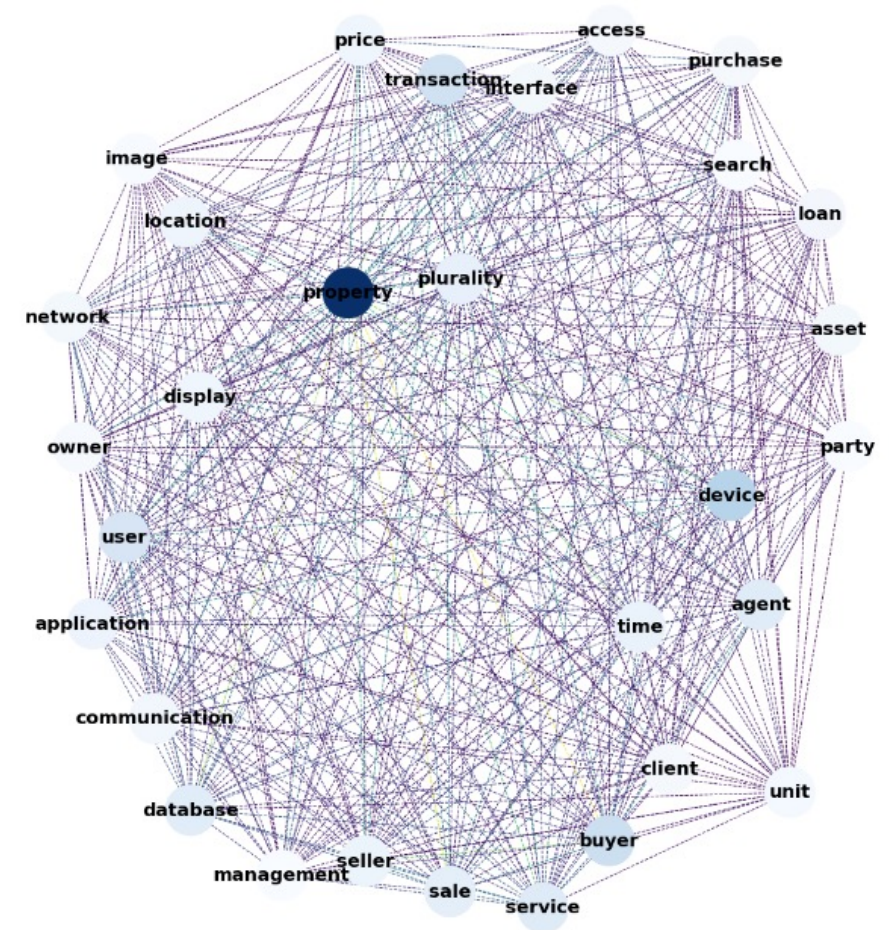
2.2 Data Collection and Preprocessing

3. Final Keyword List

Keyword
Property, Transaction , Virtual Reality, Display, Listing, Data base, Sale, Finance, Mortgage, Location, Network, Interface, Device, Server, Loan, Listing, Client, Time, Agent, Asset, Purchase, Price, Equity, Market, Contract, Fund, Lender, Bid

5. Data Preprocessing
- Summary of Patent Data is used for further analysis
- Text Preprocessing Steps
  - Remove stopwords
  - Lemmatization
  - Get Nouns list

4. Patent Data
- 위 keyword 를 기반으로, Wipson 에서 patent data 보충 (b)
- Remove Duplicate data
- Remove patent applied before than 2000
- Remove unrelated patents

-> (a) + (b) :Total 2668 patent data

6. Final Data

```
['system method image insurance quote device capture image video stream house feature determine insurability replacement cost device capture location
assessment characterization feature homeowner analyze generate insurance quote insurance application transmit insurance quote application device
presentation user device user term limit coverage homeowner insurance policy',
 'semi determination valuation patent application entity machine representation patent application dimension database ontology code device figure
market finance company time communication protocol exchange network societal disparity transformation transformation economy business ecosystem
damage environment power shift urbanisation representation patent application dimension layer training machine model patent application layer patent
input interaction representation patent application user input refine machine model valuation patent application machine machine model',
 'reality view edit store display feature visualization display location person house vacation stay editor user edit input element database user
upload assign element database viewer database location view direction viewer display image time field view availability feature icon shape avatar
display picture video avatar video audio file database',
```

2.3 Methodology – Topic Modeling

✓ **Topic Modeling**

<u>**BERTopic**</u>

- SBERT를 이용한 Topic model

- BERT embeddings과 클래스 기반(class-based) TF-IDF를 활용하여 주제 설명에서 중요한 단어를 유지하면서도 쉽게 해석할 수 있는 조밀한 클러스터를 만드는 토픽 모델링 기술

- LDA(Latent Dirichlet Allocation)과 달리 사전에 topic 수를 정해주지 않아도 된다는 장점

- Noise Topic 을 허용하여 기존의 토픽에 속하지 않는 데이터를 따로 분류할 수 있음

Dynamic Topic Modeling

- 시간에 따른 Topic의 등장 빈도수를 분석할 수 있음

- 해당 토픽에 포함되는 단어들이 시간에 따라 등장하는 빈도수를 기반으로 분석

2.3 Methodology – Technology Level Map

✓ **Technology Level Map**

- 기술의 상대적인 발전 정도를 평가하기 위한 척도
- 기술의 정량적 요소와 정성적 요소를 모두 고려 -> Identify technologies of high growth potentials and effectiveness

- X-axis: **Technology Activity Index(TAI)**
  Patent's Quantitative level

$$Technology\ Growth = \frac{N_i}{N}$$

$$N = Number\ of\ total\ patents\ in\ proptech\ patents$$

$$N_i = Number\ of\ total\ patents\ in\ each\ topic$$

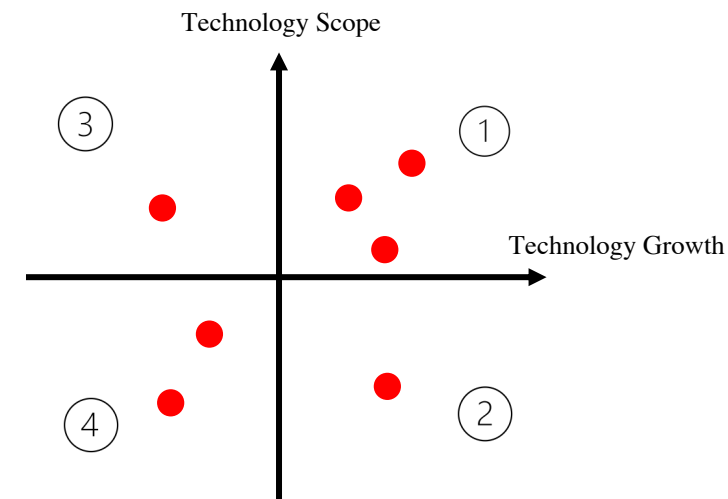-> 토픽 내에 속하는 총 특허 개수가 많을 수록 연구가 많이 된 분야로 해석 가능

- Y-axis: **Technology Impact index(TII)**
  Patent's Qualitative Level

$$Technology\ Scope = \frac{C_i}{C}$$

$$C = Number\ of\ total\ claims\ in\ proptech\ patents$$

$$C_i = Number\ of\ total\ claims\ in\ topic$$

-> 청구항 수는 특허의 법적 권한을 말하며, 청구항 수가 많을 수록 기술경쟁력과 독점권이 높아지므로 품질이 우수한 특허로 해석가능
-> (추가적인 지표 고려 필요)

①: quan/qual superiority technology - Analysis Target
②: quan superiority/ qual inferiority technology
③: quan inferiority/ qual superiority technology
④: quan/qual inferiority technology

2.3 Methodology – Technology Level Map

✓ **Technology Level Map**

- **특허의 수준을 평가하기 위한 추가적인 요소**

-> **Patent Quality Index**

Patent Quality Index (PQI), as proposed by OECD, measures the technological and commercial values of technological innovation using patent information, with six sub-indexes including forward citation, family size, the number of claims, generality index, backward citation and grant lag.

세부지표명(영문)	세부지표명(국문)	의미
Forward Citations	피인용지수	The number of references used by a patent applied later than a target patent and generally indicates technology's influence. The better the quality of the patent, the higher the number of references.
Family Size	패밀리 특허 지수	Number of countries in which patent applications of the same invention have been applied due to a target patent and a priority claim. The better the quality of patent, the more family size it has.
Number of Claims	청구항 지수	Number of claims and the patent protection coverage for technology. The better the quality of patent, the higher the number of claims it has.
Generality Index	범용성 지수	Number of technical fields which have an influence on technology and technology diffusion in other fields
Backward Citation	인용지수	It shows a patent reference used before a target patent is applied. The better the quality of patent, the higher the number of preceding patent references.
Grant Lag	특허등록 소요 기간	It means the period of time after which a patent has been applied for and registered. The better the quality of patent, the earlier the patent is registered.
PQI Index	특허품질지수	Each index is fully qualified, having a weight value of 1 Korea/Japan : (Family size+ Grant lag+ Number of claims)/3 The United States: (Family size + Grant lag + Number of claims + Backward citation +Grant lag)/6

3.

Empirical Analysis
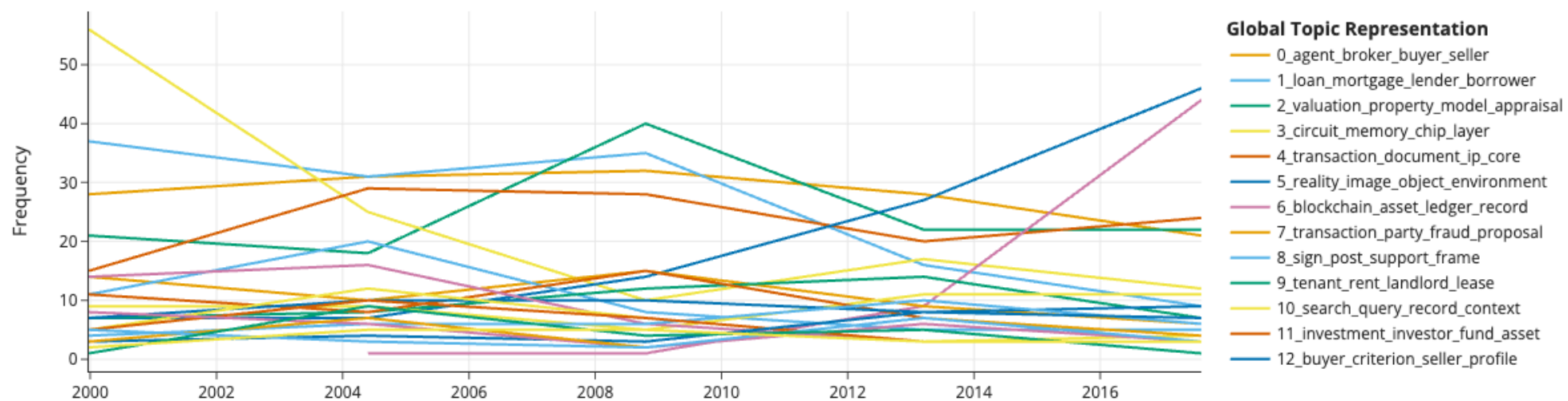
3.1 Topic Modeling

✓ **BERTopic**

- 총 25개의 topic 으로 분류됨

Topic	Name	Topic	Name
0	Agent, broker, buyer, seller	13	Equity, ownership, investor
1	Loan, mortgage, lender	14	Display, screen, menu
2	Valuation, property, model	15	Location, celebrity, device
3	Circuit, memory, chip	16	Insurance, polity, title
4	Transaction, document, IP	17	Patent, business, IP
5	Reality, Image, object	18	Content, product, consumer
6	Blockchain, asset, ledger	19	Game, player, play
7	Transaction, party, fraud	20	Device, wireless, transmit
8	Sign, post, support	21	Lock, access, lockbox
9	Tenant, rend, landlord	22	Project, construction, development
10	Search, query, record	23	Auction, bid, bidder
11	Investment, investor, fund	24	Asset, population, analysis
12	Buyer, criterion, seller		

3.2 Time series Analysis

✓ **Dynamic Topic Modeling**

Topics over Time

Global Topic Representation
- 0_agent_broker_buyer_seller
- 1_loan_mortgage_lender_borrower
- 2_valuation_property_model_appraisal
- 3_circuit_memory_chip_layer
- 4_transaction_document_ip_core
- 5_reality_image_object_environment
- 6_blockchain_asset_ledger_record
- 7_transaction_party_fraud_proposal
- 8_sign_post_support_frame
- 9_tenant_rent_landlord_lease
- 10_search_query_record_context
- 11_investment_investor_fund_asset
- 12_buyer_criterion_seller_profile

Frequency 기반 최근 5년간 상승세 Topic

Topic6: Blockchain, asset, ledger (블록체인 기술)
Topic5: Reality, image, object, environment (VR 기술)
Topic4: Transaction, document, IP (거래 관련 기술)
Topic2: Valuation, property, model, appraisal (부동산 가치평가 관련 기술)
Topic10: Search, query, record, context (기록 관련 기술)
Topic 12: Buyer, seller, criterion, profile (매매자-매도자 간 평가 관련 기술)

3.3 Technology Level Evaluation

✓ **Technology Level Map**
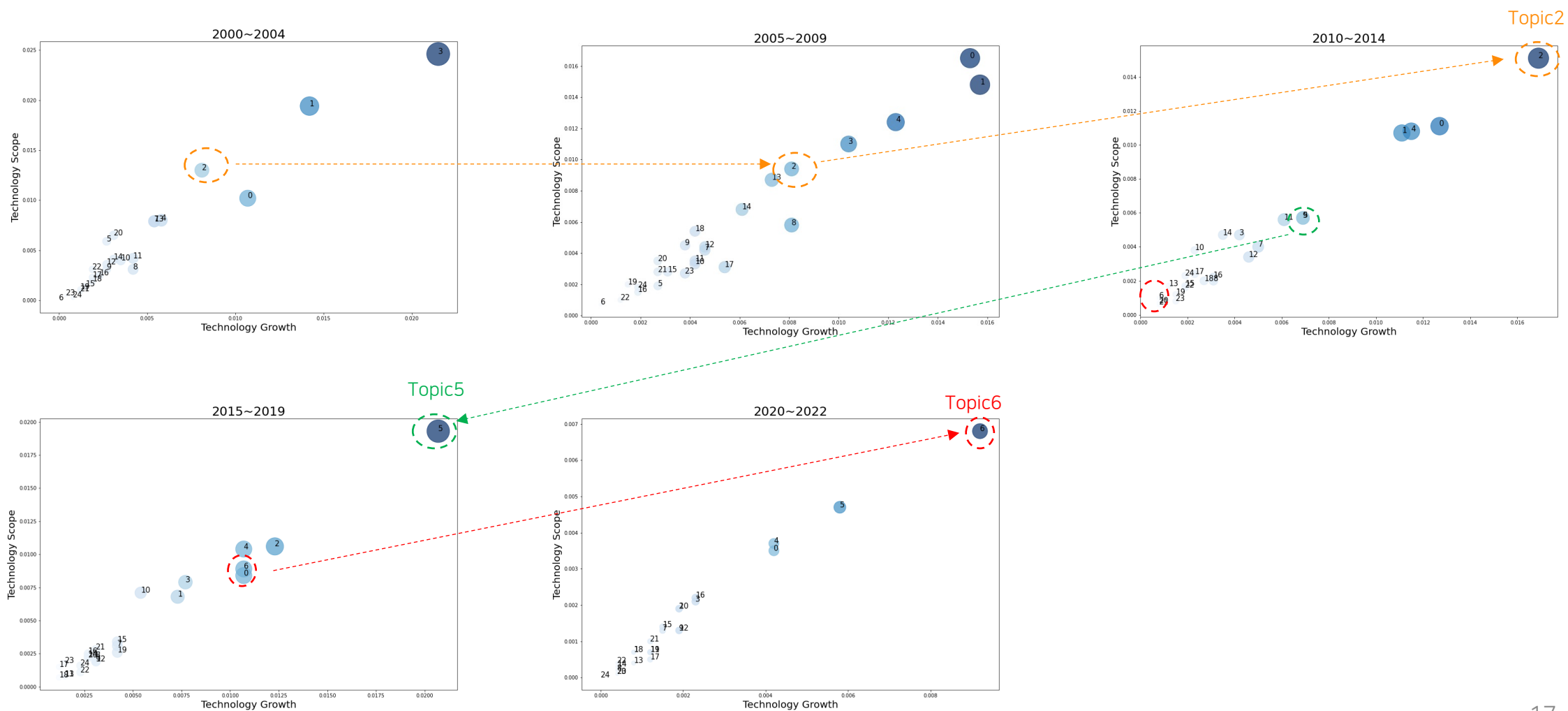
3.5. Conclusion & Limitation

✓ **Conclusion**
- Dynamic Topic Modeling 과 Technology Level map 을 모두 고려하였을 때, Proptech 분야의 promising technology는 다음과 같다.
    - Topic6:   Blockchain, asset, ledger (블록체인 기술)
    - Topic5: Reality, image, object, environment (VR 기술)
    - Topic2: Valuation, property, model, appraisal (부동산 가치평가 관련 기술)

✓ **Limitation**
- Topic 들의 명칭을 정하는 한계
- Technology Level Map 관련 추가 지표 고려 필요
- Technology Level Map 의 분면을 나누는 명확한 기준 부재

3.4. Reference

- Luque.(2021), Proptech Global Trends 2021 Barometer
- Moehrle, M. G., & Caferoglu, H. (2019). Technological speciation as a source for emerging technologies. Using semantic patent analysis for the case of camera technology. *Technological Forecasting and Social Change*, 146, 776–784.
- Kwon, K., Jun, S., Lee, Y.-J., Choi, S., & Lee, C. (2022). Logistics Technology Forecasting Framework Using Patent Analysis for Technology Roadmap, *Sustainability,* 30.
- Xu, L., Lu, X., Yang, G., & Shi, B. (2020). Identifying FinTech Innovations with Patent Data: A Combination of Textual Analysis and Machine-Learning Techniques. *Sustainable Digital Communities* (Vol. 12051, pp. 835–843).
- Chun, E., Jun, S., & Lee, C. (2021). Identification of Promising Smart Farm Technologies and Development of Technology Roadmap Using Patent Map Analysis. *Sustainability*, 13(19), 10709.
- Chun, E., Jun, S., & Lee, C. (2021). Identification of Promising Smart Farm Technologies and Development of Technology Roadmap Using Patent Map Analysis. *Sustai nability*, *13*(19), 10709.

감사합니다.