



기업의 잠재력 및 미래의 융합 트랜드를 고려한 빅데이터 기반의 고급 기술인력 탐색 프레임워크 제안

연세대학교
산업공학과
ISL

이종욱

Nov. 01, 2019



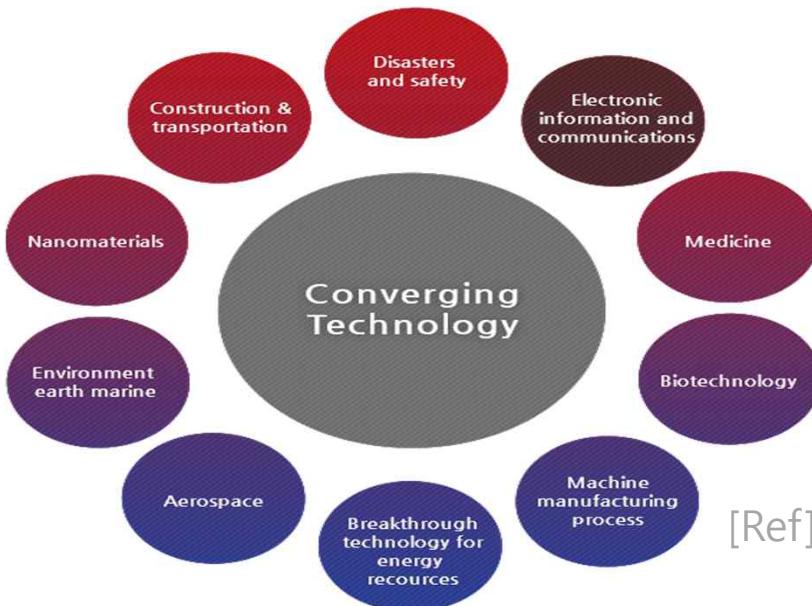
Contents

- 1. Introduction**
- 2. Literature Review**
- 3. Framework**
- 4. Case Study**
- 5. Conclusion**

Introduction

☒ 기술 융합

- ✓ 첨단 기술 발전에 따른 **서로 다른 기술간의 거리 감소**로 인해 기술 융합이 큰 관심을 받고 있음.
(Caviggioli, 2016)
- ✓ 따라서 기업은 자신의 주력 기술 뿐만 아니라, **다른 분야와의 융합**을 고려한 미래 전략 수립이 필수!
- ✓ 전략 수립에 앞서, 융합 대상 분야의 **고급 인재를 먼저 선점**하는 것이 바탕이 되어야 함.
- ✓ 본 연구에서는 융합 대상 분야를 **ICT 분야**로 Case Study.

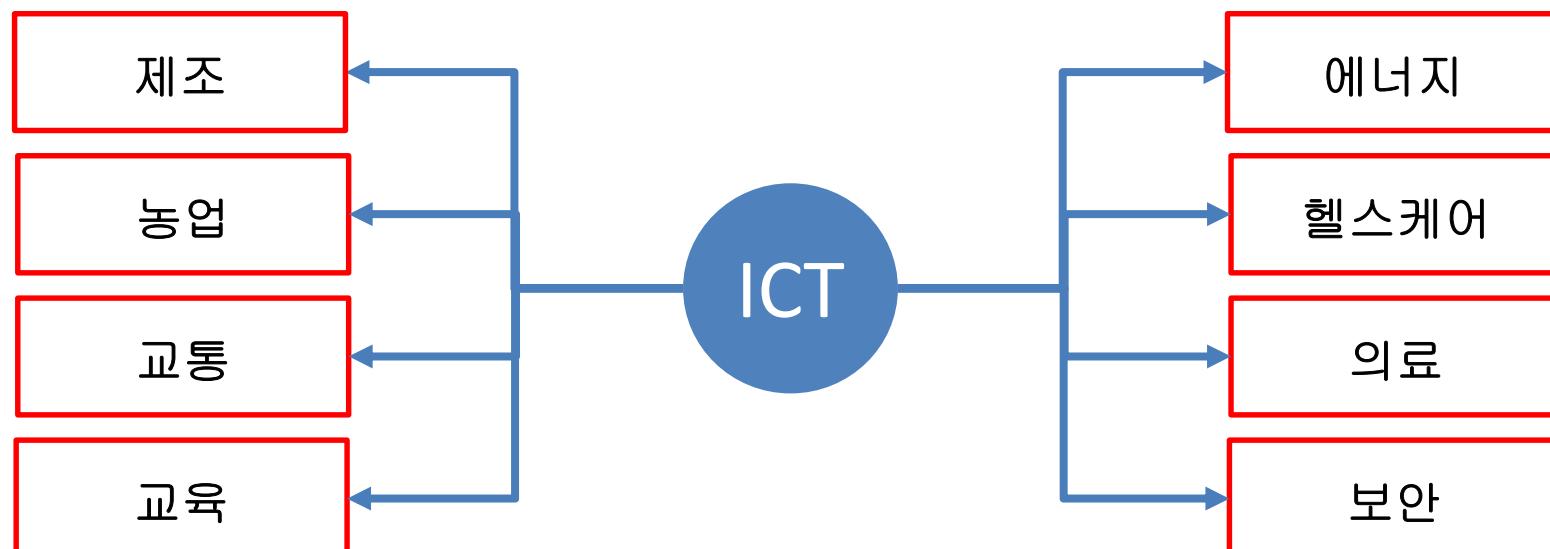


[Ref] <https://crpc.kist.re.kr/>

Introduction

☒ ICT 란?

- ✓ OECD : “전자적으로 데이터 및 정보를 수집하거나 전송하는 제조 산업과 서비스 산업의 융합”
- ✓ 현재 우리는 ICT 산업 속에서 살아가고 있다는 것은 과언이 아님.
(Buyya et al., 2009)
- ✓ ICT 산업은 제조업을 넘어서, 농업, 교통, 교육, 에너지, 헬스 케어, 보안 등 대부분의 산업에 융합되고 있음.
(Ma et al., 2003; Gollakota, 2008; Zhang and Li, 2017; Ali and Kumar, 2011)



Introduction

☒ ICT 고급 인력 모집의 어려움

- ✓ 많은 기업들이 시대적 흐름에 따라 **ICT 산업과의 융합**을 고려하며, 고급 인력을 탐색하고자 함.
- ✓ 하지만, 대부분의 나라에서 ICT 고급 인력의 수요에 비해 **공급이 매우 부족**한 상태.
- ✓ 한국에서는 **정부 차원**에서 ICT 인재를 창출하기 위해 다양한 정책을 벌이고 있음.
- ✓ 또한, **기업 차원**에서도 고급 인력을 찾기 위해 많은 노력을 하고 있음.
- ✓ 정부 및 기업의 많은 노력에도 불구하고, ICT 산업 내의 고급 인력을 찾아내는 것은 **큰 어려움**이 있음.



Introduction

☑ 연구의 목적 및 의의

- ✓ 본 연구는 기업의 보유 기술력 뿐만 아니라, 기업의 주력 기술 분야와 ICT 기술 분야와의 미래 융합 트렌드를 고려한 특허 연구자를 추천하는 프레임 워크를 제시하고자 함.
- ✓ 본 연구를 통해, 기업이 융합 대상 분야의 R&D 인력을 탐색 및 모집하는데 투자 되는 시간적 비용을 크게 줄여줄 수 있을 것이라 기대됨.
- ✓ 기업의 보유 기술 중심으로 미래의 융합 트렌드가 함께 고려되기 때문에, 기업의 미래 발전 방향에 부합하는 고급 인력이 탐색 될 것으로 기대됨.



Literature Review

인력 모집 전략

- ✓ **인적 자원**은 기업의 발전에 있어서 가장 중요한 요소 (Rao et al., 2010)
- ✓ Stochastic Programming model for **optimal recruitment allocation** (Jeeva et al., 2004)
- ✓ LP model to determine the **optimal workforce size** (Akinyele et al., 2007)
- ✓ **Workforce planning strategy** based on LP (Kareem and Aderoba, 2008)
- ✓ **Risk-based** workforce planning model (Jaillet et al., 2018)

본 연구의 차별점

- ✓ 기존의 연구는 모두 기업의 현 상황만 고려할 뿐, **기술 융합 관점**에서의 인력 모집 전략은 없었음.
- ✓ 기존의 연구는 Macro한 전략만 제시할 뿐, **Data 기반의 전략**은 제시되지 않았음.



따라서 본 연구는, 기술 융합 관점에서 빅데이터 기반의 인력 탐색 전략을 제시.

Literature Review

기술 융합

- ✓ 기술 융합은 산업 및 **기술 혁신의 가장 중요한 요소**를 여겨짐. (Curran and Leker, 2011)
- ✓ 기술 융합은 기업으로 하여금, 경로 의존성을 탈피시키고, **새로운 시장을 선점하게 함.** (Lei, 2000; Yoffie, 1996)
- ✓ 기존의 기술 융합 연구는 기술 혁신의 근본있는 지표라 할 수 있는 **특허 데이터**를 활용 (Ernst, 2003)
- ✓ 기술 융합의 측정하기 위한 지표로, **공동 발현 IPC** (국제 특허 분류) 혹은 **공동 발현 Citation**이 활용됨.
- ✓ 공동 발현 Citation의 경우, 특허의 기술적 성질을 직접적으로 나타내지 않는 단점이 있어, 공동 발현 IPC에 비해 큰 **한계점**을 가지고 있음. (Choi et al., 2015)

공동 발현 IPC

- ✓ IPC는 특허가 포함하는 **기술의 영역**에 대한 정보를 제공.
- ✓ IPC는 넓은 범위의 기술 영역으로써 **4자리** 부터 좁은 범위의 기술 영역으로써 **8자리**의 표현을 가짐.
- ✓ 특히, 4자리 IPC는 특허의 기술 영역의 합리적이고, **분석적인 범주를 제공함.** (Van Zeebroeck et al., 2006; Park and Yoon, 2014)
- ✓ 한 개의 특허에서 **서로 다른 두 개의 기술 영역**이 내재되어 있는 경우, 공동 발현 IPC로 정의.

Literature Review

기술 융합 예측

- ✓ Lee et al. (2015) 는 IPC 공동발현 정보에 연관 규칙을 적용한 후, Link prediction을 사용해, 기술 융합 예측함
- ✓ Han and Sohn (2016) 은 IPC 정보에 엔트로피 및 Gravity 개념을 도입해, 기술 융합 트렌드를 파악함.
- ✓ Curran and Leker (2011) 는 IPC 공동 발현 정보를 사용해, 화장품과 영양학의 융합 트렌드를 관찰함.

기존 연구의 한계점

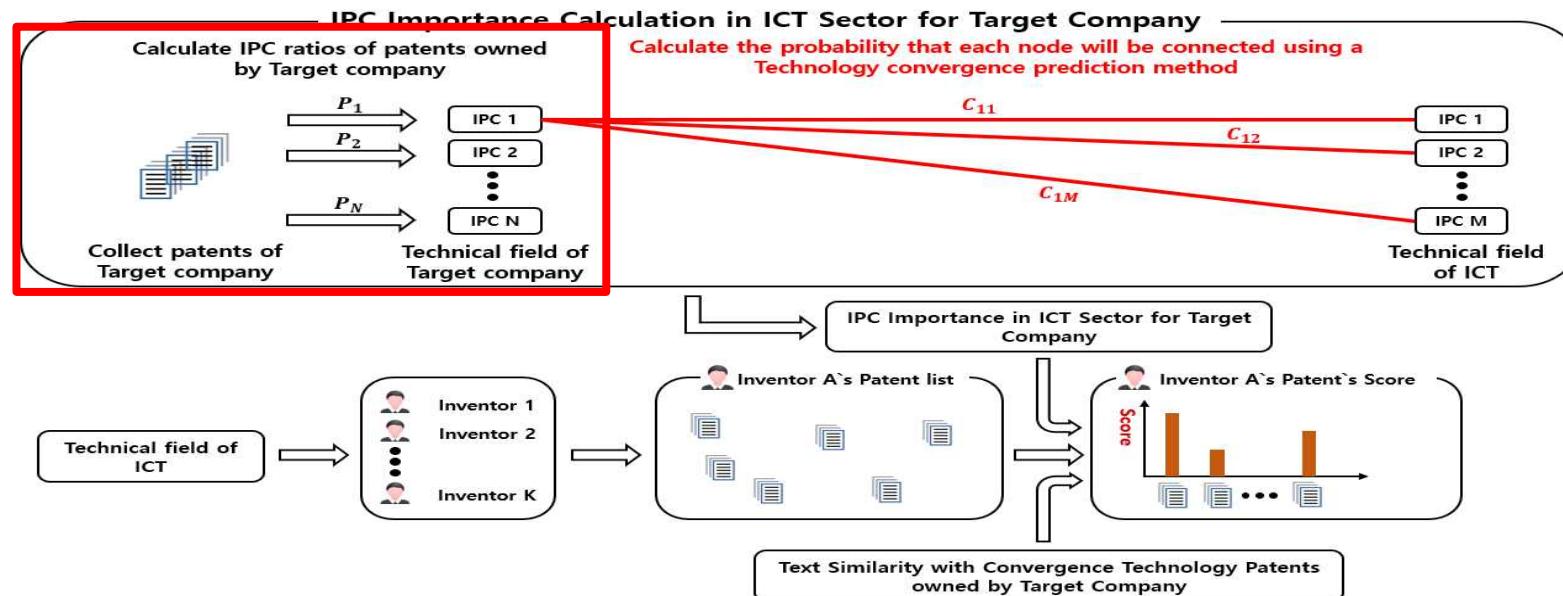
- ✓ IPC 분류 체계는 완벽하지 않기 때문에, 오분류가 있을 수 있음.
- ✓ 미래의 기술 융합 트렌드는 과거의 융합 정보에 의해서만 결정되는 것이 아님.
- ✓ 따라서 IPC 정보에만 의존한 분석은 신뢰성이 부족할 수 있음.



따라서 본 연구는, IPC 정보 뿐만 아니라 특허의 서지학적 정보를 추가적으로 활용.
(Kim et al., 2019)

Framework

✓ R&D 고급 인력 탐색 프레임워크

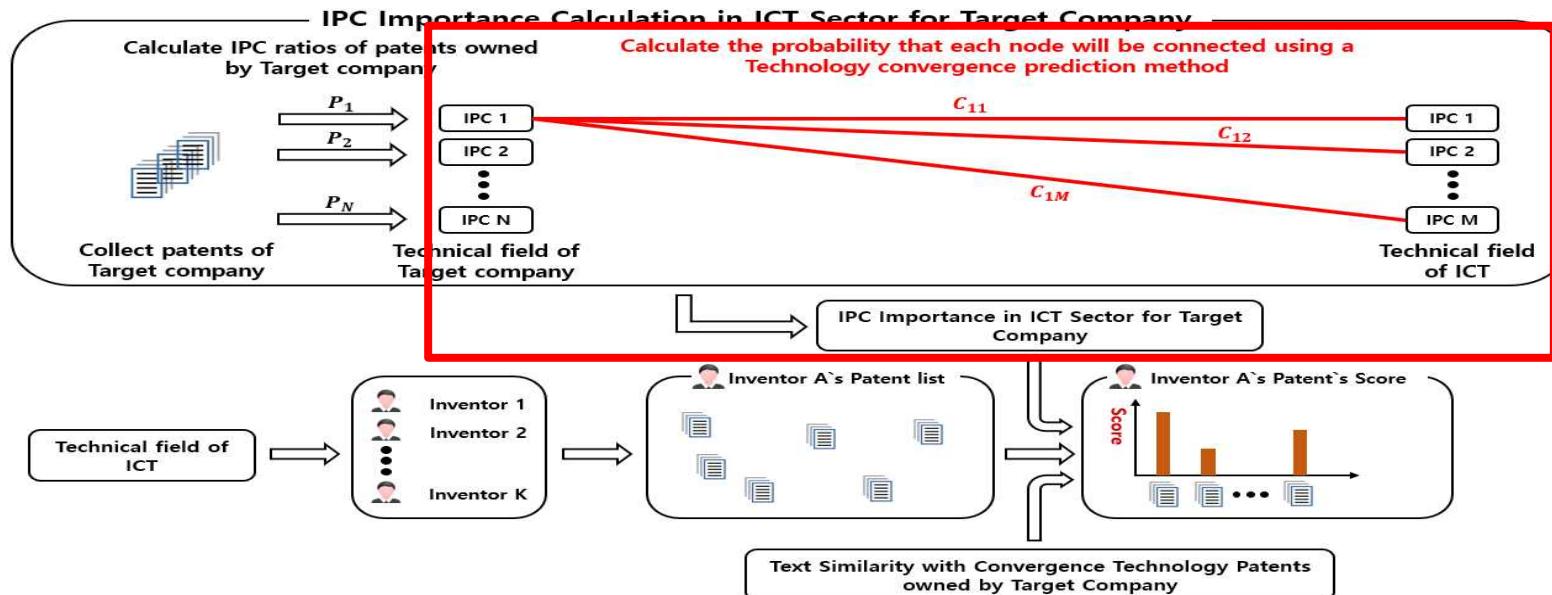


Step 1 (기업의 보유 기술 정도 파악)

- ✓ 기업의 주력 기술 분야에서 기업이 보유한 특허의 IPC4자리 정보를 수집.
- ✓ 기업이 보유한 특허를 IPC 4자리 기준으로 IPC i 에 해당하는 비율 P_i 계산.
- ✓ 예를 들어, 기업이 100개의 특허를 보유하고 있을 때, IPC i 에 속한 특허가 30개라면, $P_i = 0.3$

Framework

✓ R&D 고급 인력 탐색 프레임워크

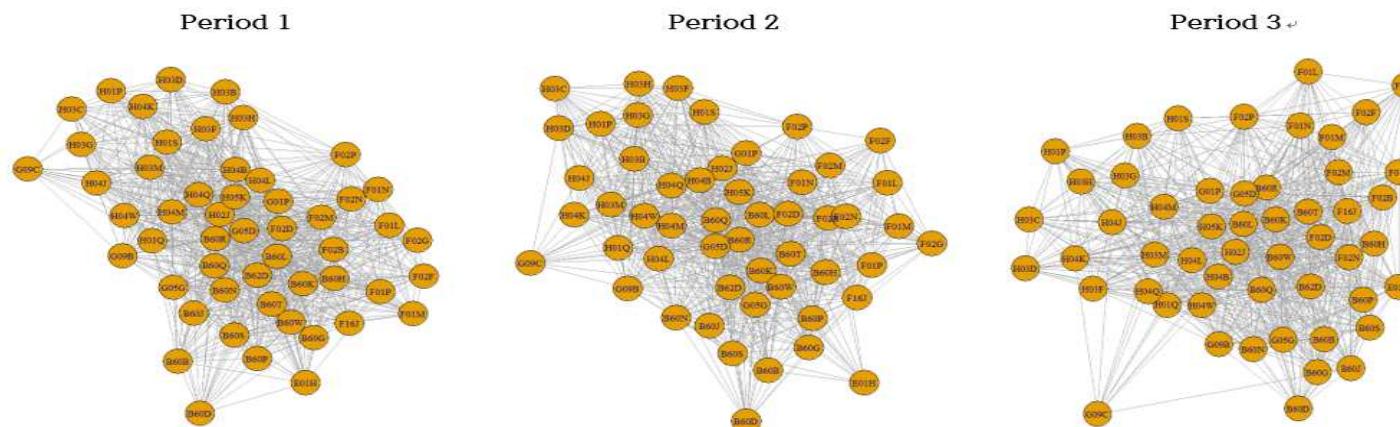
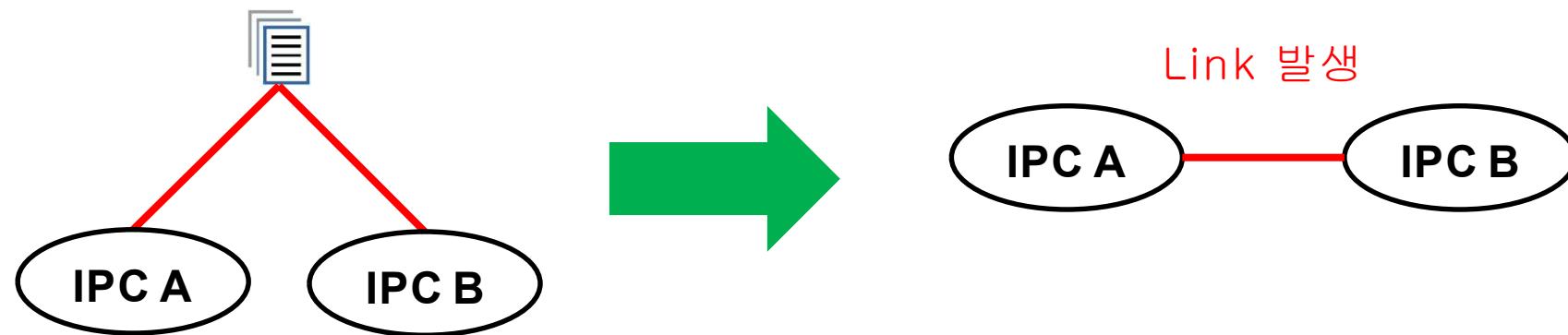


Step 2 (ICT 분야의 IPC 중요도 계산)

- ✓ USPTO에서 T기간 동안 기업의 주력 기술 분야와 ICT 분야에 출원된 모든 특허를 수집.
- ✓ 전체 T 구간을 동일한 길이의 3개의 구간으로 분할.
- ✓ 머신러닝 알고리즘을 사용하여, 미래에 기업의 주력 기술 분야에 속하는 IPC i 와 ICT 분야에 속하는 IPC j 사이에 융합이 발생할 확률 C_{ij} 를 계산.
- ✓ 이 때, 기업의 입장에서 ICT 분야에 속하는 IPC j 의 중요도는 $Importance (ICT_j) = \sum_{i=1}^N P_i \times C_{ij}$ 로 계산됨.

Framework

1. 각 구간별로 IPC 공동발현 네트워크 생성



Framework

☒ 2. 기술 융합 확률 C_{ij} 계산

Period 1 Input Feature Period 2 IPC 페어 Link
머신러닝 알고리즘 훈련

Period 2 Input Feature Period 3 IPC 페어 Link
훈련된 알고리즘 성능 검증

Period 3 Input Feature 미래의 IPC 페어 Link
예측

IPC i 와 IPC j 사이에
융합이 발생할 확률 C_{ij}

Framework

(1) Input Feature 구성

Link Prediction 기반의 18가지 유사성 지표

- ✓ Link Prediction 기반의 유사성 지표는 크게 Local, Global, Quasi-Local로 분류됨. (Lu and Zhou, 2011)
- ✓ Local : 노드 페어들의 이웃들만 고려, Global : 네트워크 전체 고려, Quasi-Local : 중간 단계

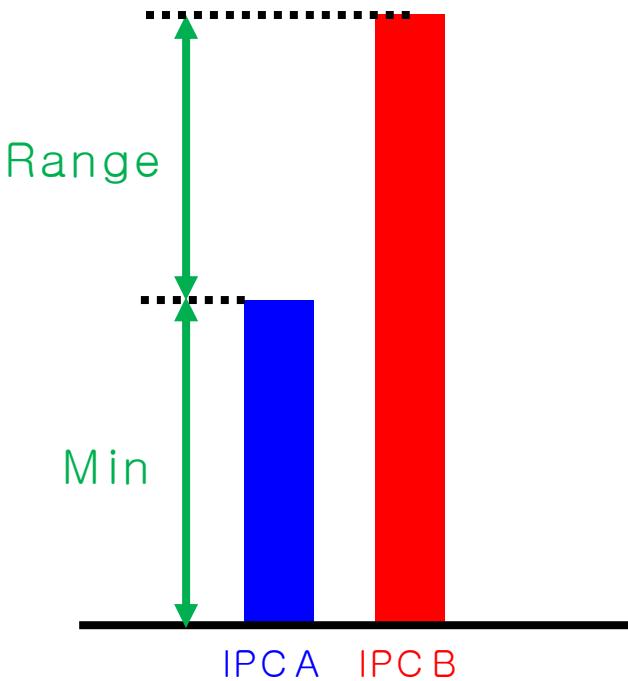
Local		Global		Quasi-Local	
Common Neighbors	Hub Depressed Index	Katz Index	Normalized ACT	Shortest Paths	Local Path Index
Salton Index	Leicht-Holme-Newman (Local)	Leicht-Holme-Newman (Global)	Random Walk with Restart		
Jaccard Index	Preferential Attachment	Average Commute Time	Matrix Forest Index		
Sorensen Index	Adamic-Adar Index				
Hub Promoted Index	Resource Allocation Index				

Framework

(1) Input Feature 구성

19가지 서지학적 정보

IPC의 특성 변수	
특허의 개수	IPC 성장 지표
특허의 성장율	
평균 발명자 수	
평균 청구항 수	
평균 후방 특허 인용 수	IPC 복잡성 지표
평균 후방 비특허 인용 수	
평균 Family 크기	
평균 전방 인용 수	IPC 지식 유동성 지표

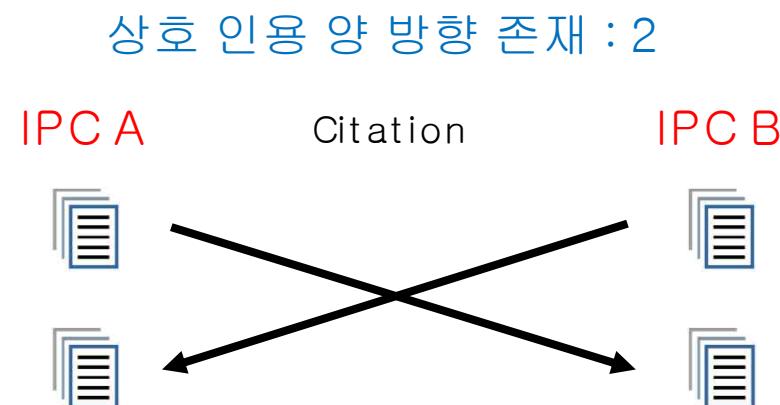
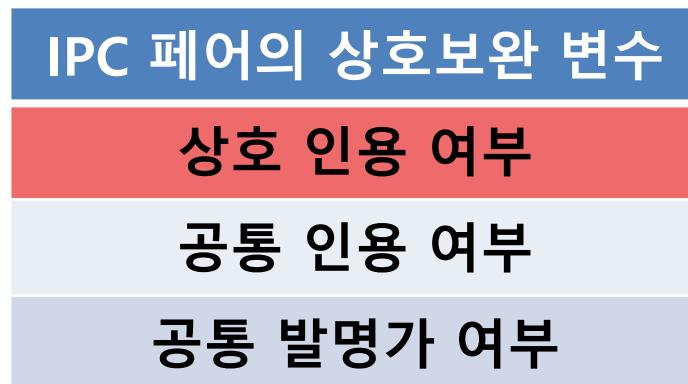


IPC 페어간의 변수로 Range 와 Min 추가

Framework

☒ (1) Input Feature 구성

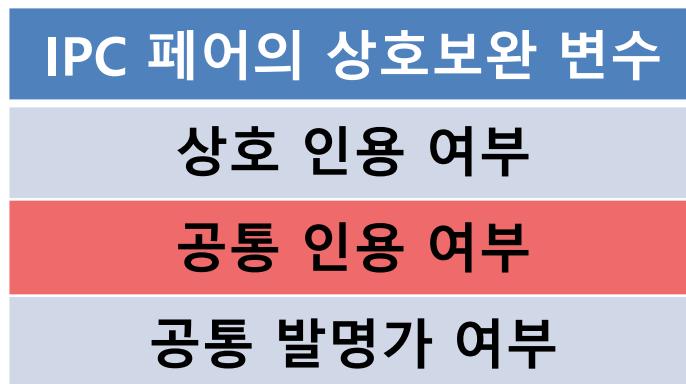
19가지 서지학적 정보



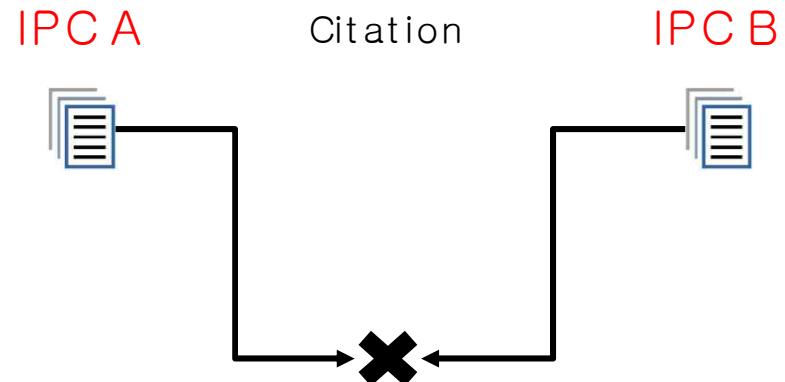
Framework

(1) Input Feature 구성

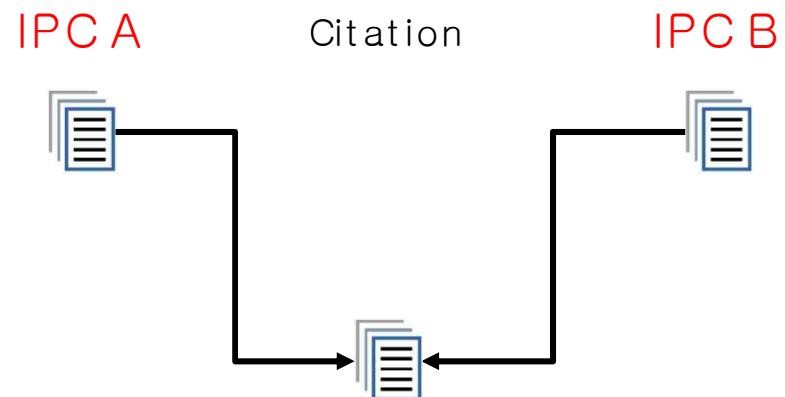
19가지 서지학적 정보



공통 인용 없으면 : 0



공통 인용 존재 : 1



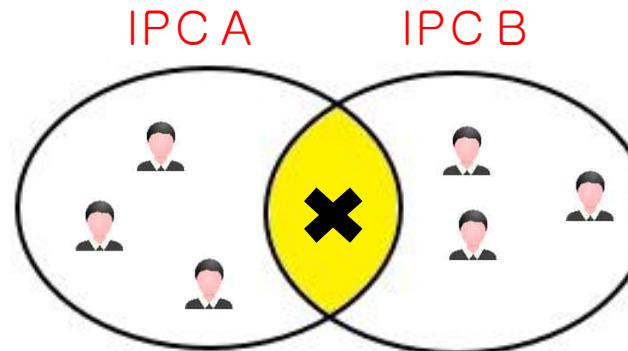
Framework

(1) Input Feature 구성

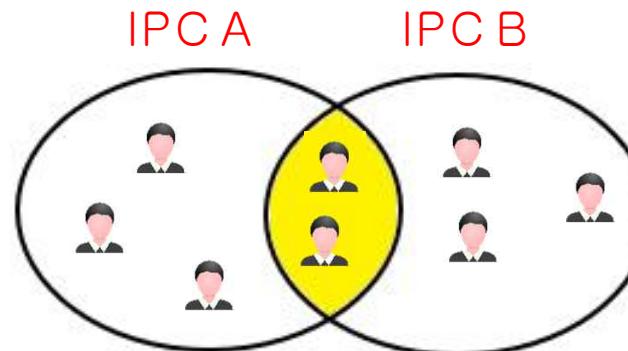
19가지 서지학적 정보

IPC 페어의 상호보완 변수
상호 인용 여부
공통 인용 여부
공통 발명가 여부

공통 발명가 없으면 : 0



공통 발명가 존재 : 1



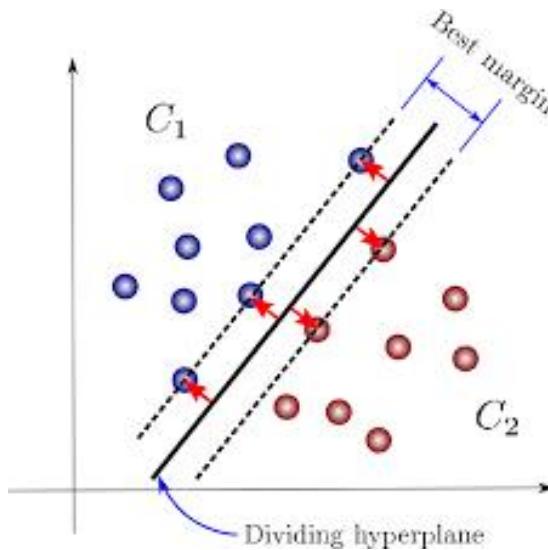
Framework

☒ (2) 머신러닝 알고리즘

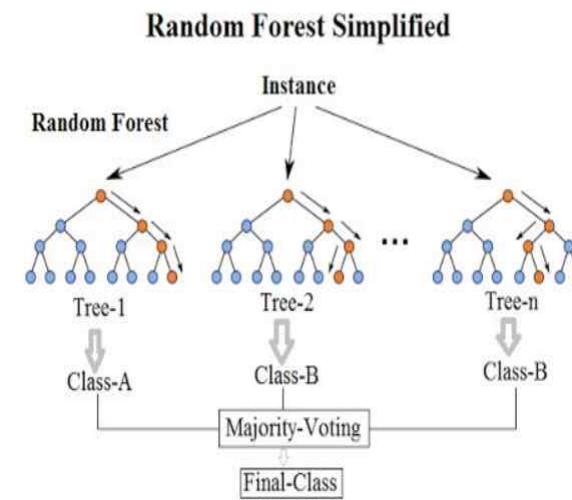
의사 결정 트리



SVM



RandomForest



Framework

☒ (2) 머신러닝 알고리즘

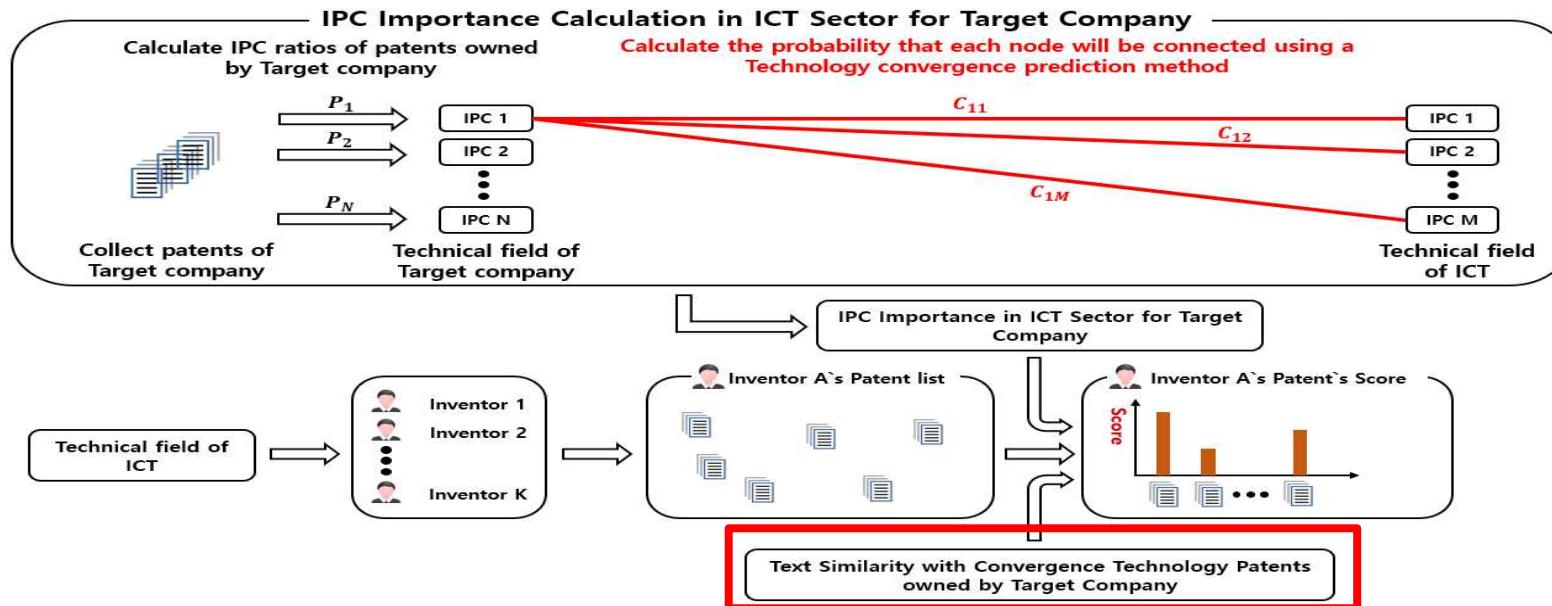
- ✓ 성능 평가 지표 : Accuracy, Precision, Recall, F1 Score
- ✓ 최적의 Hyperparameter를 찾기 위해 Grid Search가 수행됨. (Accuracy 기준)

Grid Search

Algorithm	Hyperparameter	Value
C5.0	Model	Tree, Rule
	Predictor Winnowing	T / F
	Fuzzy	T / F
	Trial	1 to 100
SVM	Gamma	$2^{-5, -4, \dots, 5}$
	Cost	$2^{0, 1, 2, 3, 4, 5}$
RandomForest	Number of candidate variables at each split	3 to p
	Number of Tree	100, 200, ..., 2500

Framework

✓ R&D 고급 인력 탐색 프레임워크

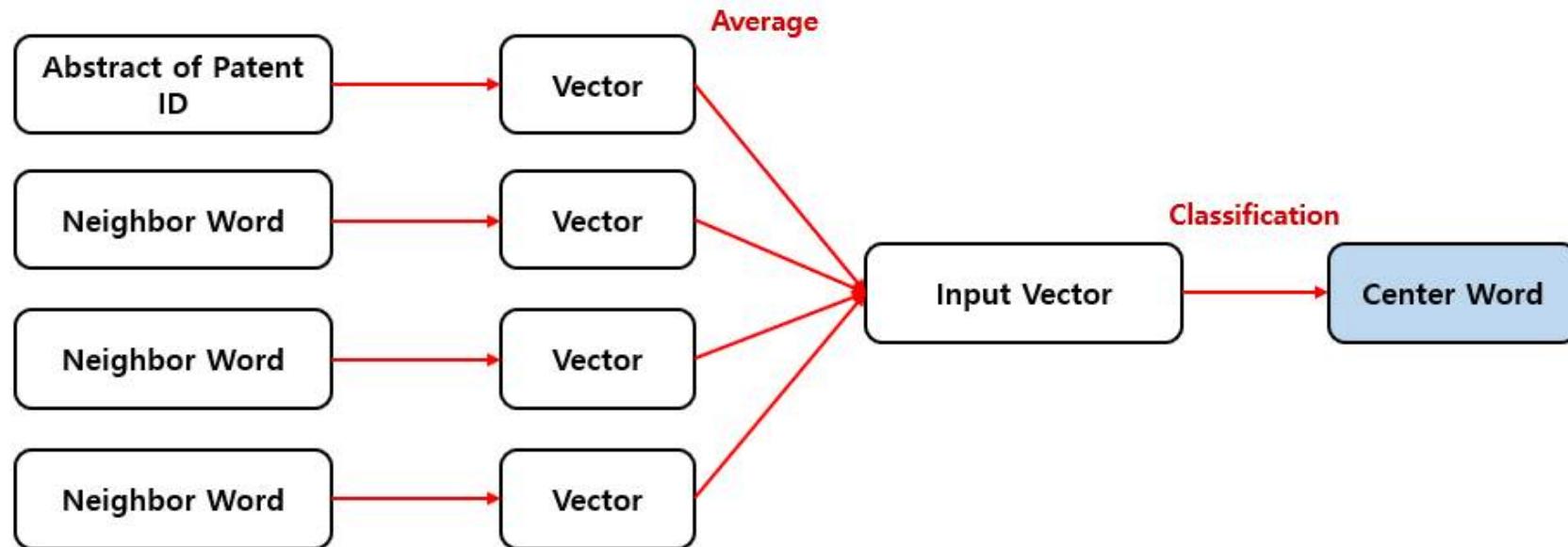


Step 3 (기업이 보유한 융합 기술과의 유사성)

- ✓ IPC 중요도 뿐만 아니라, 기업이 이미 소유한 융합 기술 특허와의 유사성 또한 고려되어야 함.
- ✓ IPC 중요도가 높더라도, 유사성이 낮으면 기업의 기술과의 융합이 현실적으로 발생하기 어려움.
- ✓ 특허의 초록을 Doc2vec 방법을 통해 벡터화 한 후, 두 특허의 초록 벡터의 Cosine 유사성에 의해 두 특허의 유사성이 계산됨.

Framework

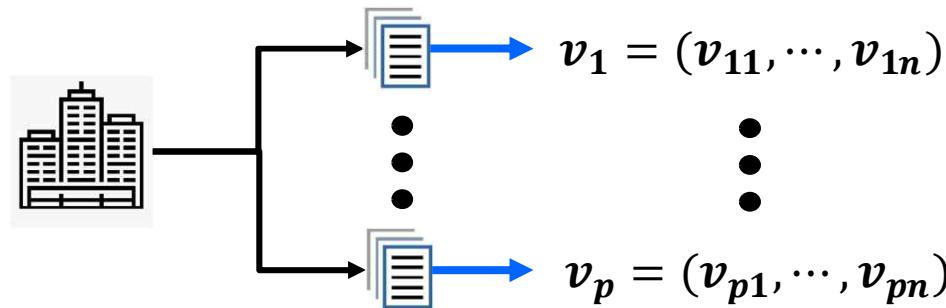
Doc2vec (Le & Mikolov, 2014)



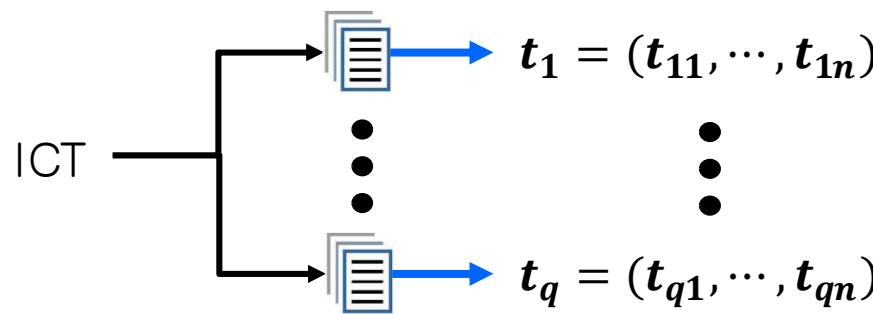
- ✓ Word2vec (Mikolov et al., 2013)을 기반으로 함.
- ✓ Word2vec은 주변단어로부터 중심단어가 나올 확률을 최대화하는 방향으로 단어 벡터의 유사성을 부여.
- ✓ Doc2vec은 Word2vec의 학습 과정에서, Document ID를 추가한 방법.

Framework

☒ 두 특허의 유사성



기업의 융합기술
특허 초록 수집



ICT 기술
특허 초록 수집

특허 k (ICT)

$$t_k = (t_{k1}, \dots, t_{kn})$$

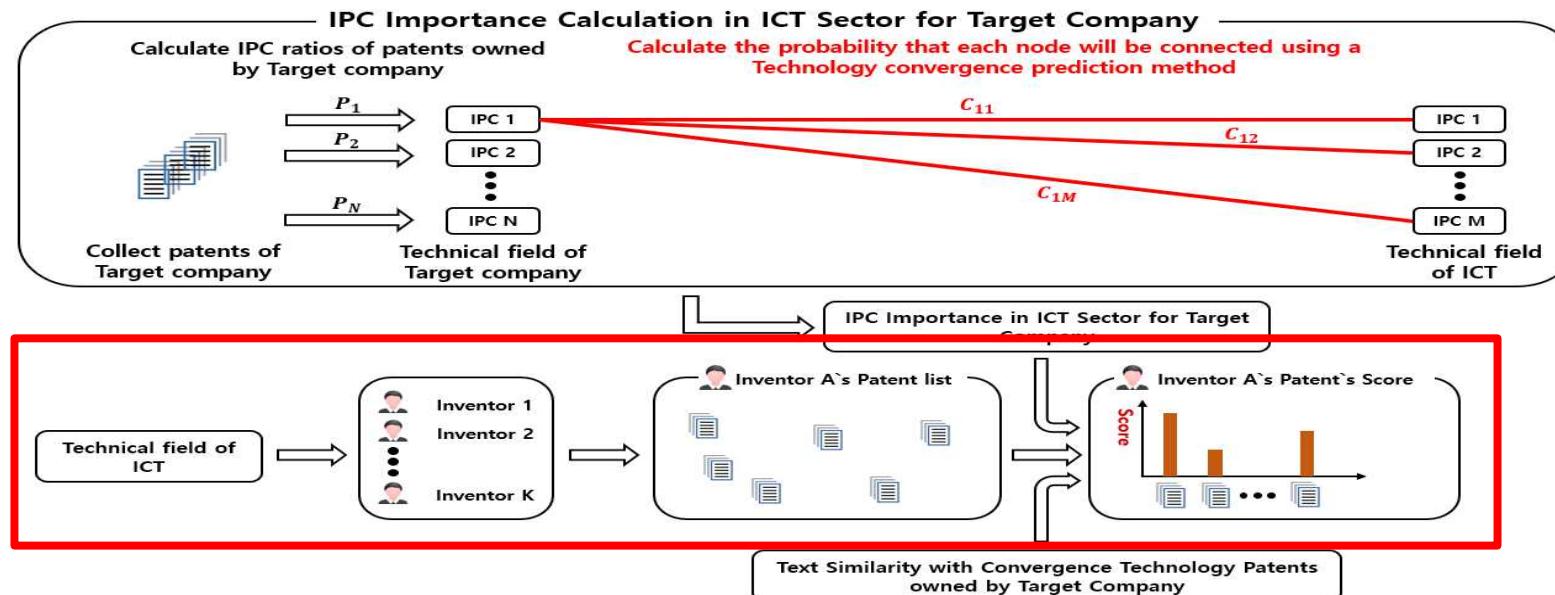
$$\text{Similarity} (\text{Patent}_k) = m \max_{w=1,2,\dots,p} \left(\frac{v_w \cdot t_k}{\|v_w\| \times \|t_k\|} \right)$$

Doc2vec을 이용한
벡터화

기업이 보유한 융합기술들과의
유사성 계산

Framework

✓ R&D 고급 인력 탐색 프레임워크



Step 4 (특허 발명가에 대한 평가)

- ✓ IPC 중요도와 융합 기술 특허와의 유사성에 의해 특허의 점수가 평가됨.

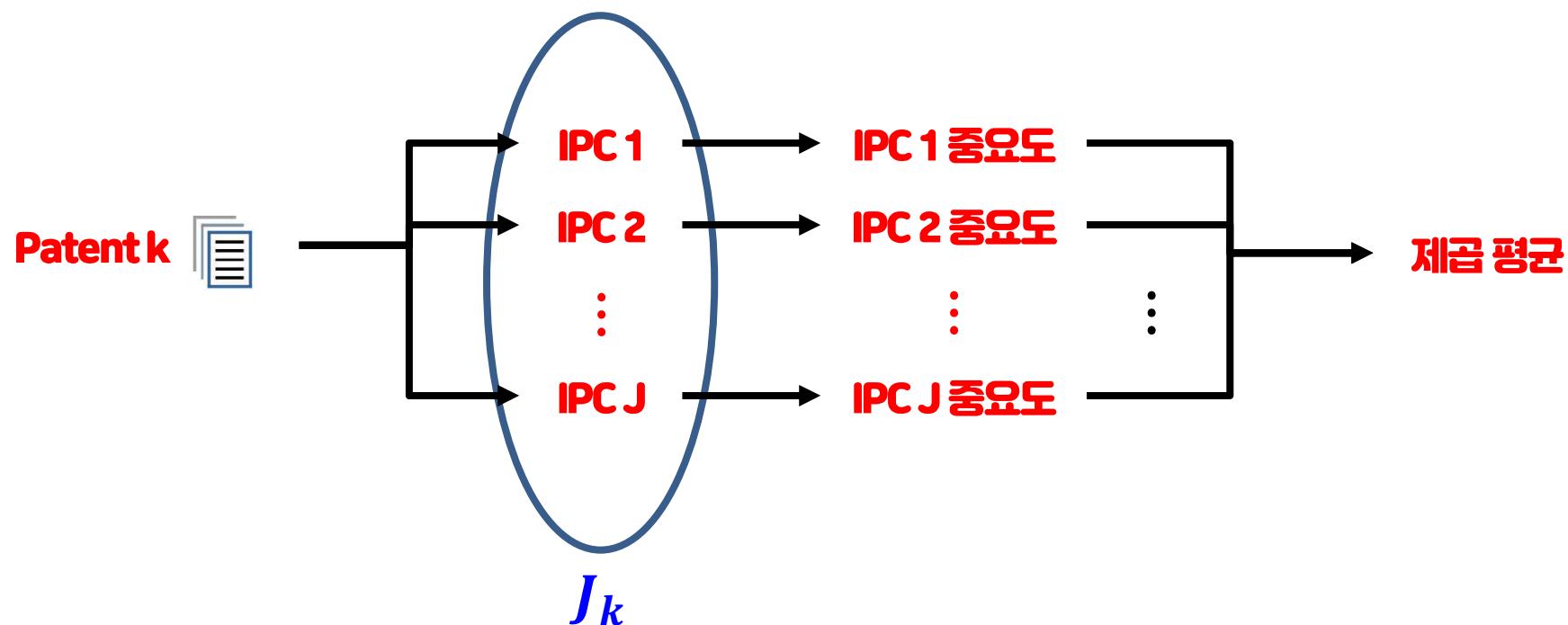
$$\checkmark Score (Patent_k) = \sqrt{\frac{\sum_{j \in J_k} Im\ portance (ICT_j)^2}{|J_k|}} \times Sim\ larity (Patent_k) \times c_k$$

✓ 특허 발명가는 그들이 발명에 참여한 특허들의 점수를 합산하여 평가됨.

Framework

▣ 특허의 점수

✓ Score (Patent_k) = $\sqrt{\frac{\sum_{j \in J_k} \text{Importance} (\text{IC}T_j)^2}{|J_k|}} \times \text{Similarity} (\text{Patent}_k) \times c_k$

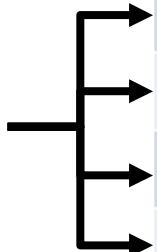


Framework

☑ 특허의 점수

✓ $Score(Patent_k) = \sqrt{\frac{\sum_{j \in J_k} Importance(CT_j)^2}{|J_k|}} \times Sim\;sim(Patent_k) \times c_k$

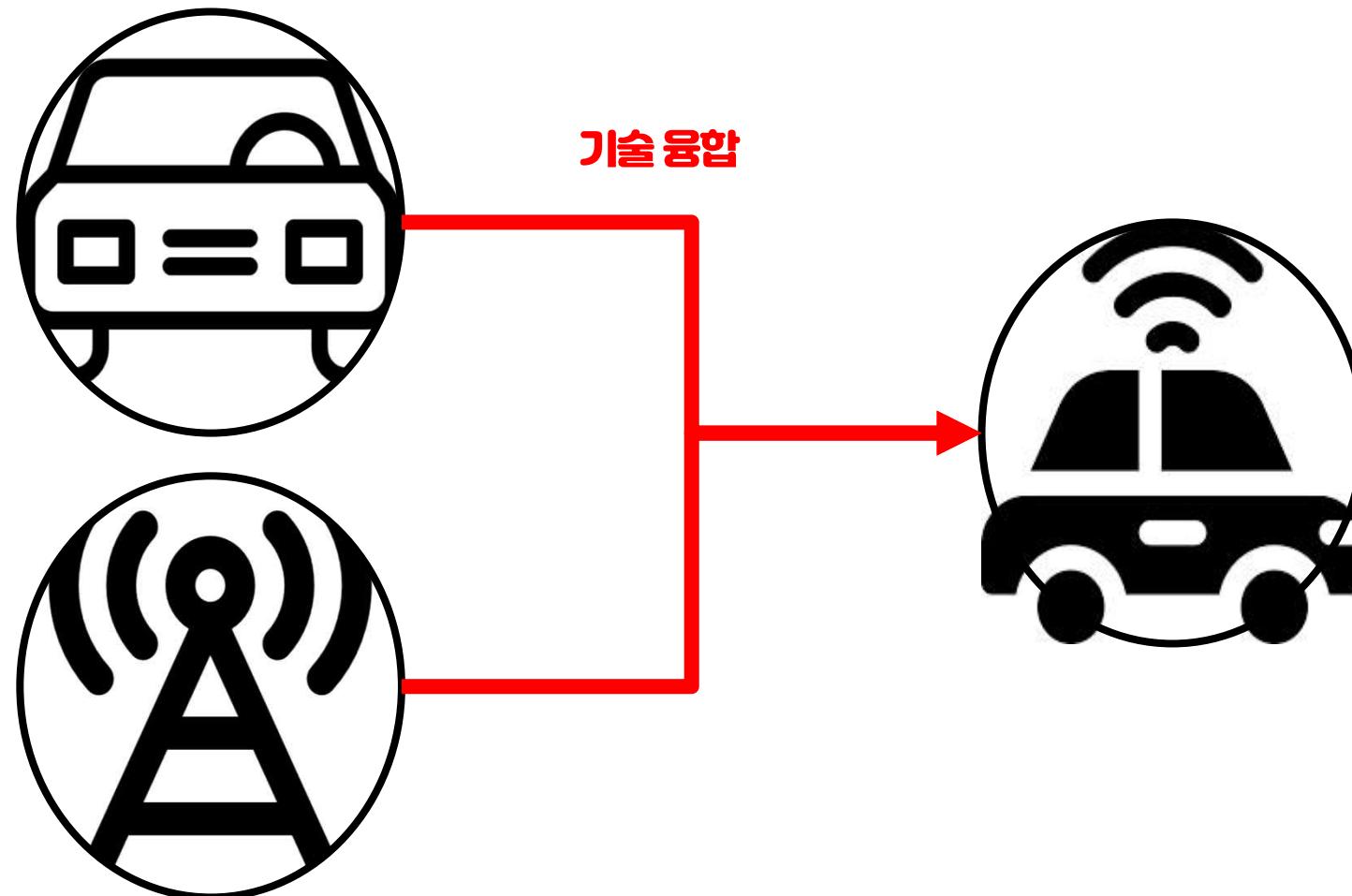
Patent k



특허의 가치 평가 지표 c_k	의미
$c_k = 1$	모든 특허의 가치는 동일
$c_k =$ 청구항 수	기술 중요성 (Lanjouw and Schankerman, 2004)
$c_k =$ 후방인용 수	기술 복잡성 (Harhoff et al., 2003)
$c_k =$ 전방인용 수	경제적 잠재성 (Fischer and Leidinger, 2014)

Case Study

☒ 자율주행 자동차



Case Study

데이터 수집

- ✓ 현대자동차가 KIPO에 출원한 자동차 분야의 특허 62,848 개
- ✓ 2002년부터 2016년 (T=15) USPTO에 출원된 자동차 분야와 ICT 분야의 특허 279,884, 768,989 개



기술 융합 예측 머신러닝 알고리즘 Grid Search 결과

Condition	Model	Optimal Hyperparameter Setting	Accuracy
Link prediction based similarity	C5.0	Model = tree, Predictor Winnowing = T, Fuzzy = T, Trials = 10	0.8356
	SVM	gamma = , cost =	0.8233
	Random Forest	Number of candidate variables = 7 Number of Tree = 200	0.7877
Link prediction based similarity + Bibliographic Information	C5.0	Model = tree, Predictor Winnowing = F, Fuzzy = T, Trials = 44	0.8479
	SVM	gamma = , cost = 2	0.8157
	Random Forest	Number of candidate variables = 11 Number of Tree = 1900	0.8402

Case Study

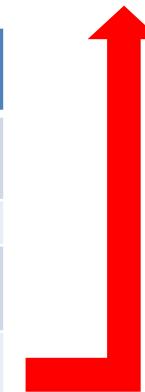
데이터 수집

- ✓ 현대자동차가 KIPO에 출원한 자동차 분야의 특허 62,848 개
- ✓ 2002년부터 2016년 (T=15) USPTO에 출원된 자동차 분야와 ICT 분야의 특허 279,884, 768,989 개

기술 예측 머신러닝 알고리즘

Performance indicators	Value
Accuracy	0.8479
Precision	0.901
Recall	0.723
F1 score	0.802

Condition	Model	Optimal Hyperparameter Setting	Accuracy
Link prediction based similarity	C5.0	Model = tree, Predictor Winnowing = T, Fuzzy = T, Trials = 10	0.8356
	SVM	gamma = , cost =	0.8233
	Random Forest	Number of candidate variables = 7 Number of Tree = 200	0.7877
Link prediction based similarity + Bibliographic Information	C5.0	Model = tree, Predictor Winnowing = F, Fuzzy = T, Trials = 44	0.8479
	SVM	gamma = , cost = 2	0.8157
	Random Forest	Number of candidate variables = 11 Number of Tree = 1900	0.8402



Case Study

특허 발명가 탐색 전략

- ✓ 최근 3년 동안 특허 발명 활동이 있는 265,650명의 발명가에 대해서 탐색 수행.
- ✓ 일반적으로, 특허 발명 횟수가 많은 발명가일수록 고용 비용이 증가함.
- ✓ 따라서 기업의 인사 책임자는 특허 발명가의 평가와 고용 비용 모두를 고려하여야 함.
- ✓ 본 연구에서는, 발명 특허의 개수를 기준으로 발명가 탐색을 수행함.
- ✓ Case Study로 특허 발명 수가 1개, 5개인 경우만 탐색.

Patent invention number	Number of inventors	Proportion (%)
1	135,831	51.13
2	47,330	17.82
3	23,048	8.68
4	13,780	5.19
5	9,108	3.43
6	6,332	2.38
7	4,692	1.77
8	3,588	1.35
9	2,757	1.04
Otherwise	19,184	7.21

Case Study

특허의 모든 가치 동일

Patent	Rank	Name	Score	Affiliation
1	1	Jang, Deok Suk, Lee, Yong Sam	0.6603	Samsung Electro-Mechanics Co.
	2	Kim, Young Hwan, Sung, Chang Hyun	0.6302	LG Chem Ltd.
	3	Coste, Franc Ois, Jean, Guillaume	0.6187	Intelligent Electronic Systems
5	1	Yun, Sung Jin	2.7631	LG Innotek Co.
	2	Kimura Naoki, Morimoto, Toyata, Matsumoto Manabu, Masubuchi, Hayato	2.3613	Toshiba Memory Co.
	3	Im, Hyun Gu	2.3363	LG Innotek Co.

- ✓ 이 관점은 현대자동차에게 연구의 프레임워크에 가장 적합한 후보군이 탐색됨.
- ✓ 한국 기업에 소속된 발명가들이 높은 순위에 위치함.

Case Study

기술 중요성

Patent	Rank	Name	Score	Affiliation
1	1	Wang, Tiehong A.	27.7463	Virginia Innovation Sciences
	2	Bhatti, Sohaib, Smith, Nicholas, Howard, Jeffrey	26.8094	Ionic Security
	3	Wolzien, Thomas A., Milne, William, Maisey, Alexander, Thaler, Laurence, Porpiglia, Tom	26.8315	The Video Call Center
5	1	Gazdzinski, Mark J.	181.6564	West View Research
	2	Mcgowan, Daniel	122.7609	Continental Circuits
	3	Geremia, Peter P.	90.6262	Centripetal Networks

- ✓ 이 관점은 현대자동차에게 높은 수준의 기술 전문성을 보유한 후보군이 탐색됨.
- ✓ 미국의 아틀란타에 위치하는 Ionic Security 에 속한 발명가가 높은 순위에 있는 것을 보아, 자율 주행 자동차에 데이터 보안이 큰 이유인 것이 반영된 것으로 보임.

Case Study

기술 복잡성

Patent	Rank	Name	Score	Affiliation
1	1	Piernot Philippe P., Naik Devang K.	1349.84	Apple
	2	Sinha Anoop K.	1231.51	Apple
	3	Jin, Robert Y.	1032.66	Abbott Diabetes Care
5	1	Yoscovich, Ilan	1944.82	Solaredge Technologies
	2	Leonard, Chantal M,	1447.97	Microsoft Technology Licensing
	3	Zaman, Nazia	1447.24	Microsoft Technology Licensing

- ✓ 이 관점은 현대자동차에게 다양한 기술을 적용할 수 있는 능력을 보유한 후보군이 탐색됨.
- ✓ 한 개의 특허를 발명한 발명가 중에서는 Apple에 속한 발명가들이 상대적으로 높은 점수를 받음.
- ✓ 5 개의 특허를 발명한 발명가 중에서는 이스라엘 국적의 Yoscovich, Ilan이 압도적으로 높은 점수를 받음.
- ✓ 또한, Microsoft에 속한 발명가들도, 유능한 것으로 보임.

Case Study

경제적 잠재성

Patent	Rank	Name	Score	Affiliation
1	1	Float, Jamison J., Scheib, Charles, J., Miller, Christopher C., O'kelly, Matthew E.	10.7356	Ethicon Endo-Surgery
	2	Overmyer, Matthew E., Lytle, IV, Thomas W., Adams, Shane R.	10.7133	Ethicon Endo-Surgery
	3	Lau, Vincent Kin Nang	6.4116	Hong Kong University of Science and Technology
5	1	John, Howard Hall	26.5564	Oracle International Corporation
	2	Laracey, Kevin	23.6179	Paypal
	3	Barnickel, Jr. Donald J.	23.3231	AT&T Inc.

- ✓ 이 관점은 현대자동차에게 **기술적 우위를 바탕으로 경제적 이익을 창출**할 수 있는 후보군이 탐색됨.
- ✓ 의료 기계 제조업체인 **Ethicon Endo-Surgery**가 높은 순위에 위치한 것을 보아, 최근에 의료 분야 또한 ICT 분야와의 융합이 주요한 것에 영향을 받은 것으로 보임.
- ✓ 또한, 기존 탐색과는 다르게 **대학 소속의 발명가**가 높은 점수를 받은 것으로 관측 됨.

Conclusion

본 연구를 통해서,

기업이 융합 대상 분야의 R&D 인력을 탐색 및 모집 하는 데 투자 되는 시간적 비용을 크게 줄여줄 수 있을것이라 기대됨.



기업의 보유 기술 중심으로 미래의 융합 트렌드가 함께 고려되기 때문에, 기업의 미래 발전 방향에 부합하는 고급 인력이 탐색 될 것으로 기대됨.



Conclusion

본 연구의 한계점은,

본 연구에서 반영되는 미래의 융합 트렌드는 2017년부터 2021사이로, 2019년인 현재 시점에서 실용성이 떨어질 수 있음.

→ 만약, 2019년까지의 데이터를 수집할 수 있다면, 2020년부터 2024년 사이의 융합트렌드를 예측함으로써, 보다 실용적인 결과를 얻을 수 있음.

특히 가치를 바라보는 네 가지 관점에서, 탐색이 수행되었지만 기업은 포괄적인 하나의 후보군 탐색을 필요로 할 수 있음.

→ 본 연구에서 수행되지 않았지만, 기업 입장에서 Analytic Hierarchy Process 등 방법을 사용하여, 각 관점에 대한 Weight를 설정하면 포괄적인 하나의 후보군을 제시할 수 있음.

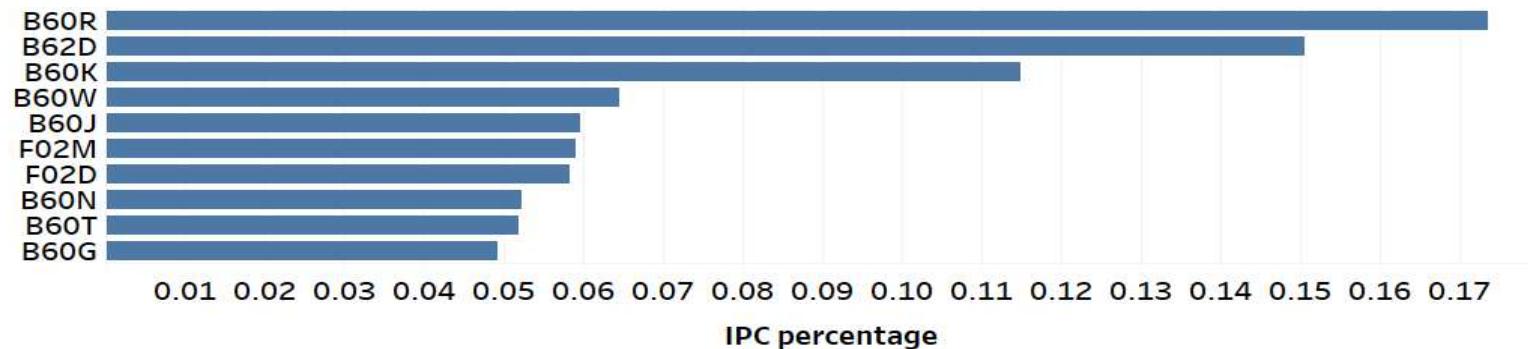
실제로 기업에서 인력 탐색 전략을 수립할 때, 발명가의 나이, 경력, 직급 등의 다양한 정보가 필요함.

→ 결여되어 있는 다양한 정보가 추가된다면, 보다 현실적인 후보군이 제시 될것으로 보임.

감사합니다

Appendix

현대자동차 소유 특허의 IPC 비율



IPC Importance

IPC	G09B	G09C	H01P	H01Q	H01S	H02J	H03B	H03C
Importance	0.77	0.204	0.277	0.735	0.499	1.058	0.332	0.126
IPC	H03D	H03F	H03G	H03H	H03M	H04B	H04J	H04K
Importance	0.153	0.173	0.362	0.349	0.459	0.893	0.522	0.46
IPC	H04L	H04M	H04Q	H04W	H05K			
Importance	0.972	0.824	0.753	0.965	1.02			

Reference

- Abramovici, M., Göbel, J. C., & Neges, M. (2015). Smart engineering as enabler for the 4th industrial revolution. In Integrated systems: Innovations and applications (pp. 163-170). Springer, Cham.
- Adamic, L. A., & Adar, E. (2001, May). You are what you link. In 10th annual International World Wide Web Conference, Hong Kong. Retrieved June (Vol. 19, p. 2001).
- Akinyele, S. T. (2007). Determination of the optimal manpower size using linear programming model. Research Journal of Business Management, 1(1), 30-36.
- Ali, J., & Kumar, S. (2011). Information and communication technologies (ICTs) and farmers' decision-making across the agricultural supply chain. International Journal of Information Management, 31(2), 149-159.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509-512.
- Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
- Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems, 25(6), 599-616.
- Cano, V., & Lind, N. C. (1991). Citation life cycles of ten citation classics. Scientometrics, 22(2), 297-312.
- Caviggioli, F. (2016). Technology fusion: Identification and analysis of the drivers of technology convergence using patent data. Technovation, 55, 22-32.
- Chebotarev, P., & Shamis, E. (2006). The matrix-forest theorem and measuring relations in small social groups. arXiv preprint math/0602070.
- Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. Genomics, 99(6), 323-329.

Reference

- Choi, H. S., Sohn, S. Y., & Yeom, H. J. (2018). Technological composition of US metropolitan statistical areas with high-impact patents. *Technological Forecasting and Social Change*, 134, 72-83.
- Choi, J., Jeong, S., & Kim, K. (2015). A study on diffusion pattern of technology convergence: Patent analysis for Korea. *Sustainability*, 7(9), 11546-11569.
- Curran, C. S., Bröring, S., & Leker, J. (2010). Anticipating converging industries using publicly available data. *Technological Forecasting and Social Change*, 77(3), 385-395.
- Curran, C. S., & Leker, J. (2011). Patent indicators for monitoring convergence—examples from NFF and ICT. *Technological Forecasting and Social Change*, 78(2), 256-273.
- Ernst, H. (2003). Patent information for strategic technology management. *World Patent Information*, 25(3), 233-242.
- Fischer, T., & Leidinger, J. (2014). Testing patent value indicators on directly observed patent value—An empirical analysis of Ocean Tomo patent auctions. *Research Policy*, 43(3), 519-529.
- Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10), 906-914.
- Gollakota, K. (2008). ICT use by businesses in rural India: The case of EID Parry's Indiagriline. *International Journal of Information Management*, 28(4), 336-341.
- Grimaldi, M., Cricelli, L., Di Giovanni, M., & Rogo, F. (2015). The patent portfolio value analysis: A new framework to leverage patent information for strategic technology planning. *Technological Forecasting and Social Change*, 94, 286-302.

Reference

- Guerry, M. A., & De Feyter, T. (2012). Optimal recruitment strategies in a multi-level manpower planning model. *Journal of the Operational Research Society*, 63(7), 931-940.
- Han, E. J., & Sohn, S. Y. (2016). Technological convergence in standards for information and communication technologies. *Technological Forecasting and Social Change*, 106, 1-10.
- Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, 32(8), 1343-1363.
- Haupt, R., Kloyer, M., & Lange, M. (2007). Patent indicators for the technology life cycle development. *Research Policy*, 36(3), 387-398.
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat*, 37, 547-579.
- Jaillet, P., Loke, G. G., & Sim, M. (2018). Risk-Based Manpower Planning: A Tractable Multi-Period Model. Available at SSRN 3168168.
- Jeeva, M., Rajagopal, R., Charles, V., & Yadavalli, V. S. S. (2004). An application of stochastic programming with Weibull distribution–cluster based optimum allocation of recruitment in manpower planning. *Stochastic Analysis and Applications*, 22(3), 801-812.
- Kareem, B., & Aderoba, A. A. (2008). Linear programming based effective maintenance and manpower planning strategy: a case study. *J Comput Internet Manag*, 16(2), 26-34.
- Karnouskos, S. (2018). Self-Driving Car Acceptance and the Role of Ethics. *IEEE Transactions on Engineering Management*. 1-14

Reference

- Karvonen, M., Kapoor, R., Uusitalo, A., & Ojanen, V. (2016). Technology competition in the internal combustion engine waste heat recovery: a patent landscape analysis. *Journal of Cleaner Production*, 112, 3735-3743.
- Karvonen, M., & Kässi, T. (2013). Patent citations as a tool for analysing the early stages of convergence. *Technological Forecasting and Social Change*, 80(6), 1094-1107.
- Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1), 39-43.
- Kim, T.S., & Sohn, S.Y. (2019). Machine-Learning-Based Deep Semantic Analysis Approach for Forecasting New Technology Convergence. Manuscript submitted for publication.
- Kodama, F., & Gardiner, P. (1996). Emerging patterns of innovation: Sources of Japan's technological edge. *R and D Management*, 26(2), 179-180.
- Lai, K. K., & Wu, S. J. (2005). Using the patent co-citation approach to establish a new patent classification system. *Information processing & management*, 41(2), 313-330.
- Lanjouw, J. O., & Schankerman, M. (2004). Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal*, 114(495), 441-465.
- Laurens, P., Villard, L., Schoen, A., & Larédo, P. (2018). The artificial patents in the PATSTAT database: How much do they matter when computing indicators of internationalisation based on worldwide priority patents?. *Scientometrics*, 114(1), 91-112.
- Le, Q., & Mikolov, T. (2014, January). Distributed representations of sentences and documents. In *International Conference on Machine Learning* (pp. 1188-1196).
- Lee, J., & Sohn, S. Y. (2017). What makes the first forward citation of a patent occur earlier?. *Scientometrics*, 113(1), 279-298.

Reference

- Lee, W. J., Lee, W. K., & Sohn, S. Y. (2016). Patent network analysis and quadratic assignment procedures to identify the convergence of robot technologies. *PloS one*, 11(10), e0165091.
- Lee, W. S., Han, E. J., & Sohn, S. Y. (2015). Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. *Technological Forecasting and Social Change*, 100, 317-329.
- Lei, D. T. (2000). Industry evolution and competence development: the imperatives of technological convergence. *International Journal of Technology Management*, 19(7-8), 699-738.
- Leicht, E. A., Holme, P., & Newman, M. E. (2006). Vertex similarity in networks. *Physical Review E*, 73(2), 026120.
- Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications*, 390(6), 1150-1170.
- Ma, J. X., Buhalis, D., & Song, H. (2003). ICTs and Internet adoption in China's tourism industry. *International Journal of Information Management*, 23(6), 451-467.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Newman, M. E. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2), 025102.
- Pandya, R., & Pandya, J. (2015). C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications*, 117(16), 18-21.
- Park, H., & Yoon, J. (2014). Assessing coreness and intermediarity of technology sectors using patent co-classification analysis: the case of Korean national R&D. *Scientometrics*, 98(2), 853-890.

Reference

- Rao, P. T., Reddy, P. R. S., & Suhasini, A. V. S. (2010). Optimal manpower recruitment by stochastic programming in graded manpower systems. International Journal of Systems Assurance Engineering and Management, 1(2), 178-185.
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. Science, 297(5586), 1551-1555.
- Ruddock, L. (2006). ICT in the construction sector: Computing the economic benefits. International Journal of Strategic Property Management, 10(1), 39-50.
- Salton, G., & McGill, M. J. (1983). Introduction to modern information retrieval. mcgraw-hill.
- Schellekens, M. (2015). Self-driving cars and the chilling effect of liability law. Computer Law & Security Review, 31(4), 506-517.
- Scotchmer, S. (1996). Protecting early innovators: should second-generation products be patentable? The Rand Journal of Economics, 322-331.
- Shang, M. S., Lü, L., Zeng, W., Zhang, Y. C., & Zhou, T. (2010). Relevance is more significant than correlation: Information filtering on sparse data. EPL (Europhysics Letters), 88(6), 68008.
- Song, C. H., Elvers, D., & Leker, J. (2017). Anticipation of converging technology areas—A refined approach for the identification of attractive fields of innovation. Technological Forecasting and Social Change, 116, 98-115.
- Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. Shanghai Archives of Psychiatry, 27(2), 130.
- Sørensen, T. J. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. I kommission hos

Reference

- Tong, H., Faloutsos, C., & Pan, J. Y. (2006, December). Fast random walk with restart and its applications. In Sixth International Conference on Data Mining (ICDM'06) (pp. 613-622). IEEE.
- Urmson, C. (2008). Self-driving cars and the urban challenge. *IEEE Intelligent Systems*, 23(2), 66-68.
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157), 468-472.
- Van Zeebroeck, N., de la Potterie, B. V. P., & Han, W. (2006). Issues in measuring the degree of technological specialisation with patent data. *Scientometrics*, 66(3), 481-492.
- Yen, L., Fouss, F., Decaestecker, C., Francq, P., & Saerens, M. (2007, May). Graph nodes clustering based on the commute-time kernel. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 1037-1045). Springer, Berlin, Heidelberg.
- Yoffie, D. B. (1996). Competing in the age of digital convergence. *California Management Review*, 38(4), 31.
- Zhang, Y., & Li, X. (2017). Uses of information and communication technologies in HIV self-management: A systematic review of global literature. *International Journal of Information Management*, 37(2), 75-83.
- Zhou, T., Lü, L., & Zhang, Y. C. (2009). Predicting missing links via local information. *The European Physical Journal B*, 71(4), 623-630.