
깃허브를 이용한 혁신 생태계 분석 : 자율주행 사례를 중심으로

김인조(석사과정)¹, 이학연(교수)²

¹서울과학기술대학교 데이터사이언스학과

²서울과학기술대학교 산업공학과

¹deulseong1121@ds.seoultech.ac.kr

²hylee@seoultech.ac.kr



Index

1. 서론
2. 선행연구
3. 연구방법
 - 1) 프레임워크
 - 2) 분석데이터 수집
 - 3) Co-contribution 기반 네트워크
 - 4) 네트워크 구축
4. 자율주행 프로젝트 네트워크
5. 자율주행 개발자 네트워크
6. 결론
7. 참고문헌

1. 서론

• 연구 배경

- 기술 혁신을 통한 경쟁력 확보는 글로벌 시장에서 경쟁해야 하는 기업들의 필수 생존 전략
- 기술 혁신은 단일 조직이 아닌 여러 기업, 대학들의 리소스가 하나로 결합될 때 발생하기 때문에 비즈니스 생태계(ecosystem) 파악은 혁신을 예측하는데 매우 중요한 역할(Huhtanaki & Rubens, 2016)

• 연구 동기

- 기술 및 사회 환경의 불확실성, 변동성 등으로 인해 장기적인 관점의 기술 예측보다는 단기적 관점의 기술 나우캐스팅이 필요
- 기존의 특허 기반의 기술 예측은 출원 이후 등록까지의 시간이 오래 걸린다는 특징으로 인해 단기적인 기술 생태계 분석에는 부적합 (Wang & Hagedoorn, 2014)
- 기업, 학교, 개인이 직접 개발에 참여하는 오픈소스 데이터는 그 자체로 거대한 기술 생태계를 이루고 있으며, 최신의 기술 트렌드를 반영(Munaiah et al., 2017)
- 그 중 GitHub는 코드 버전관리 소프트웨어인 Git을 통해 관리되는 코드를 호스팅하고 협업 기능을 지원하는 웹서비스로 5천만명에 개발자들이 이용중인 플랫폼이며 다양한 오픈소스의 도서관(Kalliamvakou et al., 2015; Mergel, 2015)

• 연구 목적

- 깃허브 데이터를 활용하여 특정 산업군을 구성하는 기업들의 생태계(ecosystem) 및 요소 기술들을 분석하는 방법론 개발
- 자율주행분야는 이미지 처리, 통신, 제어 시스템 등 다양한 분야의 기술이 집약된 융합 기술이고, 최근 IT 기업들을 중심으로 활발히 연구되고 있는 분야이므로 본 연구의 케이스 스터디로 선정

2. 선행 연구

- Innovation ecosystem 분석

- 조직 및 개인의 혁신적 성과를 위한 행위 및 제도를 innovation ecosystem으로 정의 (Granstrand & Holgersson, 2020)
- 네트워크 분석 기법을 이용한 innovation ecosystem 분석 (Porter, 2020; Cai et al., 2019; Huhtanaki & Rubens, 2016)
- SNA 기반의 혁신 트렌드 분석 기법 제안 (Kim et al., 2014)
- 혁신 네트워크의 시각화 기법에 관한 연구 (Faber et al., 2018; Reimhofer et al., 2018)



혁신 네트워크 내에서 향후 유망 기술을 예측하기 위해 SNA 기법을 활용

- GitHub 데이터 분석

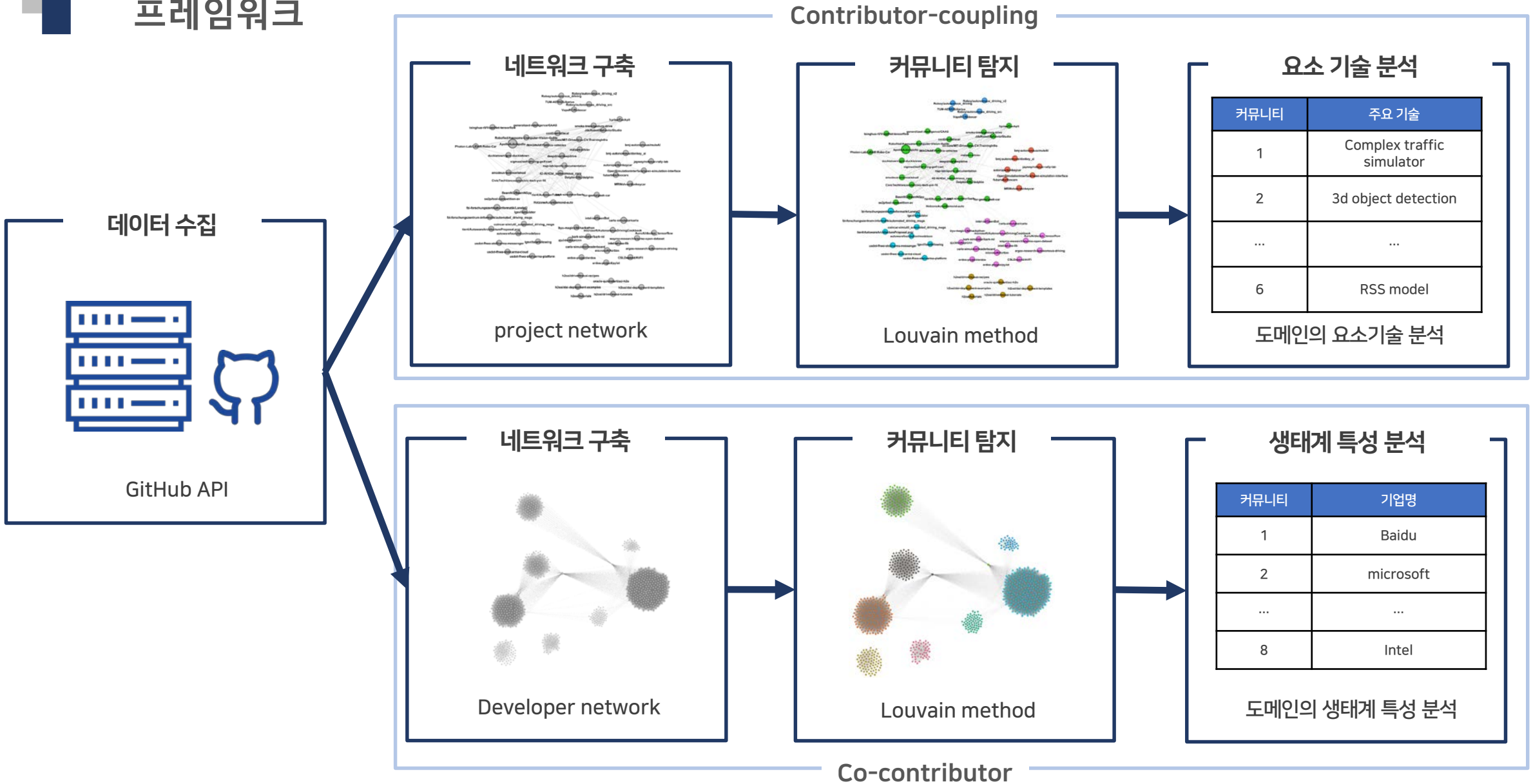
- GitHub는 개발자 간의 원활한 협업을 위해 pull request, fork, contribution과 같은 기능을 지원 (Bleiel, 2016)
- Readme에서 추출한 topic을 기반으로 repository를 기술, 연구분야별로 분류 (Sharma et al., 2017)
- LSTM을 이용하여, 오픈소스의 도메인 트렌드 변화 분석 (Varuna & Mohan, 2019)
- 지역별로 자주 사용되는 프로그래밍 언어 및 도메인 영역 분석 (Rusk & Coady, 2014)



오픈 소스 개발 데이터를 이용한 혁신 생태계 분석 연구는 미흡

3. 연구방법(1/3)

프레임워크



3. 연구 방법(2/3)

분석 데이터 수집

keyword	repository 수
Self-driving car	13,958
Self-driving vehicle	1,252
Autonomous car	2,973
Autonomous vehicle	2,658
Autonomous driving	2,176
Total	23,017

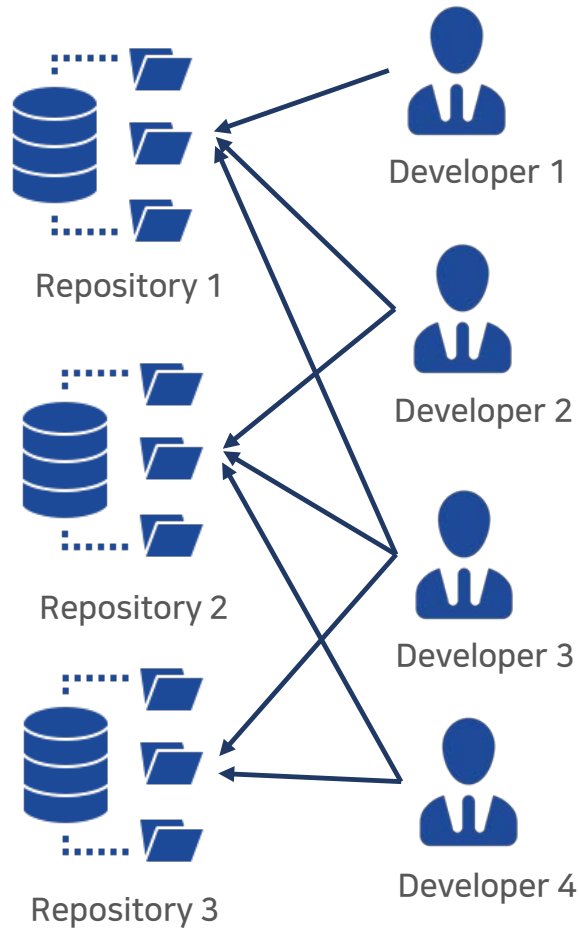
필터링 절차	repository 수
Raw data	23,017
Remove duplicate	21,180
User 소유의 repository 제거	737
Contributor가 두명 이하인 repository 제거	385
최종 데이터	385



- 데이터 수집 기간 : 2012.05.01 ~ 2021.03.10
- 자율주행 관련 여러 문헌을 참고하여 실제 용례를 기반으로 검색 키워드 선정 (Badue et al., 2021; Shladover, 2017; Schoettle & Sivak, 2014)

3. 연구 방법(3/3)

Co-contribution 기반 네트워크



Adjacency matrix (A)

	Repo1	Repo 2	Repo 3
Dev 1	1	0	0
Dev 2	1	1	0
Dev 3	1	1	1
Dev 4	0	1	1

- $a_{ij} = 1$: if developer i contributes to repository j

$A^T A$

$A A^T$

- Project network (contributor-coupling network)

	Repo1	Repo 2	Repo 3
Repo1	3	2	1
Repo 2	2	3	2
Repo 3	1	2	2

- Developer network (Co-contributor network)

	Dev1	Dev 2	Dev 3	Dev4
Dev 1	1	1	1	0
Dev 2	1	2	2	1
Dev 3	1	2	3	2
Dev 4	0	1	2	2

4. 자율주행 프로젝트 네트워크(2/2)

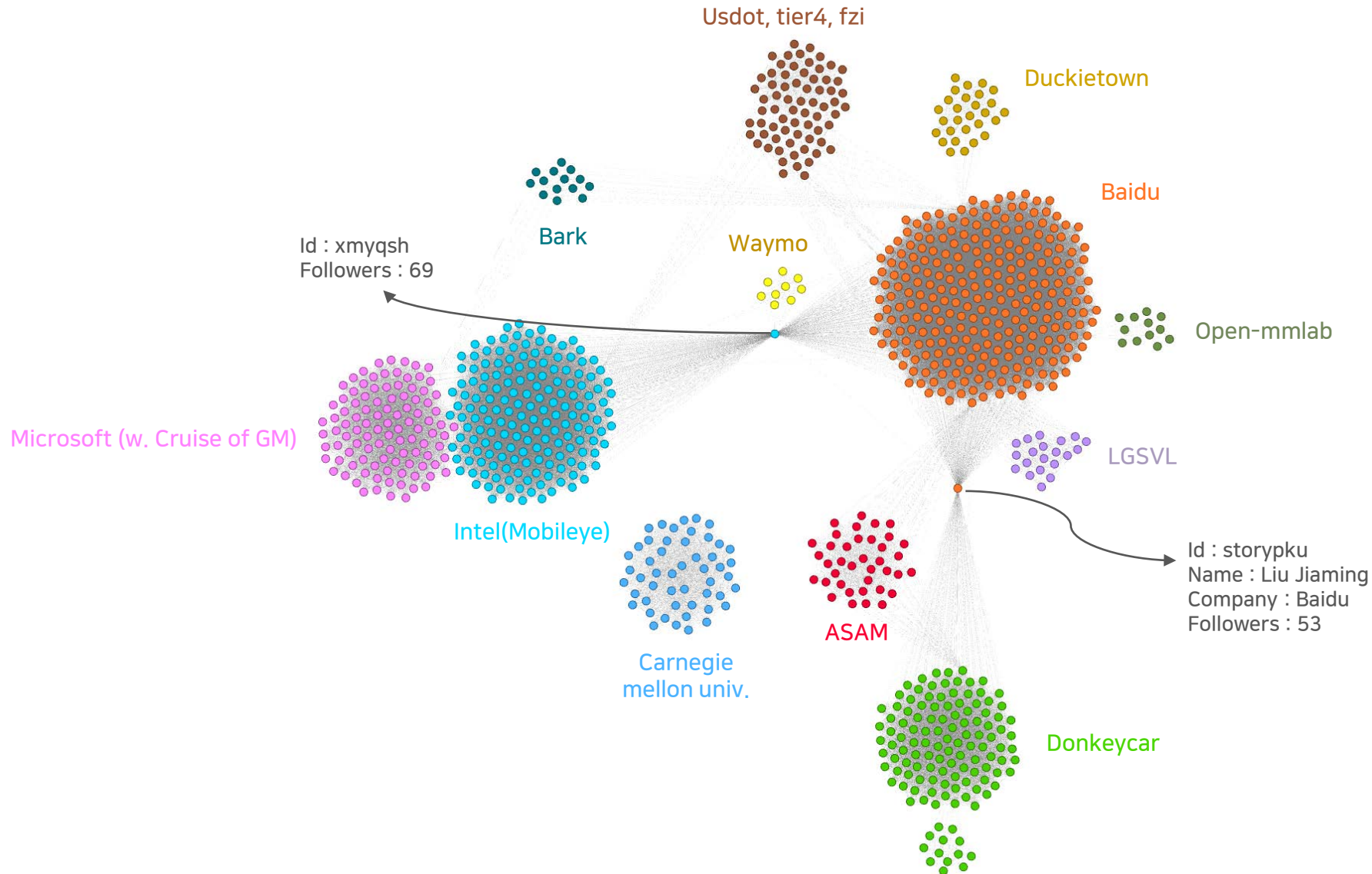
요약

Community 명	네트워크 내 비율(%)	Key player
Actuator control	37.31	JdeRobot, smoke trees, imaginary, hotzone
Virtual simulator	22.39	Intel, Microsoft
MOT(Moving object tracker)	17.92	Fzi, usdot, tier4, open-mmlab, LGSVL
Autonomous RC car controller	10.45	donkeycar
Motion planning	7.46	Roboy, vapeRC
Underwater self-driving car	4.48	srmauvsoftware

- 네트워크 내에서 시뮬레이션과 인지기술(이미지 처리, 센서 데이터 처리) 관련 커뮤니티가 높은 비중을 차지
 - 2021년 현재, 자율주행 오픈소스 분야는 시뮬레이션과 인지기술 위주로 발전
- 네트워크 내에서 Apollo, Carla, Airsim 3개의 시뮬레이터 노드가 높은 중앙성을 가짐
 - 실제 시뮬레이션 기술은 대부분 위 시뮬레이터들에 호환되도록 구현

5. 자율주행 개발자 네트워크(1/3)

Co-developer network



Node	908
Edge	64,368

색깔	회사명
Orange	Baidu
Pink	Microsoft
Cyan	Intel
Green	Donkeycar
Purple	LGSVL
Olive	Open-mmlab
Gold	Duckietown
Brown	Usdot, tier4, fzi
Red	ASAM
Light Blue	Carnegie mellon univ.
Yellow	Waymo
Dark Teal	Bark

5. 자율주행 개발자 네트워크(2/3)

Co-developer network

Community 명	네트워크 내 비율(%)	주요 기술
Baidu	29.74	Complex traffic simulator, DSL for self-driving car
Intel	18.83	Fault injection test, RSS model
Donkeycar	13.66	Autonomous RC car controller
Microsoft	11.34	3d object detection, smartphone-robot transfer
Usdot & Fzi & Tier4	7.60	Traffic awareness system using sensor and camera
Carnegie mellon university	5.51	Database for self-driving car
ASAM	4.19	simulator
Duckietown	3.08	Gym API, simulator
LGSVL	2.20	simulator
Bark	1.54	Decision making, monte carlo simulation
Open-mmlab	1.32	3d object detection, lidar
Waymo	0.99	Dataset for self-driving

5. 자율주행 개발자 네트워크(3/3)

요약

- 자율주행 developer network에 community detection을 진행한 결과, 기업 커뮤니티가 생성
- 자율주행 오픈소스 시뮬레이터로 유명한 Apollo의 Baidu, Carla의 Intel, Airsim의 Microsoft 가 네트워크 내에서 높은 비율을 차지
 - LGSVL, Dukietown, Bark등 시뮬레이터 관련 오픈소스를 보유한 기업들이 전체 네트워크의 71%
 - 현재 자율주행 오픈소스 개발의 트렌드는 실제 트랙에서 시험주행보다는 컴퓨터 시뮬레이션에 더 높은 비중
- MOT(Moving Object Tracking), 라즈베리 파이 기반 RC 카 시뮬레이터, 자율주행용 데이터베이스를 연구하는 기업들도 네트워크 내에 위치
 - MOT : Open-mmlab, Usdot, Fzi, Tier4
 - RC 카 시뮬레이터 : Donkey car
 - 자율주행 용 데이터베이스 : Carnegie mellon university
 - 도로 및 차량 데이터셋 : Waymo

6. 결론

- 요약 및 의의

- GitHub의 project 및 developer 네트워크를 이용하여 자율주행 오픈소스 분야의 혁신 생태계 파악
- Community detection을 통해 자율주행 분야의 key player와 요소 기술로 분류
- 관련 기업에게 전체 로드맵 및 향후 R&D 전략 수립 방안에 대한 청사진 제공
- 본 연구의 프레임워크를 활용하여 다른 종류의 산업군에서도 유의미한 분석결과를 제시할 수 있을 것으로 기대

- 한계점 및 향후 연구

- 테슬라, 애플등 오픈소스 공개에 폐쇄적인 스탠스를 취하고 있는 기업들은 본 네트워크에 미등장
- 향후 GitHub 내의 다른 social 지표인 issue 또는 pull requests 기반 네트워크를 통해 오픈소스 시장의 추가적인 특징 분석 가능
- 최근 활발히 연구되고있는 graph neural network 알고리즘을 이용한 graph embedding 및 link prediction을 이용해 기술 동향 분석을 넘어 forecasting으로 확장

7. 참고문헌

- Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., Jesus, L., Berriel, R., Paixão, T. M., Mutz, F., de Paula Veronese, L., Oliveira-Santos, T., & De Souza, A. F. (2021). Self-driving cars: A survey. In *Expert Systems with Applications* (Vol. 165). Elsevier Ltd. <https://doi.org/10.1016/j.eswa.2020.113816>
- Bleiel, N. (2016). Collaborating in GitHub. *IEEE International Professional Communication Conference, 2016-Novem*, 3–5. <https://doi.org/10.1109/IPCC.2016.7740497>
- Cai, Y., Ferrer, B. R., & Lastra, J. L. M. (2019). Building university-industry co-innovation networks in transnational innovation ecosystems: Towards a transdisciplinary approach of integrating social sciences and artificial intelligence. *Sustainability (Switzerland)*, 11(17). <https://doi.org/10.3390/su11174633>
- Faber, A., Rehm, S. V., Hernandez-Mendez, A., & Matthes, F. (2018). Modeling and visualizing smart city mobility business ecosystems: Insights from a case study. *Information (Switzerland)*, 9(11). <https://doi.org/10.3390/info9110270>
- Faber, A., Riemhofer, M., Hernandez-Mendez, A., & Matthes, F. (2018). Visualizing an Emerging Mobility Business Ecosystem. *Colloquium in Information Science and Technology, CIST, 2018-October*, 88–93. <https://doi.org/10.1109/CIST.2018.8596472>
- Granstrand, O., & Holgersson, M. (2020). Innovation ecosystems: A conceptual review and a new definition. *Technovation*, 90–91(November 2019). <https://doi.org/10.1016/j.technovation.2019.102098>
- Huhtamaki, J., & Rubens, N. (2016). Exploring innovation ecosystems as networks: Four european cases. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2016-March*, 4505–4514. <https://doi.org/10.1109/HICSS.2016.560>
- Kalliamvakou, E., Damian, D., Blincoe, K., Singer, L., & German, D. M. (2015). Open source-style collaborative development practices in commercial projects using GitHub. *Proceedings - International Conference on Software Engineering*, 1, 574–585. <https://doi.org/10.1109/ICSE.2015.74>
- Kim, K., Lee, W. R., & Altmann, J. (2015). SNA-based innovation trend analysis in software service networks. *Electronic Markets*, 25(1), 61–72. <https://doi.org/10.1007/s12525-014-0164-8>
- Mergel, I. (2015). Open collaboration in the public sector: The case of social coding on GitHub. *Government Information Quarterly*, 32(4), 464–472. <https://doi.org/10.1016/j.giq.2015.09.004>
- Munaiah, N., Kroh, S., Cabrey, C., & Nagappan, M. (2017). Curating GitHub for engineered software projects. *Empirical Software Engineering*, 22(6), 3219–3253. <https://doi.org/10.1007/s10664-017-9512-6>
- Rusk, D., & Coady, Y. (2014). Location-based analysis of developers and technologies on GitHub. *Proceedings - 2014 IEEE 28th International Conference on Advanced Information Networking and Applications Workshops, IEEE WAINA 2014*, 681–685. <https://doi.org/10.1109/WAINA.2014.110>
- Schoettle, B., & Sivak, M. (2014). A SURVEY OF PUBLIC OPINION ABOUT AUTONOMOUS AND SELF-DRIVING VEHICLES IN THE U.S., THE U.K., AND AUSTRALIA. <http://www.umich.edu/~umtriswt>
- Sharma, A., Thung, F., Kochhar, P. S., Sulistya, A., & Lo, D. (2017). Cataloging GitHub repositories. *ACM International Conference Proceeding Series, Part F1286*, 314–319. <https://doi.org/10.1145/3084226.3084287>
- Shladover, S. E. (2018). Connected and automated vehicle systems: Introduction and overview. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 22(3), 190–200. <https://doi.org/10.1080/15472450.2017.1336053>
- Varuna V, st T., & Anuraj Mohan, nd. (2019). Trend Prediction of GitHub using Time Series Analysis. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.
- Wang, N., & Hagedoorn, J. (2014). The lag structure of the relationship between patenting and internal R&D revisited. *Research Policy*, 43(8), 1275–1285.