

Machine Learning 기반 게임 봇 탐지

거래 네트워크와 유저 행동 패턴을 활용한 게임 작업장 탐지

NCSoft I&I실 분석모델링팀 서상덕

목차

I. 실 소개

II. 작업장이란?

III. 작업장의 특징

IV. 전체 작업 구성도

V. 거래 네트워크를 활용한 작업장 추정

VI. 행동 정보 요약

VII. 네트워크 시각화 및 집단 내 동질성 정량화

VIII. 성능 평가

I&I (Intelligence & Insight) 실은?

- 회사에 존재하는 다양한 데이터에 대한 분석을 통해 인사이트를 찾고 기계학습 및 통계 모델링을 이용해 고도화된 지표 및 시스템을 개발하는 조직



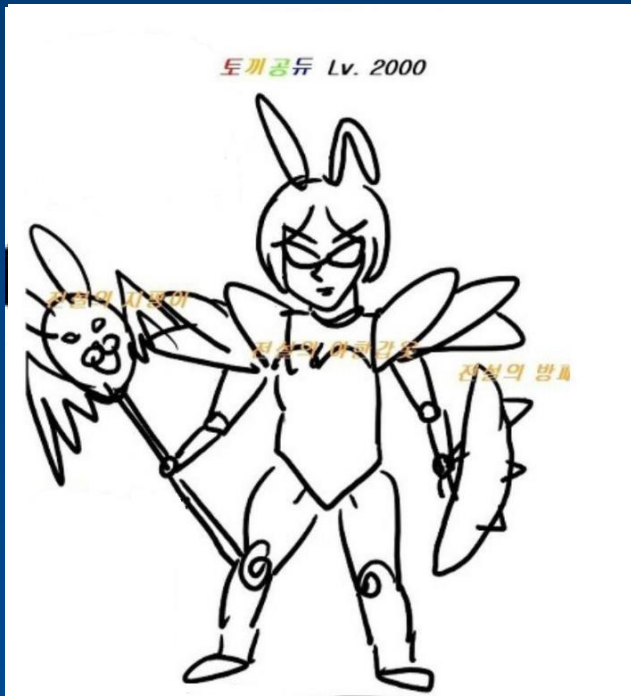
작업장이란?

- 다수의 캐릭터를 효율적으로 운영하여 아이템과 재화를 현금화하는 전문적인 집단



일반유저와 작업장의 차이?

일반 유저



작업장 캐릭터



작업장의 두 가지 특징을 탐지에 활용

1. 재화를 한곳으로 취합

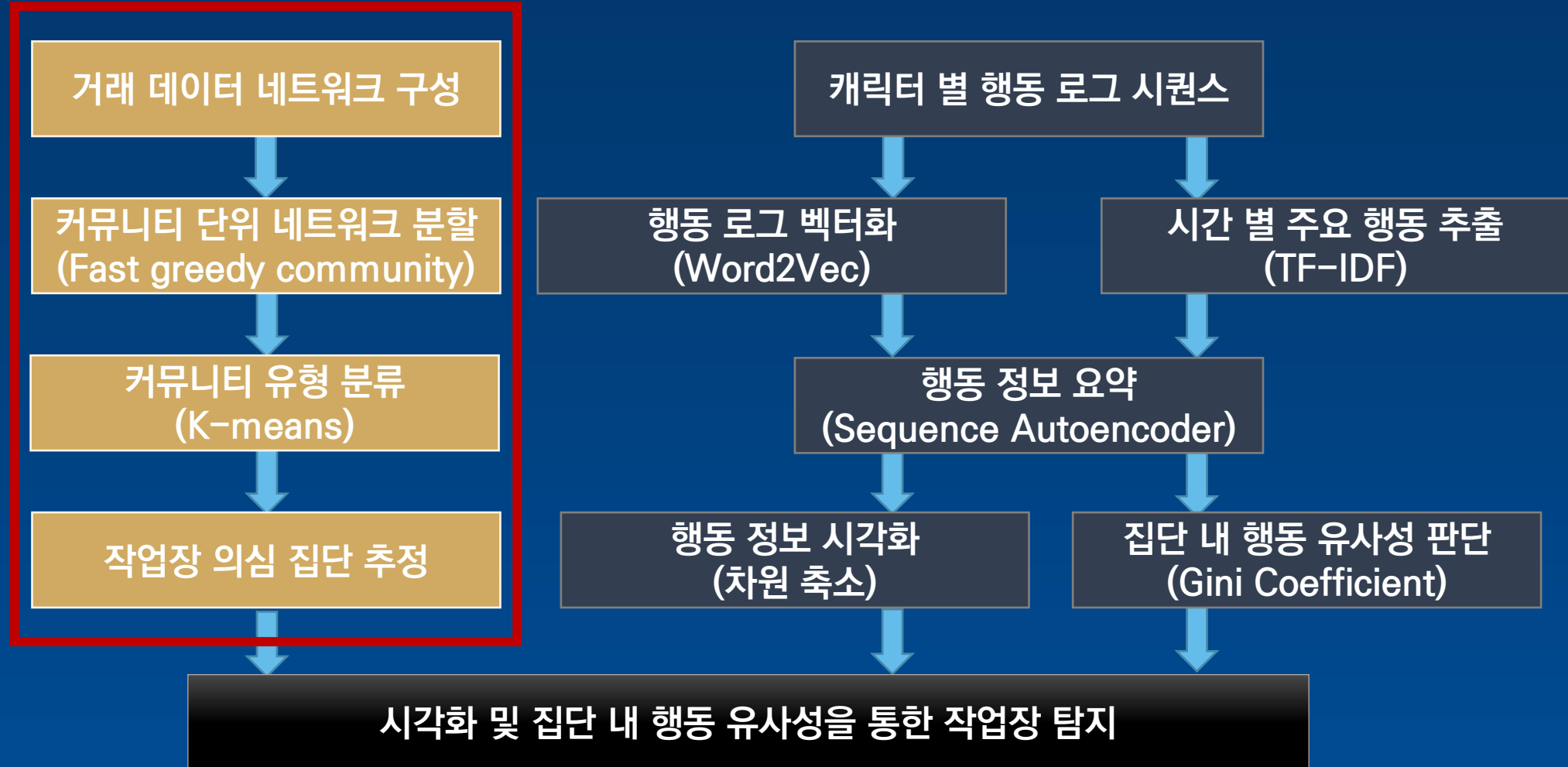
- 거래 네트워크를 활용하여 재화를 취합하는 커뮤니티 추출
 - 발표 논문
 - Crime Scene Reconstruction: Online Gold Farming Network Analysis – IEEE

2. 작업장 캐릭터의 유사한 행동 패턴

- 행동 정보를 요약하여 캐릭터의 행동 패턴 추출
 - 참고 논문
 - NGUARD: A Game Bot Detection Framework for NetEase MMORPGs – KDD

3. 작업장 집단 내 행동 유사성 확인

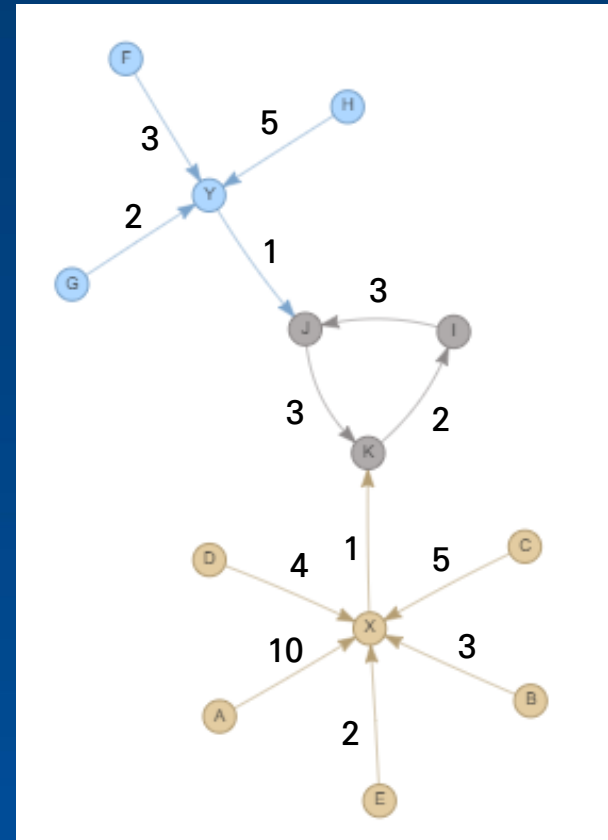
4. 전체 작업 구성도



거래 데이터 네트워크 구성

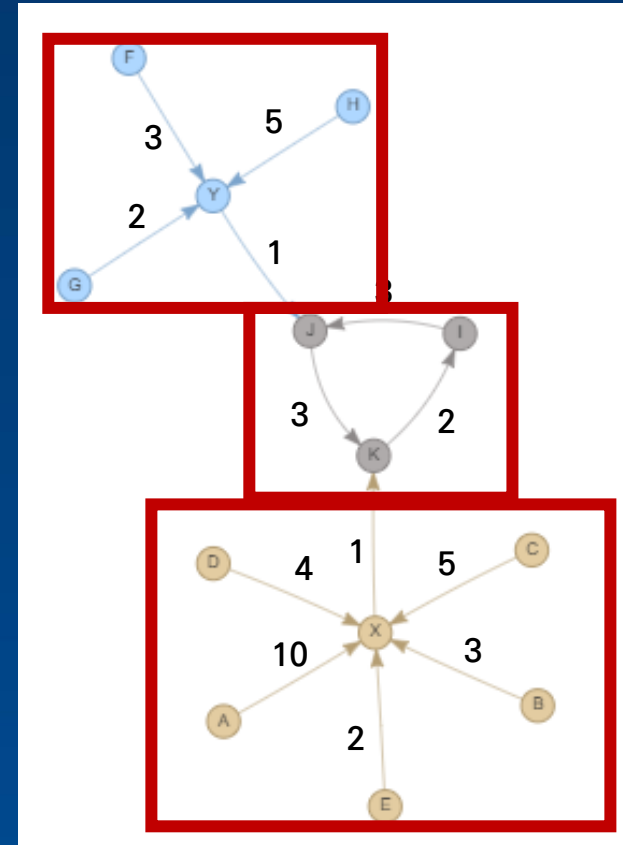
- 각 캐릭터마다 1주일간 수행한 거래 정보를 집계 후 네트워크 구성
- 자주 사용되는 아이템 및 재화에 대한 모든 거래 정보 활용

전달 캐릭터	받은 캐릭터	게임 재화
A	X	1000
A	X	2000
A	X	2500
...
E	X	4000
F	Y	2300
...
H	Y	2500
I	J	4800
...
K	I	1300
X	K	2000
Y	J	1000



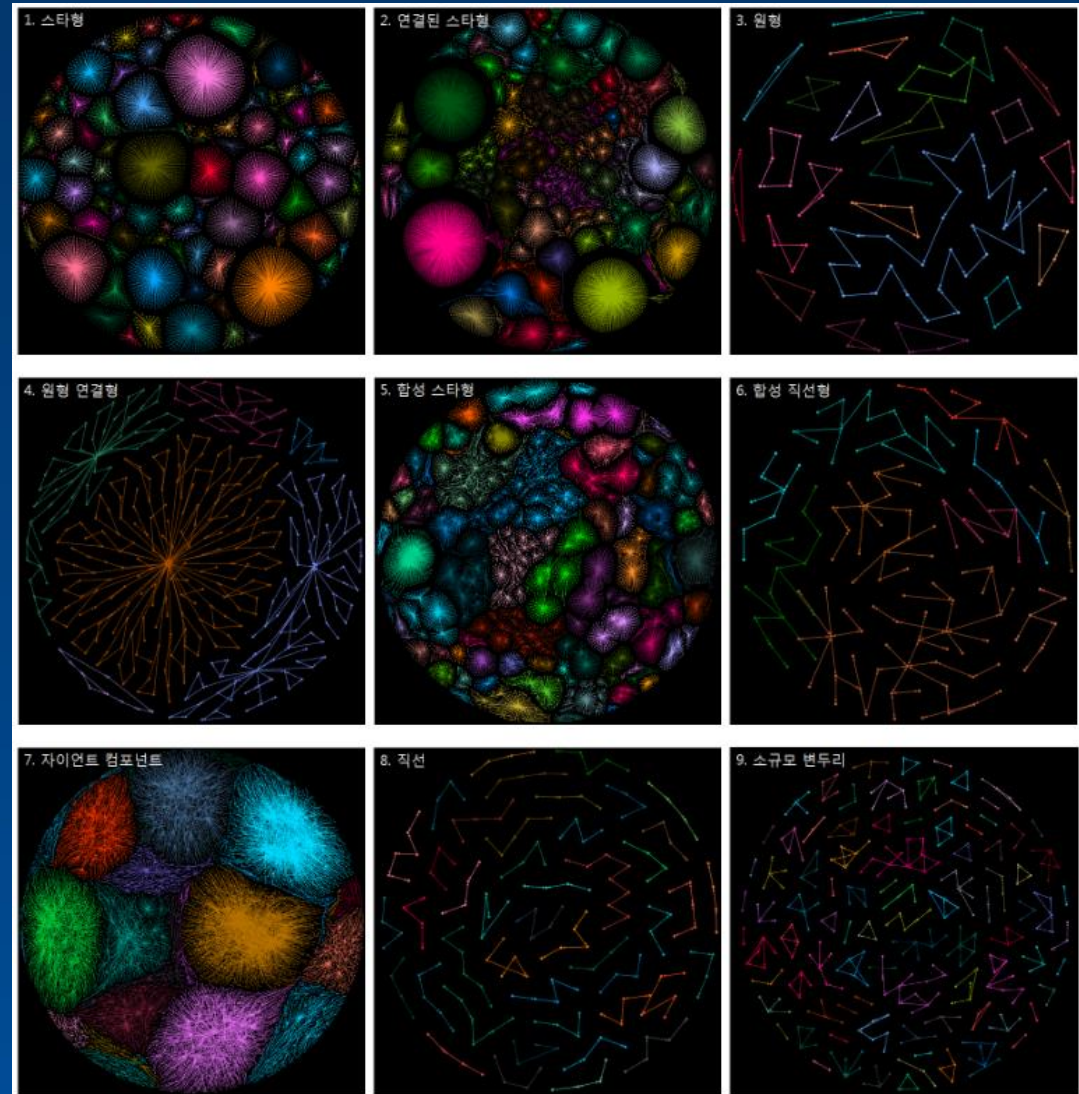
커뮤니티 단위 네트워크 분할

- Graph Clustering 알고리즘을 활용하여 분할
 - 분리된 커뮤니티가 최대한 밀집하게 연결되도록 그룹화
 - 캐릭터 간 거래 횟수를 가중치로 활용



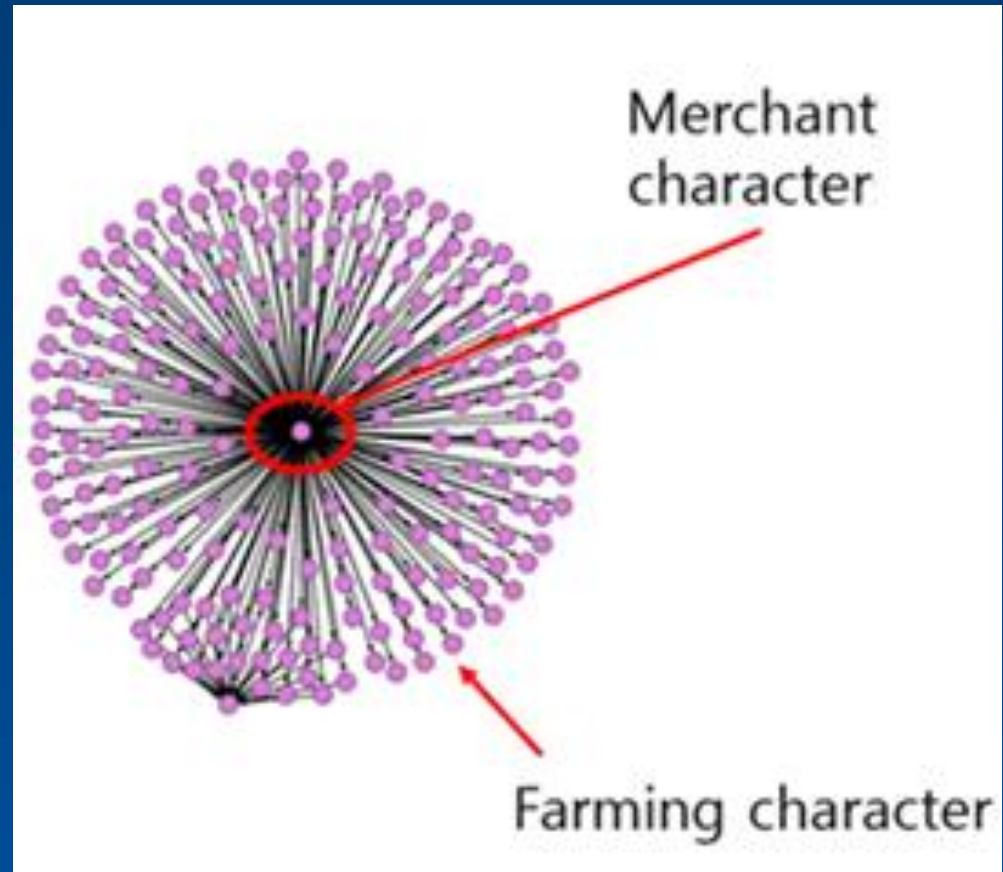
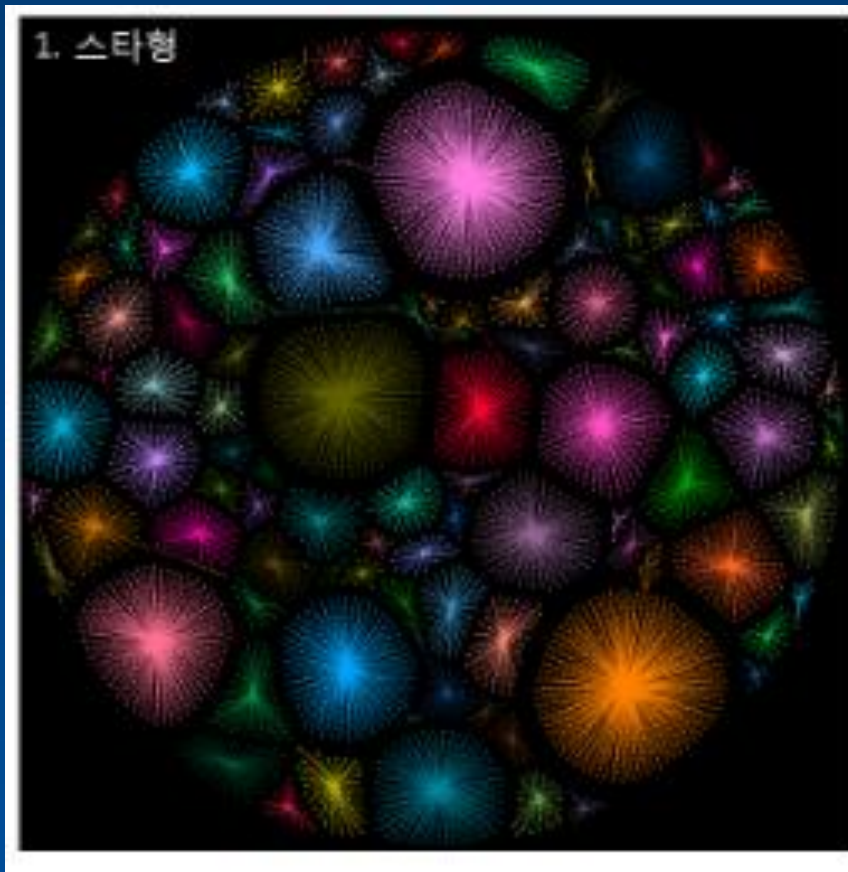
커뮤니티 유형 분류

- 분리된 커뮤니티의 구조와 관련된 특징 추출
 - 연결된 노드의 수
 - 그래프의 너비
 - 그래프의 노드 개수 등
- 추출된 특징을 활용하여 네트워크 유형 분류
 - K-means 클러스터링



작업장 유형 추정

- 재화를 한곳으로 모으는 유형을 작업장으로 추정

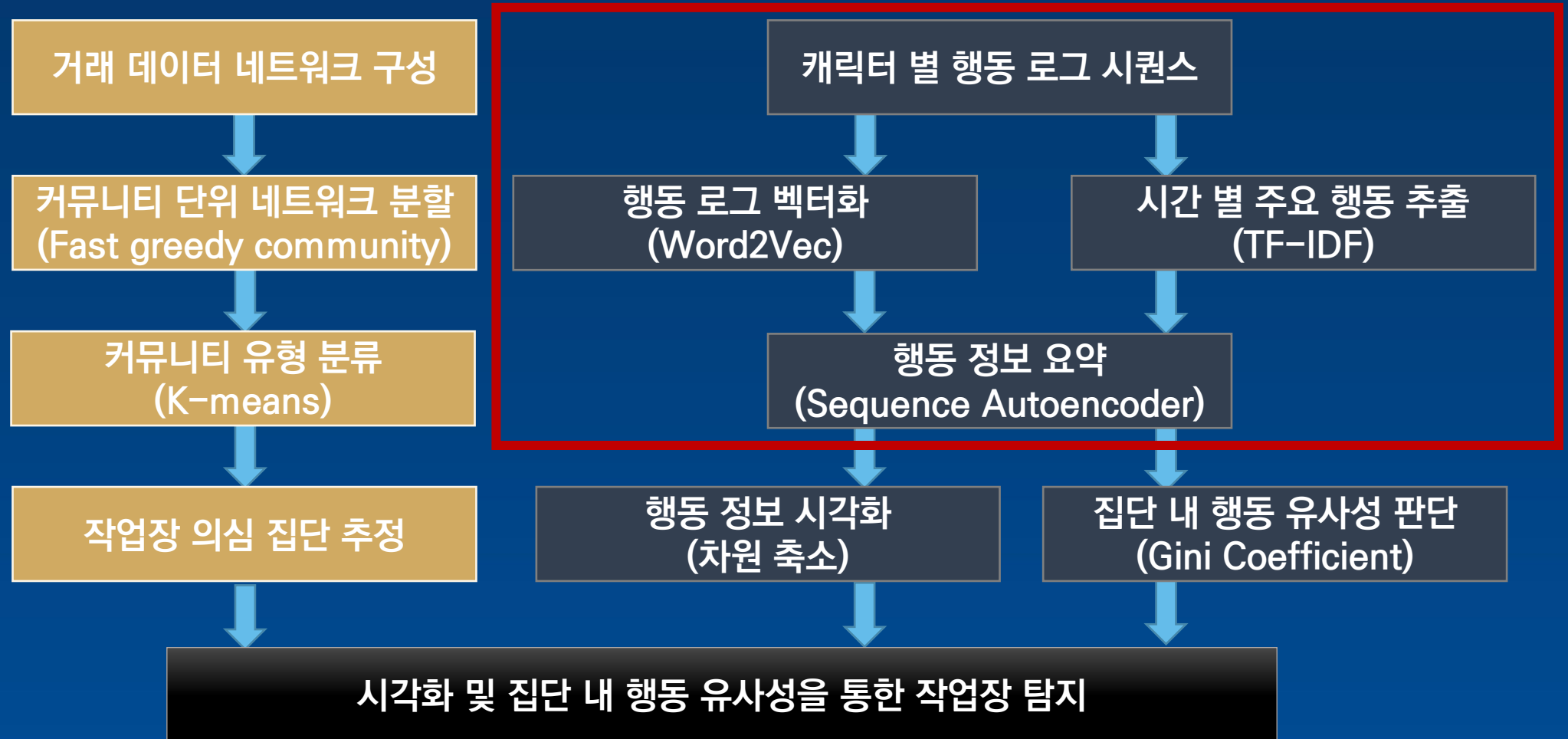


재화를 한곳으로 모은다는 이유만으로 작업장이라 볼 수 없다!

- 거래 정보에 캐릭터의 행동 정보를 추가로 확인하여 오탐을 방지



4. 전체 작업 구성도



Word2Vec을 활용한 행동 로그 벡터화 (1)

- CBOW(Continuous Bag of Words) 기법을 사용하여 행동 로그를 Vector로 변환

일요일은	내가	짜파게티	요리사	!
참깨빵	위에	순쇠고기	패티	두장

Word2Vec을 활용한 행동 로그 벡터화 (1)

- CBOW(Continuous Bag of Words) 기법을 사용하여 행동 로그를 Vector로 변환



Word2Vec을 활용한 행동 로그 벡터화 (2)

- 캐릭터의 행동 정보는 일반적인 문장과 다르게 시퀀스 사이에 시간차가 존재



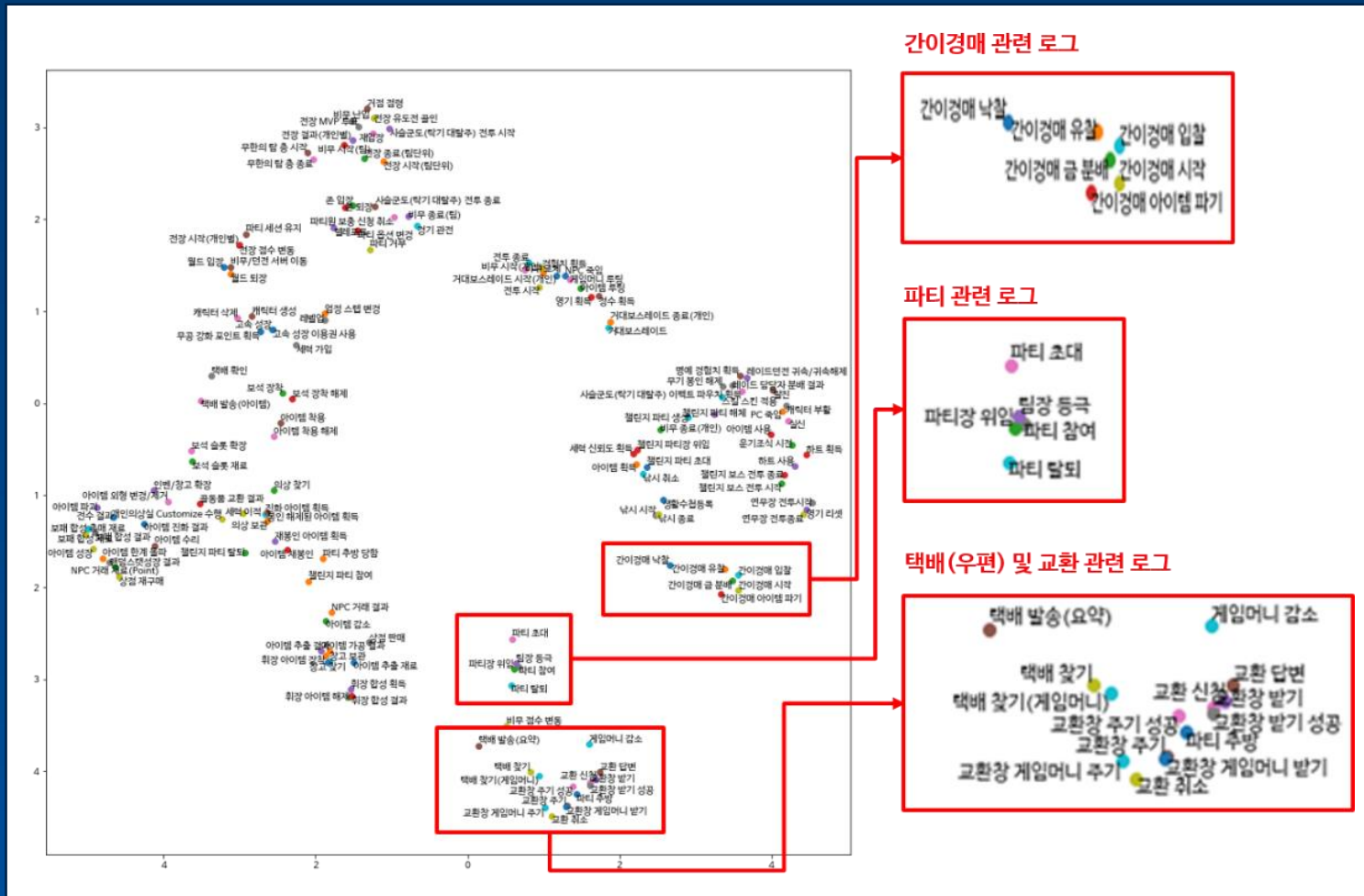
Word2Vec을 활용한 행동 로그 벡터화 (1)

- 행동 로그간의 시간차 가중치 부여
 - One-hot-encoding에서 1을 α^t 로 대체하여 시간 가중치 반영
 - α : 하이퍼 파라미터 ($0 < \alpha < 1$) t : 맥락 로그와 예측 로그 사이의 시간 간격



Word2Vec 학습 결과 시각화

- 행동 성향이 유사한 로그끼리 잘 군집 되어있는 것을 확인



TF-IDF를 활용한 시간별 주요 행동 추출 (1)

- MMORPG 특성 상 사냥 관련 로그는 대부분의 캐릭터에 존재
- 특징적인 행동이 학습에 반영되지 않을 수 있음

캐릭터 1	몬스터 죽임	아이템 획득	경험치 획득	...	물약 사용	우편 발송	재화 감소	거래	재화 감소
캐릭터 2	파티 초대	몬스터 죽임	아이템 획득	경험치 획득	...	물약 사용	몬스터 죽임	간이 경매	파티 탈퇴
캐릭터 3	몬스터 죽임	아이템 획득	경험치 획득	...	물약 사용	스킬 사용	스킬 사용	PC 죽임	캐릭터 사망

TF-IDF를 활용한 시간별 주요 행동 추출 (2)

- 단어의 빈도(TF)와 역 문서 빈도(IDF)를 활용하여 특정 단어의 중요도를 구하는 방법

- $\text{idf}(d, t) = \log\left(\frac{n}{df(t)}\right)$

- 1시간 단위로 캐릭터의 행동 수를 집계하여 TF-IDF 값이 높은 행동을 주요 행동으로 사용

캐릭터 1

행동	행동 수
몬스터 죽임	10
아이템 획득	4
경험치 획득	10
거래	2
물약 사용	5
재화 감소	2
우편발송	3

캐릭터 2

행동	행동 수
몬스터 죽임	8
아이템 획득	6
경험치 획득	8
파티 초대	1
물약 사용	8
간이경매	4
파티탈퇴	2

캐릭터 3

행동	행동 수
몬스터 죽임	13
아이템 획득	4
경험치 획득	13
PC 죽임	3
물약 사용	50
스킬 사용	10
캐릭터사망	1

n = 3

$$df(t) = 3$$

$$\text{idf}(d, t) = \log\left(\frac{3}{3}\right) = 0$$

$$df(t) = 1$$

$$\text{idf}(d, t) = \log\left(\frac{3}{1}\right) = 0.48$$

TF-IDF를 활용한 시간별 주요 행동 추출 (3)

- 1시간 단위로 캐릭터의 모든 행동을 추출한 뒤, TF-IDF값이 높은 행동으로 시퀀스를 재구성

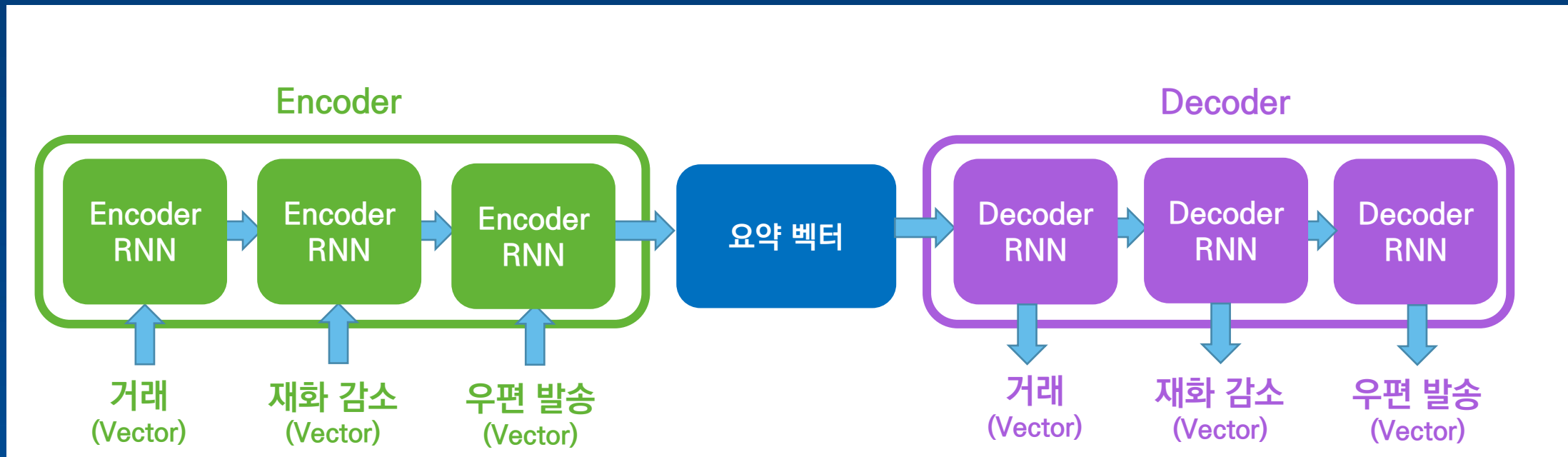
캐릭터 1	몬스터 죽임	아이템 획득	경험치 획득	...	물약 사용	우편 발송	재화 감소	거래	재화 감소
캐릭터 2	파티 초대	몬스터 죽임	아이템 획득	경험치 획득	...	물약 사용	몬스터 죽임	간이 경매	파티 탈퇴
캐릭터 3	몬스터 죽임	아이템 획득	경험치 획득	...	물약 사용	스킬 사용	스킬 사용	PC 죽임	캐릭터 사망



캐릭터 1	우편 발송	재화 감소	거래
캐릭터 2	파티 초대	간이 경매	파티 탈퇴
캐릭터 3	스킬 사용	PC 죽임	캐릭터 사망

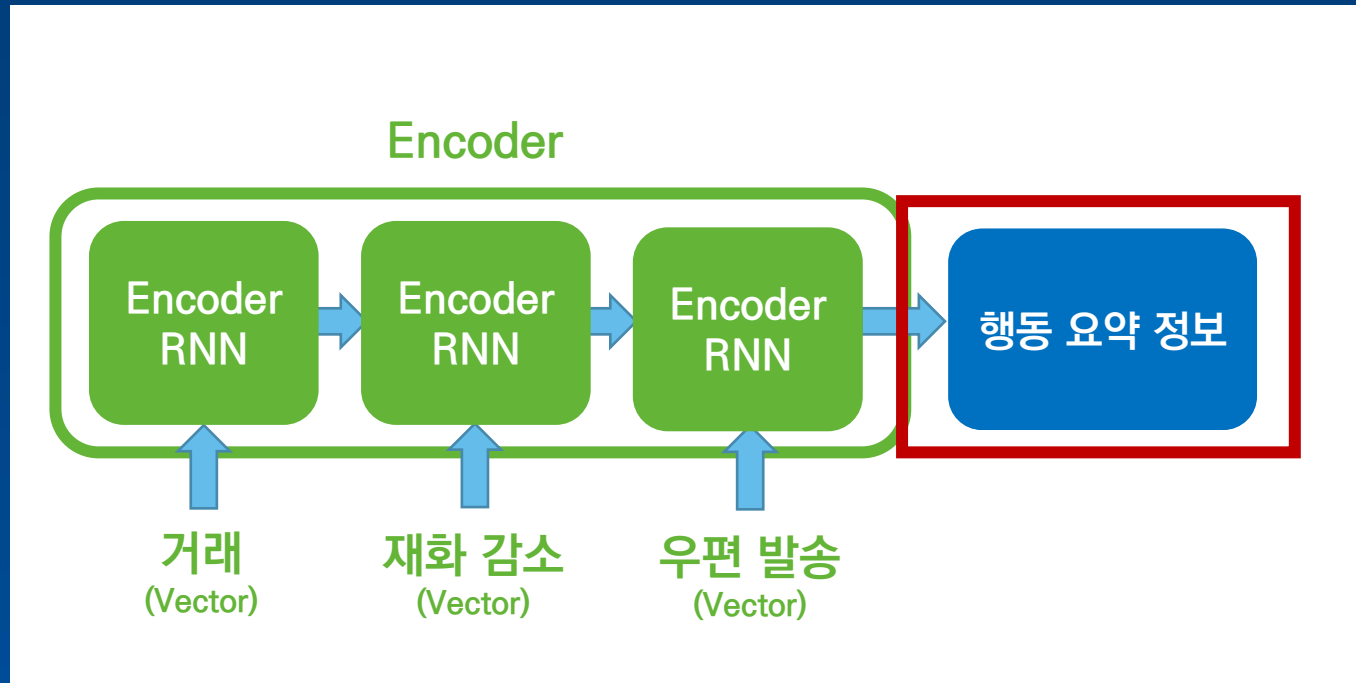
Sequence Autoencoder를 활용하여 행동 정보 요약

- 입력데이터와 출력데이터를 행동 시퀀스 데이터로 동일하게 사용하여 학습을 진행
 - 행동 로그를 Word2Vec으로 학습한 행동 벡터로 변환하여 학습에 반영
- 인코더에서 생성된 행동 요약 벡터(Context Vector)를 캐릭터의 행동 요약 정보로 활용

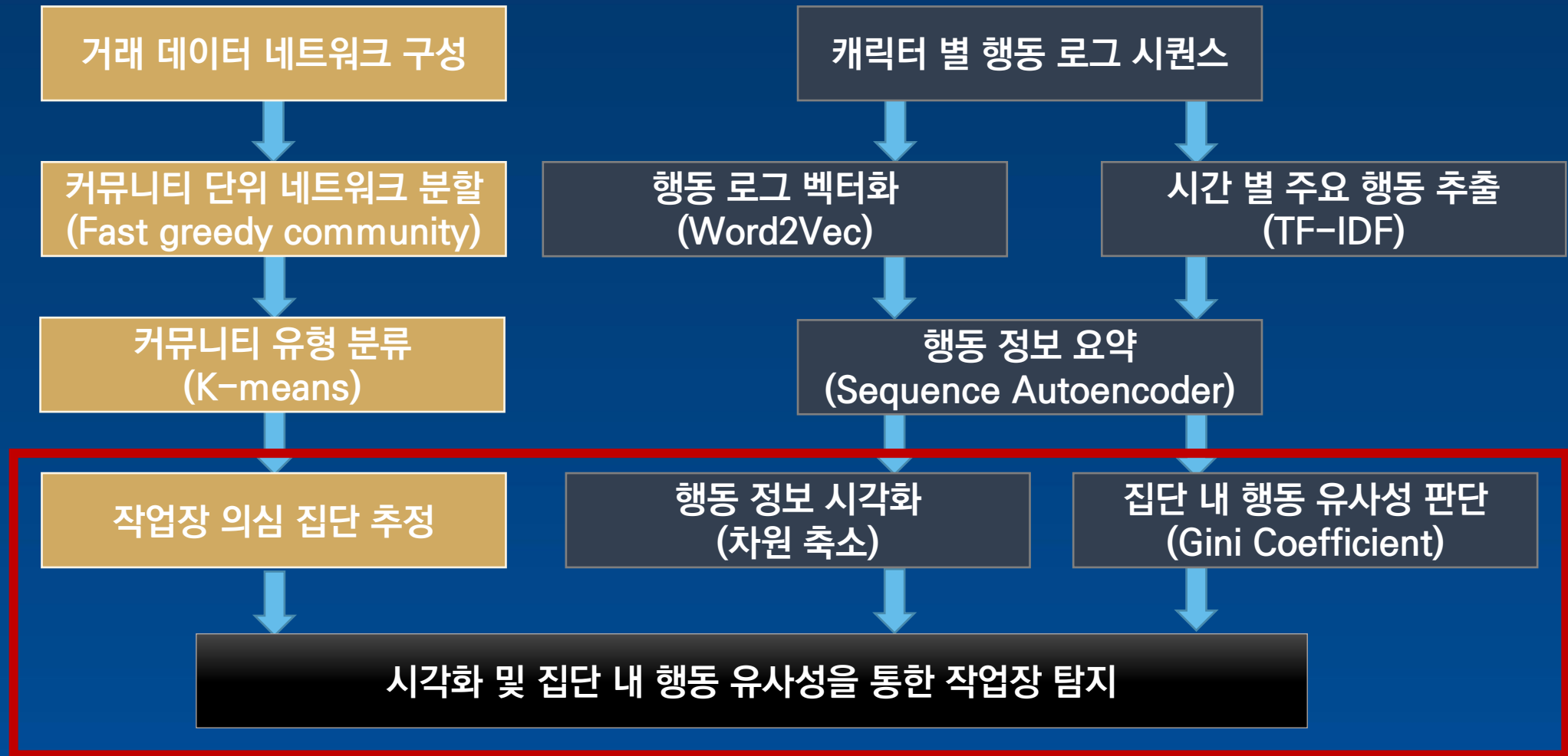


Sequence Autoencoder를 활용하여 행동 정보 요약

- 입력데이터와 출력데이터를 행동 시퀀스 데이터로 동일하게 사용하여 학습을 진행
 - 행동 로그를 Word2Vec으로 학습한 행동 벡터로 변환하여 학습에 반영
- 인코더에서 생성된 행동 요약 벡터(Context Vector)를 캐릭터의 행동 요약 정보로 활용



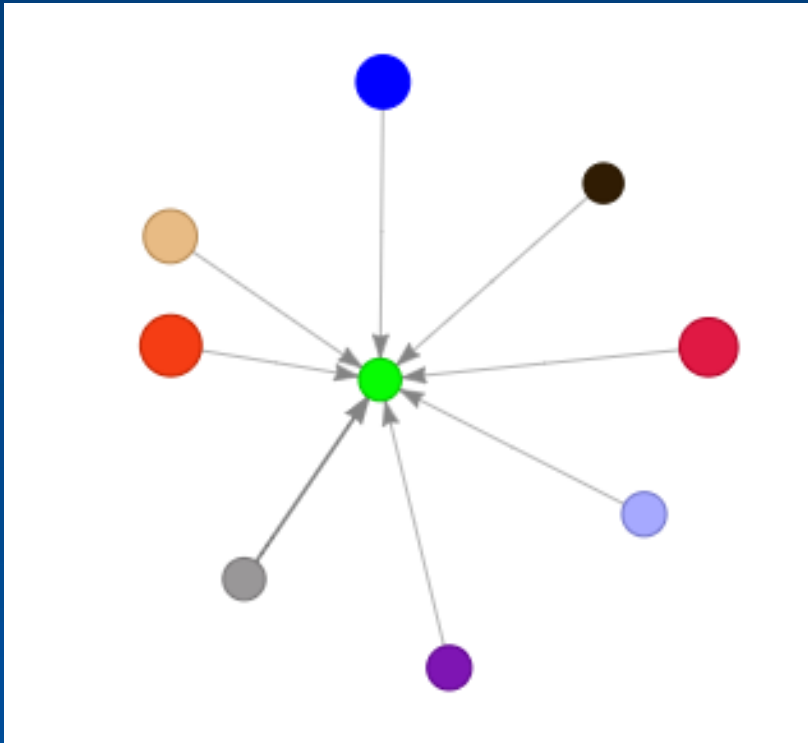
4. 전체 작업 구성도



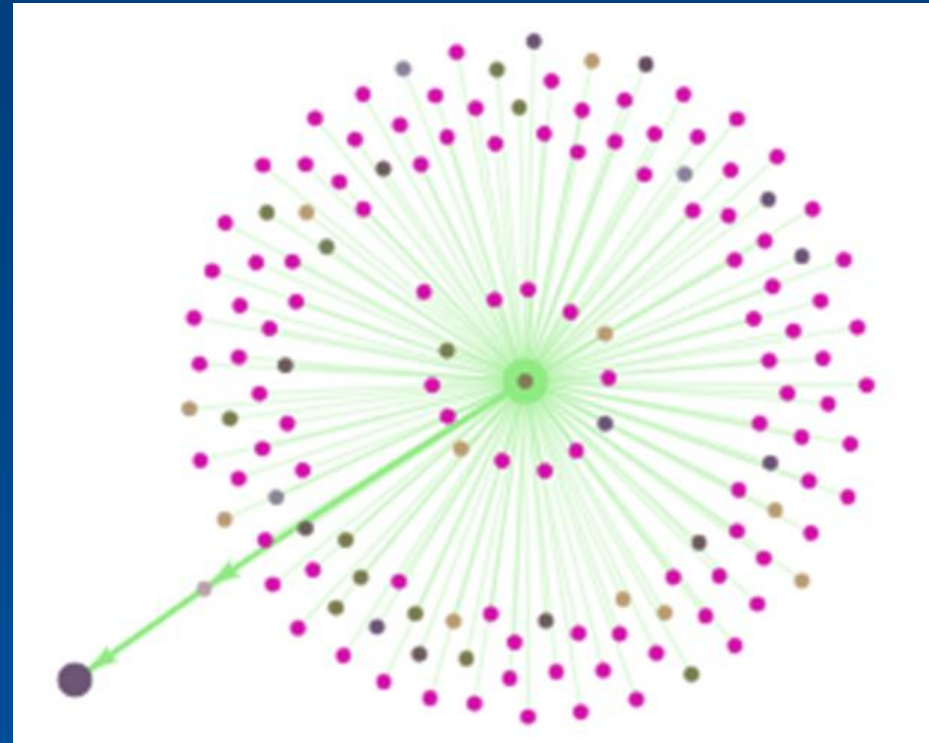
작업장으로 추정된 네트워크에 행동 정보를 활용한 시각화

- 행동 정보를 3차원으로 차원 축소하여 RGB 값으로 활용
- 일반 유저의 네트워크와는 다르게 중심 노드에 재화를 전달하는 주변 노드 색상이 대부분 유사한 것을 확인

일반 유저 거래 네트워크



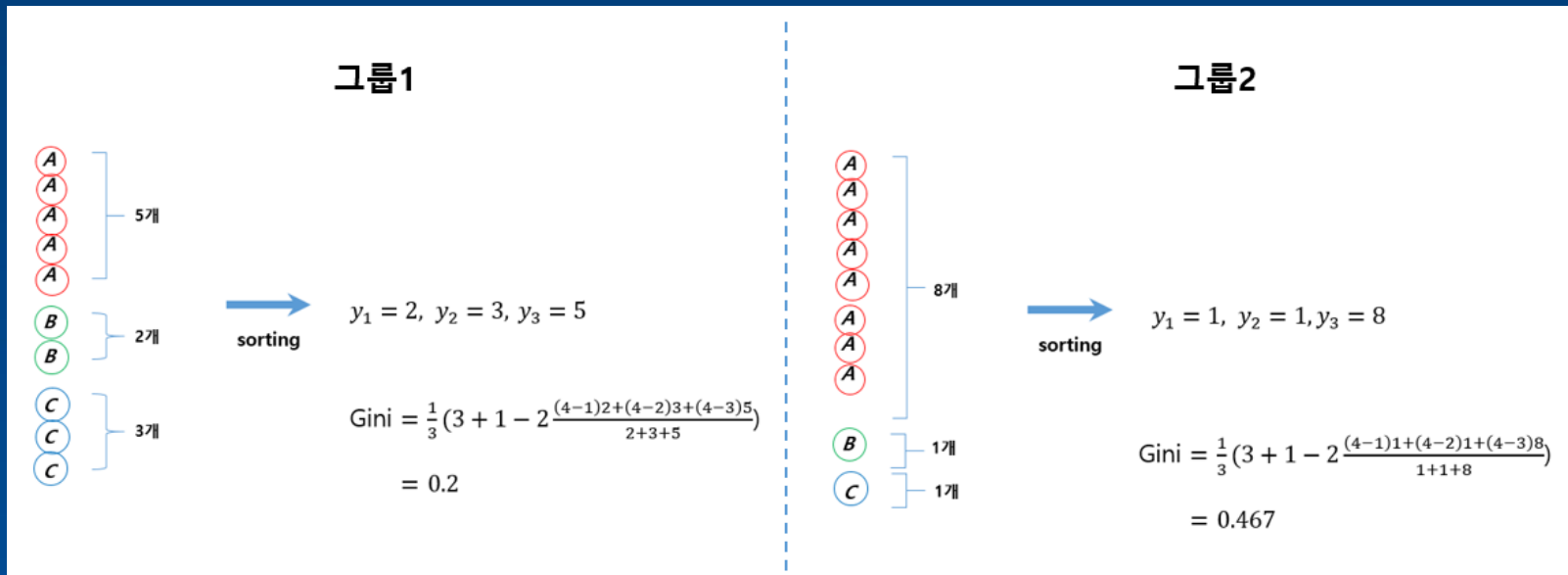
작업장 거래 네트워크



지니 계수를 활용한 집단 내 동질성 정량화

- 지니 계수 : 소득 또는 자산에 대한 불평등 정도를 나타내는 계수
- 집단 내 동일한 행동이 얼마나 불균형하게 분포되어있는지 측정

- $G = \frac{1}{n} \left(n + 1 - 2 \frac{\sum_{i=1}^n (n+1-i)y_i}{\sum_{i=1}^n y_i} \right) (y_i \leq y_{i+1})$ n : 집단 내 행동 유형 수 y_i : 같은 행동을 한 캐릭터 수.



정리

1. 재화를 한곳으로 취합

- 거래 네트워크 구성 후 네트워크 유형 분류
 - 거래 네트워크 분리 (Fast Greedy Community)
 - 네트워크 유형 분류 (K-means)

2. 작업장 캐릭터의 유사한 행동 패턴 추출

- 행동 의미 학습
 - 행동 로그 벡터화 (Word2Vec CBOW)
- 행동 정보 요약
 - 캐릭터 주요 행동 추출 (TF-IDF)
 - 행동 정보 요약 (Sequence Autoencoder)

3. 작업장 집단 내 행동 유사성 확인

- 작업장으로 추정된 커뮤니티의 행동 유사성 정량화
 - 작업장 내 행동 유사성 추출 (지니 계수)

작업장 검출에 대한 신뢰도 평가

- 사용 모델 : RandomForest 분류 모델
- 학습 데이터 : 124,837개
- 테스트 데이터 : 26,346개
- 특성
 - 거래 네트워크 유형
 - 행동 정보 지니 계수
 - 행동 요약 벡터 등

작업장 검출에 대한 신뢰도 평가

• 라벨 데이터

- 데이터를 수동으로 확인하여 작업장(B1), 작업장 의심(B2), 일반(N) 으로 분류
- B1 : 실제로 작업장으로 판단한 유저
- B2 : 작업장이라 의심이 되는 유저
- N : B1과 B2에 포함되지 않는 유저

		Predicted			
		B1	B2	N	Total
TRUE	B1	1,352	21	0	1,373
	B2	94	12,090	3,016	15,200
	N	0	1,123	8,650	9,773
	Total	1,446	13,234	11,666	26,346

END OF DOCUMENT