GNN Based Food-Drug Interaction Prediction

KUBIG
DATA SCIENCE & AI

Team | 논알콜제로디카페인소주
고윤경, 권민석, 원준혁, 이건영

KUBIG
DATA SCIENCE & AI

CONTENTS

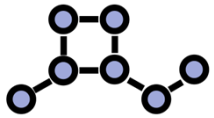
KUBIG
DATA SCIENCE & AI

01. Introduction

문제 정의 및 프로젝트 목적 설정

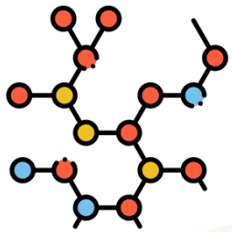
기존 Drug-Drug Interaction 예측 모델

약물 분자 구조 활용



Food-Drug Interaction 예측 모델

음식 화합물 구조 활용



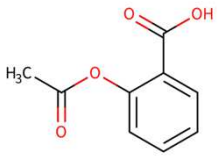
KUBIG
DATA SCIENCE & AI

02. Data

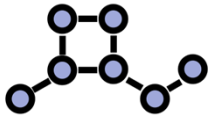
분자 구조 데이터

1) 약물 분자 구조 to 그래프

Aspirin:
CC(=O)OC1=CC=CC=C1C(O)=O



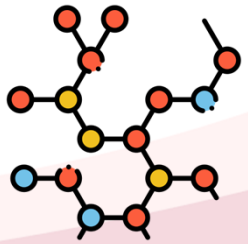
node: 원자, edge: 결합 종류



2) 음식 화합물 구조 to 그래프

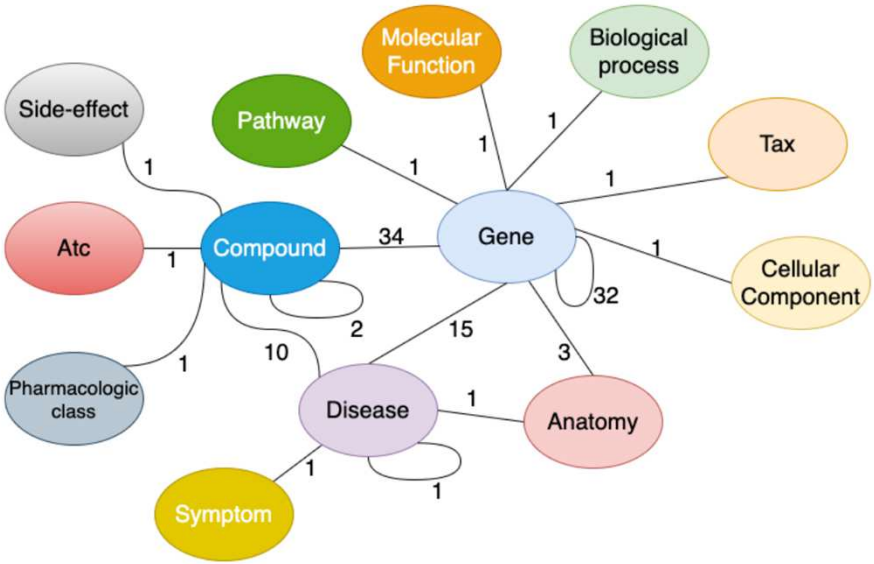
Food_Name	orig_content	name
Kiwi	1450	Pectic acid
Kiwi	1146	Potassium
Kiwi	950	Chinese tannin
Kiwi	356	L-Ascorbic acid
Kiwi	168	Phosphorus
Kiwi	103.5	Calcium
Kiwi	17.25	Sodium

node: 분자, edge: 함유량 비율



지식 그래프

약물 지식 그래프 DRKG(13개 entity type, 107개 edge type)**에 'FOOD entity' 추가**



FOOD
(connected by. compound SMILES)

상호작용 정보 라벨링

1) Drug-Drug Interaction

DrugBank : 전문가가 수동으로 정리한 86가지의 문장 기반 상호작용 유형

TWOSIDES : 부작용 데이터 signal 기반 200여가지의 특정 부작용 유형

2) Food-Drug Interaction

<u>food/drug type</u>

- SapBERT from PudMedBERT 이용해 기존 데이터셋에 있던 것들과 매핑/필터링
- threshold) embedding vector cosine similarity 0.6

<u>interaction type</u>

- S-PubMedBert-MS-MARCO 이용해 DrugBank 기준 interaction 라벨들로 매핑
- threshold) embedding vector cosine similarity 0.9

상호작용 정보 라벨링 - fdi

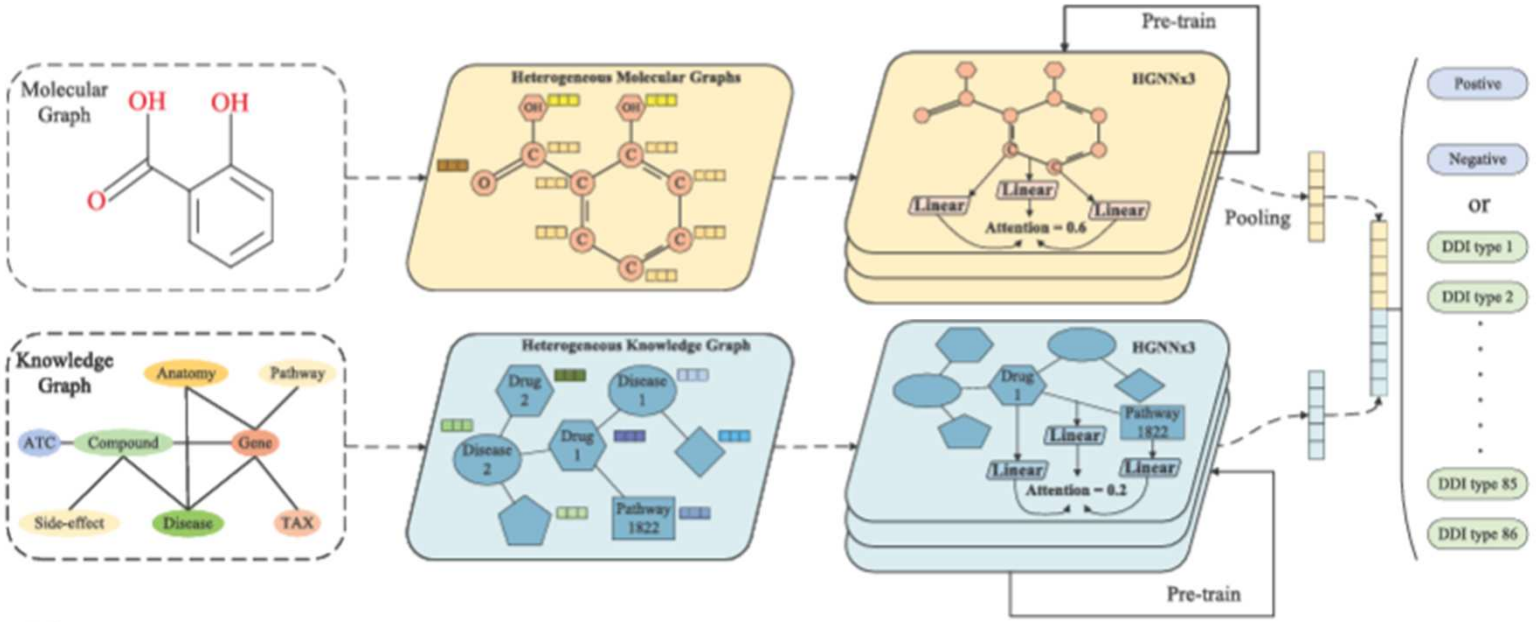
Origin_Food	Origin_Drug	Origin_Interaction_Text	New_Food	New_Drug	Food_Similarity	Drug_Similarity	New_Interaction_Text	Ineraction_Similarity
Apple	alendronate	the potential adverse interactions included decrease drug bioavailability	Apple	Alendronic Acid	1	0.91	#Drug1 can cause an increase in the absorption of #Drug2 resulting in an increased serum concentration and potentially a worsening of adverse effects.	0.9
beef	3-o-methyl-glucose	at t = 0 min, subjects ingested 100 g beef mince labeled with 20 mbq technetium-99m-sulfur colloid and 3 g 3-o-methyl-glucose (3-omg), a marker of glucose absorption	Beef	methyl beta-D-glucopyranoside	1	0.79	The serum concentration of the active metabolites of #Drug2 can be increased when #Drug2 is used in combination with #Drug1.	0.88

KUBIG
DATA SCIENCE & AI

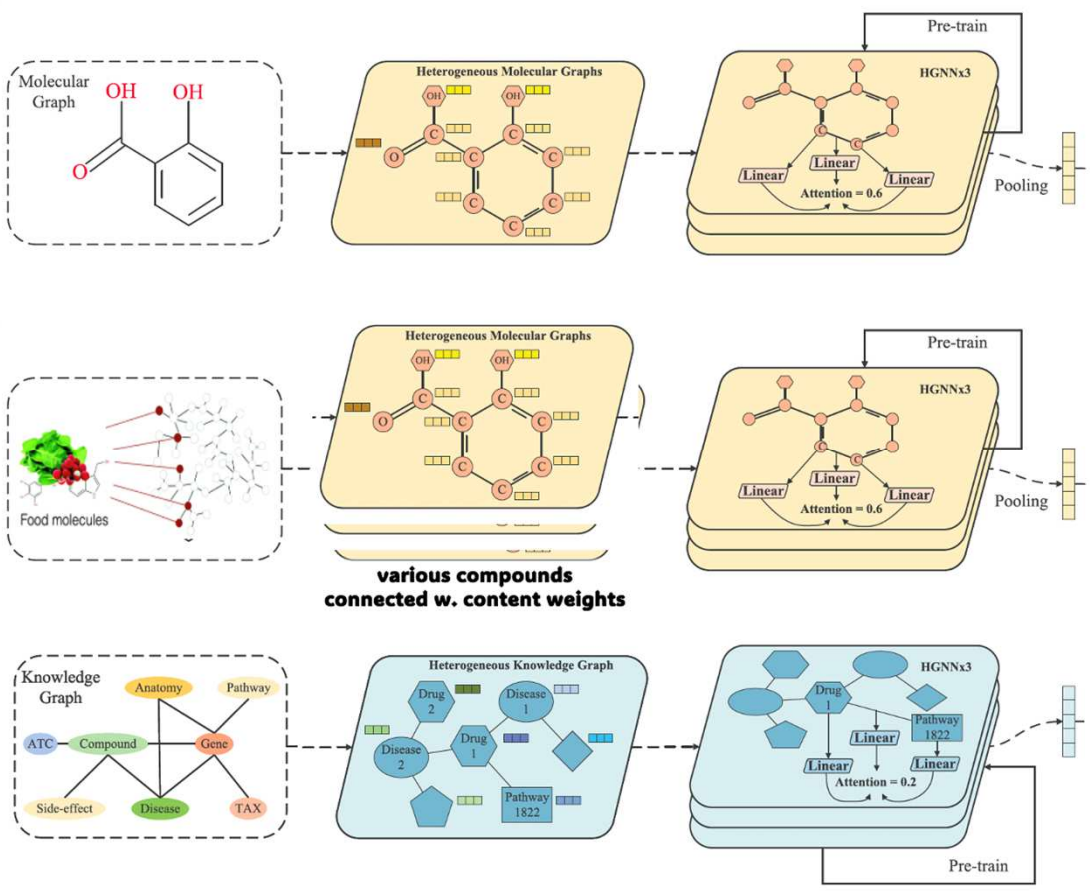
03. Model

HetDDI

HetDDI: a pre-trained heterogeneous graph neural network model for drug-drug interaction prediction



모델 구조 설명



1) drug compound info.

SMILES to molecular graph

by. RDKit library
"mol = Chem.MolFromSmiles(smiles)"

2) food compound info.

FooDB

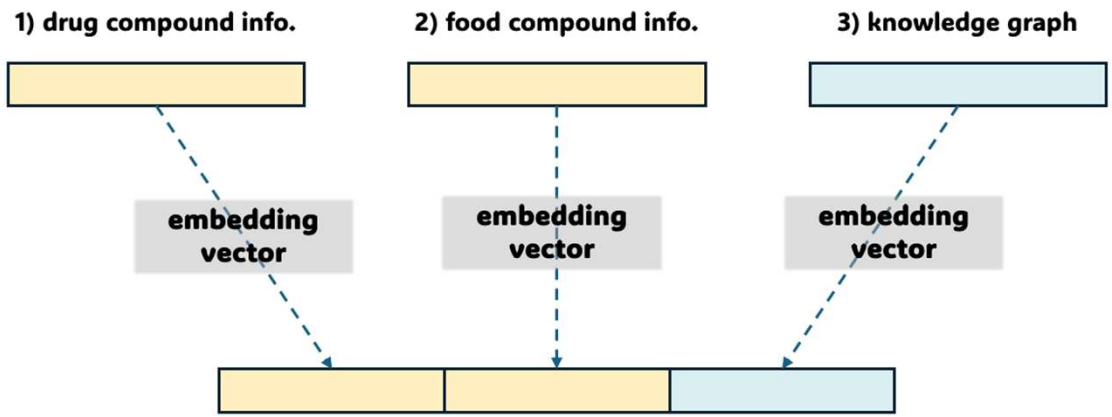
- **nodes:**
  1) food type,
  2) compound SMILES
- **edges:**
  content weight (mg/100g)

3) knowledge graph

(DRKG) + food nodes

- pretrained knowledge graph from HetDDI
- selected new food nodes to be added, which have SMILES/content info. in FooDB

모델 구조 설명



interaction text/label from DrugBank

S-PubMedBert-MS-MARCO (NLP model)
food-drug interaction info. from FooDrugs → match to labels used in DrugBank

#Drug1 can cause an increase in the absorption...

The risk or severity of hyperkalemia can be in...

. . . . . .

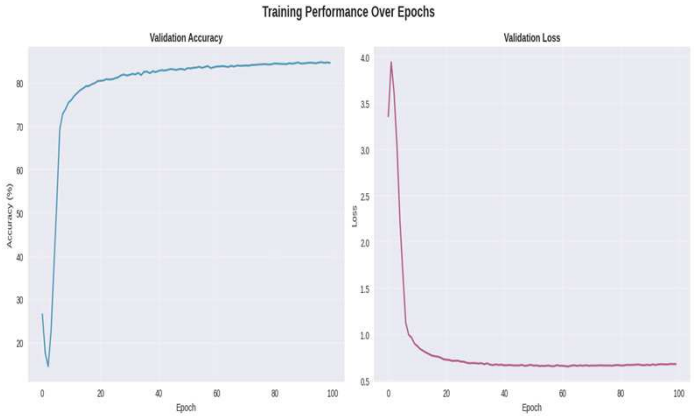
The serum concentration of the active metaboli...

KUBIG
DATA SCIENCE & AI

04. Result

실험 결과

Train/Validation Loss



Fold별 실험 결과

```
=== Training Summary (Fold 0) ===
           Metric  Initial  Final  Best  Improvement
Validation Accuracy    26.54  84.59 84.73        58.04
    Validation Loss     3.36   0.68  0.65         2.68
       Validation F1     2.23  62.10 62.27        59.87
   Validation Kappa    11.09  82.63 82.78        71.54

Total epochs: 100
Best score updates: 29
Final validation accuracy: 84.59%
Best validation accuracy: 84.73%
```

```
=== Training Summary (Fold 1) ===
           Metric  Initial  Final  Best  Improvement
Validation Accuracy    11.81  84.94 85.06        73.14
    Validation Loss     3.42   0.67  0.65         2.75
       Validation F1     0.54  61.94 62.44        61.39
   Validation Kappa     2.16  83.02 83.14        80.86

Total epochs: 100
Best score updates: 29
Final validation accuracy: 84.94%
Best validation accuracy: 85.06%
```

```
=== Training Summary (Fold 2) ===
           Metric  Initial  Final  Best  Improvement
Validation Accuracy    25.59  84.86 84.86        59.27
    Validation Loss     3.37   0.68  0.66         2.69
       Validation F1     0.57  62.78 62.78        62.21
   Validation Kappa     0.86  82.93 82.93        82.08

Total epochs: 100
Best score updates: 29
Final validation accuracy: 84.86%
Best validation accuracy: 84.86%
```

Test Set Metric

	loss	acc	f1	pre	recall	Kappa / auc
FDI	0.66	83.7	60.9	57.7	70.4	81.7
DDI	0.08	97.3	97.3	96.8	97.8	

최종 결과 및 마무리

최종 예측 결과 예시

Food	Drug	Interaction Label
Port (포트 와인)	Calcium Chloride (염화칼슘, 정맥 주사 등에 사용)	Food may increase the hypocalcemic activities of Drug (약물과 음식을 함께 복용하면 칼슘 농도가 낮아진다)
Protein supplement (프로틴)	Ibrutinib (항암 표적 치료제)	The serum concentration of the active metabolites of Food can be reduced when Food is used in combination with Drug resulting in a loss in efficacy. (혈중 약물 농도가 감소하여 효과가 줄어들 수 있다)

의의 및 한계

- DDI 프로세스를 복잡한 음식 구조로 확장
- 환자 데이터를 활용한 개인화된 음식, 약물 상호작용 예측 및 진단 가능
- 상호작용 label 추정으로 인한 정확도 감소
- 사용한 음식 범위의 제한

KUBIG
DATA SCIENCE & AI

Thank You