

数据

大致完成文本数据收集、对齐

进展

dataset1

|    | A       | B             | C   | D  | E    | F | G | H | I | J | K | L | M | N | O | P | Q |
|----|---------|---------------|---|----|------|---|---|---|---|---|---|---|---|---|---|---|---|
| 1  | id      | name          | text  | 备注 | 文本数据 |   |   |   |   |   |   |   |   |   |   |   |   |
| 2  | DB01296 | Glucosam      | Glucosamine is a common ingredient in nutritional supplements used for the relief of joint pain.  |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 3  | DB09230 | Azelinidipi   | Azelinidipine is a dihydropyridine calcium channel blocker. It is marketed by Daiichi-Sankyo pharmaceuticals, Inc. in Japan. It has a gradual onset of action and p   |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 4  | DB05812 | Abiraterol    | Abiraterone is an antiandrogen used in the treatment of metastatic castration-resistant prostate cancer and metastatic high-risk castration-sensitive prostate c      |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 5  | DB01195 | Flecainide    | Flecainide is a class Ic antiarrhythmic agent used to manage atrial fibrillation and paroxysmal supraventricular tachycardias (PSVT).                                 |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 6  | DB00201 | Caffeine      | Caffeine is a stimulant present in tea, coffee, cola beverages, analgesic drugs, and agents used to increase alertness. It is also used in to prevent and treat pulmo |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 7  | DB01020 | Isosorbide    | Isosorbide mononitrate is a nitrate used to prevent and treat angina and to treat angina caused by coronary artery disease.   |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 8  | DB00278 | Argatroba     | Argatroban is a synthetic direct thrombin inhibitor used for the prevention and treatment of thrombosis related to heparin use.                                       |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 9  | DB01403 | Methotrin     | Methotrimoprazine is a phenothiazine used in the management of psychosis, particular those of schizophrenia, and manic phases of bipolar disorder.                    |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 10 | DB00588 | Fluticason    | Fluticasone propionate is a glucocorticoid used to treat asthma, inflammatory pruritic dermatoses, and nonallergic rhinitis.  |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 11 | DB00091 | Cyclospor     | Cyclosporine is a steroid-sparing immunosuppressant used in organ and bone marrow transplants as well as inflammatory conditions such as ulcerative colitis,          |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 12 | DB00915 | Amantadi      | Amantadine is a medication used to treat dyskinesia in Parkinson's patients receiving levodopa, as well as extrapyramidal side effects of medications.                |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 13 | DB00461 | Nabumet       | Nabumetone is an NSAID used to treat osteoarthritis and rheumatoid arthritis.   |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 14 | DB01223 | Aminophy      | Aminophylline is a bronchodilator consisting of theophylline that is used for the treatment of bronchospasm due to asthma, emphysema and chronic bronchitis           |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 15 | DB00648 | Mitotane      | Mitotane is an adrenal cortex inhibitor used to treat adrenocortical tumors and Cushing's syndrome.   |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 16 | DB00783 | Estradiol     | Estradiol is an estrogenic steroid used to treat vasomotor symptoms of vulvar and vaginal atrophy in menopause, hypoestrogenism, prevention of postmenopau            |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 17 | DB00894 | Glafagab      | Glafagabine is a GABA agonist used to treat relapsed or refractory acute lymphoblastic leukemia in patients 1 to 21 years old.  |    |      |   |   |   |   |   |   |   |   |   |   |   |   |
| 18 |         | dataset1_text |   |    |      |   |   |   |   |   |   |   |   |   |   |   |   |

文本数据来自于 drugbank 数据库中对每个药物的 Summary 和 Background

DRUGBANK

ExploreFor Drug DiscoveryFor Clinical SoftwareFor Academic Research

CHAT WITH AN EXPERT

Search section headers

IdentificationPharmacologyInteractionsProductsCategoriesChemical IdentifiersReferencesClinical TrialsPharmacoeconomicsPropertiesSpectraTargets (20)Enzymes (4)

Clinical Trials

Phase 00Phase 10Phase 20Phase 31Phase 43

IDENTIFICATION

Summary

Mianserin is a tetracyclic antidepressant with therapeutic activity similar to amitriptyline used to treat depression and anxiety.

Generic Name

Mianserin

DrugBank Accession Number

DB06148

Background

A tetracyclic compound with antidepressant effects. Mianserin was previously available internationally, however in most markets it has been phased out in favour of [mirtazapine](#).


Type

Small Molecule

Groups

Approved, Investigational, Withdrawn

Structure



Weight

Average: 264.3648  
Monoisotopic: 264.16264865

|              |  |              |                 |
|--------------|--|--------------|-----------------|
| dataset2.txt | dataset3.txt   | dataset4.txt | dataset1.txt    |
| 1            | Calcitriol//Vitamin D-binding protein//Vitamin transporter activity      |              | 图数据             |
| 2            | Ergocalciferol//Vitamin D-binding protein//Vitamin transporter activity  |              |                 |
| 3            | Cholecalciferol//Vitamin D-binding protein//Vitamin transporter activity |              |                 |
| 4            | Flunisolide//Corticosteroid-binding globulin//Steroid binding            |              |                 |
| 5            | Lovastatin//Serum albumin//Toxic substance binding                       |              | 每行表示：头结点，尾结点，关系 |
| 6            | Enflurane//Serum albumin//Toxic substance binding                        |              |                 |
| 7            | Alclometasone//Corticosteroid-binding globulin//Steroid binding          |              |                 |
| 8            | Benzatropine//Serum albumin//Toxic substance binding                     |              |                 |
| 9            | Diethylstilbestrol//Transthyretin//Identical protein binding             |              |                 |
| 10           | Metoprolol//Serum albumin//Toxic substance binding                       |              | 有药物的靶向、酶信息      |
| 11           | Dicoumarol//Serum albumin//Toxic substance binding                       |              |                 |
| 12           | Argatroban//Serum albumin//Toxic substance binding                       |              |                 |
| 13           | Atomoxetine//Serum albumin//Toxic substance binding                      |              |                 |

dataset2

|   | A       | B          | C  | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---------|------------|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | id      | name       | smile  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 2 | DB01296 | Glucosamin | 9101112131415161819201781811283284285286287288289308332338339340341344345346347351352355356366367380393405406528563566567571582592614615617637 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 3 | DB09230 | Azelinidip | 91011121314151618192012913111321178182183184185189192196199283284285286299301332333335338339340341344345346351352355356365366367               |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 4 | DB05812 | Abirateron | 91011121418143147178179182183184185186192199283284285286308332333334335339341344345346351352355356358365366370371372373374                     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 5 | DB01195 | Flecainide | 910111214151819232425178180181182185283284285286287299332333338340341344345346351352355356363365366370371381382384390392393                    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 6 | DB00201 | Caffeine   | 910111415161819143148149178183184283284285286332340351352355357358359365373374375376377378379381384386387388389390391396393                    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 7 | DB01020 | Isosorbide | 91014181920143146150153283284286301308332339341344346352366401405406423455567571582614617637639662663679680681689690701                        |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 8 | DB00278 | Argatroba  | 910111214151618192033178180181182183184185192283284285286293299305308332333335338340344345351352353355365366370371374375376                    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 9 | DB01403 | Methotrin  | 910111214151833178182183184185189192283284285286293332333335340341344345351352353355365366370371374376377381382383384390393                    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

文本数据2

|   | A                         | B            | C            | D            | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---------------------------|--------------|--------------|--------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | dataset2.txt              | dataset3.txt | dataset4.txt | dataset1.txt |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 1 | Glucosamine//9//include   |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 2 | Glucosamine//10//include  |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 3 | Glucosamine//14//include  |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 4 | Glucosamine//18//include  |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 5 | Glucosamine//19//include  |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 6 | Glucosamine//20//include  |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 7 | Glucosamine//178//include |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 8 | Glucosamine//181//include |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 9 | Glucosamine//183//include |              |              |              |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

图数据2

存储smiles（化学方程式）

dataset3

|    | A        | B         | C  | D | E | F | G | H | I | J |
|----|----------|-----------|--|---|---|---|---|---|---|---|
| 1  | targetA  | targetB   | interaction                                      |   |   |   |   |   |   |   |
| 2  | Abemacic | Amiodaro  | The risk or severity of adverse effects increase |   |   |   |   |   |   |   |
| 3  | Abemacic | Apalutam  | The serum concentration decrease                 |   |   |   |   |   |   |   |
| 4  | Abemacic | Aprepitan | The serum concentration increase                 |   |   |   |   |   |   |   |
| 5  | Abemacic | Atomoxet  | The metabolism decrease                          |   |   |   |   |   |   |   |
| 6  | Abemacic | Bortezom  | The metabolism decrease                          |   |   |   |   |   |   |   |
| 7  | Abemacic | Bosentan  | The serum concentration decrease                 |   |   |   |   |   |   |   |
| 8  | Abemacic | Bosutinib | The serum concentration increase                 |   |   |   |   |   |   |   |
| 9  | Abemacic | Carbamaz  | The serum concentration decrease                 |   |   |   |   |   |   |   |
| 10 | Abemacic | Ceritinib | The risk or severity of adverse effects increase |   |   |   |   |   |   |   |

文本数据3

来源于 drugbank

|    | A        | B         | C  | D | E | F | G | H | I | J |
|----|----------|-----------|--|---|---|---|---|---|---|---|
| 1  | targetA  | targetB   | interaction                                      |   |   |   |   |   |   |   |
| 2  | Abemacic | Amiodaro  | The risk or severity of adverse effects increase |   |   |   |   |   |   |   |
| 3  | Abemacic | Apalutam  | The serum concentration decrease                 |   |   |   |   |   |   |   |
| 4  | Abemacic | Aprepitan | The serum concentration increase                 |   |   |   |   |   |   |   |
| 5  | Abemacic | Atomoxet  | The metabolism decrease                          |   |   |   |   |   |   |   |
| 6  | Abemacic | Bortezom  | The metabolism decrease                          |   |   |   |   |   |   |   |
| 7  | Abemacic | Bosentan  | The serum concentration decrease                 |   |   |   |   |   |   |   |
| 8  | Abemacic | Bosutinib | The serum concentration increase                 |   |   |   |   |   |   |   |
| 9  | Abemacic | Carbamaz  | The serum concentration decrease                 |   |   |   |   |   |   |   |
| 10 | Abemacic | Ceritinib | The risk or severity of adverse effects increase |   |   |   |   |   |   |   |

图数据3

存储药物ddi信息

dataset4

|    | A        | B   | C | D | E | F | G | H | I |
|----|----------|---|---|---|---|---|---|---|---|
| 1  | drug     | texts                                     |   |   |   |   |   |   |   |
| 2  | Glucosam | Glucosamine has a MolMR of 2.             |   |   |   |   |   |   |   |
| 3  | Glucosam | Glucosamine has a MolLogP of 4.           |   |   |   |   |   |   |   |
| 4  | Glucosam | Glucosamine has a MolWt of 4.             |   |   |   |   |   |   |   |
| 5  | Glucosam | Glucosamine has a NumRotatableBonds of 2. |   |   |   |   |   |   |   |
| 6  | Glucosam | Glucosamine has a NumAliphaticRings of 2. |   |   |   |   |   |   |   |
| 7  | Glucosam | Glucosamine has a NumAromaticRings of 2.  |   |   |   |   |   |   |   |
| 8  | Glucosam | Glucosamine has a NumSaturatedRings of 3. |   |   |   |   |   |   |   |
| 9  | Glucosam | Glucosamine has no MACCS50 key.           |   |   |   |   |   |   |   |
| 10 | Glucosam | Glucosamine has the MACCS89 key.          |   |   |   |   |   |   |   |
| 11 | Glucosam | Glucosamine has the MACCS98 key.          |   |   |   |   |   |   |   |

文本数据4

简单将图数据转化为文本描述

|    | A                                 | B | C | D | E | F | G | H | I |
|----|-----------------------------------|---|---|---|---|---|---|---|---|
| 1  | Glucosamine//MolMR//2             |   |   |   |   |   |   |   |   |
| 2  | Glucosamine//MolLogP//4           |   |   |   |   |   |   |   |   |
| 3  | Glucosamine//MolWt//4             |   |   |   |   |   |   |   |   |
| 4  | Glucosamine//NumRotatableBonds//2 |   |   |   |   |   |   |   |   |
| 5  | Glucosamine//NumAliphaticRings//2 |   |   |   |   |   |   |   |   |
| 6  | Glucosamine//NumAromaticRings//2  |   |   |   |   |   |   |   |   |
| 7  | Glucosamine//NumSaturatedRings//3 |   |   |   |   |   |   |   |   |
| 8  | Glucosamine//MACCS50//0           |   |   |   |   |   |   |   |   |
| 9  | Glucosamine//MACCS89//1           |   |   |   |   |   |   |   |   |
| 10 | Glucosamine//MACCS98//1           |   |   |   |   |   |   |   |   |
| 11 | Glucosamine//MACCS108//0          |   |   |   |   |   |   |   |   |
| 12 | Glucosamine//MACCS116//0          |   |   |   |   |   |   |   |   |
| 13 | Glucosamine//MACCS121//0          |   |   |   |   |   |   |   |   |
| 14 | Glucosamine//MACCS136//0          |   |   |   |   |   |   |   |   |
| 15 | Glucosamine//MACCS137//1          |   |   |   |   |   |   |   |   |
| 16 | Glucosamine//MACCS141//0          |   |   |   |   |   |   |   |   |
| 17 | Glucosamine//MACCS145//0          |   |   |   |   |   |   |   |   |

图数据4

存储药物是否有MACCS键值信息及7项物理特性

## 问题

- 1、dataset2 中的文本描述可以用  
CN1CCN2C(C1)C1=CC=CC=C1CC1=CC=CC=C21 该种方式表示，看后续实验效果是否修改
- 2、dataset4 中还没找到关于药物 MACCS 键值的描述，所以暂时只是简单将图数据转化成文本数据

## 模型、代码

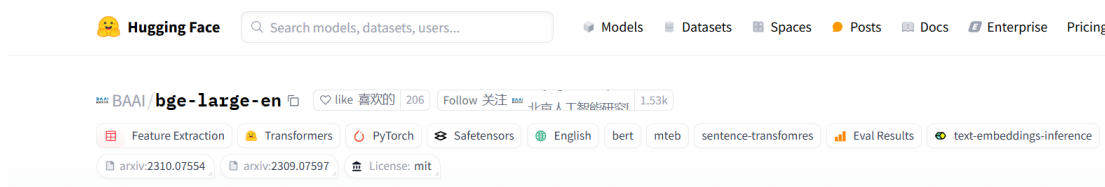
### 进展

代码方面还没做出修改

初步打算在代码的 Embedding 那边拼接文本数据的向量，然后在 Embedding 加一个 LLM 充分融合两种向量。

```
279 class FusionLayer(nn.Module):
280     def __init__(self, args):
281         nn.init.xavier_uniform_(m.weight)
282         nn.init.zeros_(m.bias)
283         elif isinstance(m, nn.BatchNorm1d):
284             nn.init.constant_(m.weight, val: 1)
285             nn.init.constant_(m.bias, val: 0)
286
287     def forward(self, arguments):
288         gnn4_embedding, gnn3_embedding, gnn2_embedding, gnn1_embedding, idx = arguments
289
290         idx = idx.numpy().tolist()
291         drugA = []
292         drugB = []
293         for i in idx:
294             drugA.append(i[0])
295             drugB.append(i[1])
296
297         Embedding = torch.cat([gnn1_embedding[drugA], gnn2_embedding[drugA], gnn3_embedding[drugA], gnn4_embedding[drugA],
298                               gnn1_embedding[drugB], gnn2_embedding[drugB], gnn3_embedding[drugB], gnn4_embedding[drugB]], dim: 1).float()
299
300         return self.fullConnectionLayer(Embedding)
```

准备使用 bge-large-en 作为 LLM



## 问题

1、文本数据通过 LLM 输出的向量是否需要动态？

目前打算先用静态的试一试，后续试试动态。

## 流程

整个项目实现的流程是清楚了

## 下一步计划

完成文本数据向量和图数据向量的拼接、融合