

## 补充方法

黄金标准 DDI 数据集的准备。 DeepDDI2 的黄金标准 DDI 数据集

使用 DrugBank1 (版本 5.1.1)制备,涵盖来自 222,041 个药物对的 222,127 个 DDI

(补充数据 1),代表 113 种 DDI 类型(补充数据 2)。新的

准备好的数据集被分成模型的训练(80%)和测试(20%)数据集

发展;在训练数据集中,10% 用于验证。

生成结构相似性概况。输入的结构相似性概况 (SSP)

药物的产生方式与之前描述的相同<sup>2</sup>,但有更多的

批准药物比上一版 DeepDDI:批准药物从 2159 个减少到 2386 个

批准的药物。从 DrugBank1 (5.1.1 版)获得了批准药物的清单。

这些批准的药物充当所谓的“固定比较目标”。简而言之,SSP 是

通过获取输入药物与 2,386 种已批准药物之间的 Tanimoto 系数生成

以成对的方式。随后,输入对中每种药物的两个 SSP 被简化为

50 维使用主成分分析,合并生成特征向量,

即具有 100 个维度的“组合 SSP”。最终处理合并的 SSP

通过 DeepDDI2 的多层感知器 (MLP) 来预测输入药物对的 DDI 类型。

开发用于预测 DDI 类型的深度学习模型。更新后的 DeepDDI2

有 113 个输出神经元,每个神经元代表一个特定的 DDI 类型。 113 DDI 的定义

类型在补充数据 2 中以句子的形式提供。DeepDDI2 被设计

执行多标签分类,其中可以为药物对分配多个 DDI

类型。为了在 DeepDDI2 中训练 DNN,DNN 的权重是根据

均匀分布。对于激活函数,整流线性单元 (ReLU) 用于

隐藏层,而 sigmoid 函数用于输出层。 DNN 被训练

最多 300 个时期,提前停止设置为 10 个时期,使用二元交叉熵作为损失

函数和 Adam 方法<sup>3</sup>进行优化。此外,蒙特卡洛 (MC) dropout<sup>4</sup>是

用于通过生成十个 MC 来衡量预测的不确定性 (即标准差)

在模型训练和新预测期间丢弃样本。平均值和标准

十个预测结果的偏差也作为 DeepDDI2 的输出生成。

超参数,包括隐藏层数,每层节点数

隐藏层、学习率和辍学率,针对 F1 分数进行了优化

贝叶斯优化<sup>5</sup> (补充表 1)。通过检查 50 种不同的架构

DNN,三个隐藏层,每层 1,024 个节点,学习率为 0.0069,dropout

0.3 的比率显示了 DeepDDI2 的最佳性能。输出阈值

神经元的活动 (预测平均值 - 标准偏差的一半)设置为 0.27;有这个

threshold 在 0 到 1 之间的阈值中产生了最好的 F1 分数

0.01 的间隔。

处理来自 DNN 的初始预测结果。最初预测的 DDI 类型

如果 DeepDDI2 满足以下两个标准<sup>6</sup>,则被认为是有效的; 1)药物应该

存在于黄金标准 DDI 数据集中,其结构与两个输入之一相似

药物 (谷本系数 > 0.7) ;和 2) 这些来自 DeepDDI2 的结构相似的药物和

数据集应与相同的 DDI 类型相关联。

从 FDA 不良事件报告中准备处方药数据集

系统 (FAERS)。在 FDA FAERS 数据库中报告的 2,159,674 例 (处方)中

从 2012 年第四季度到 2018 年第二季度 (<https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html>),只有那些涉及两到五种药物的药物被分配了 DrugBank ID,

在本研究中考虑。结果,从 301,615 个 FAERS 病例中提取了 2,248 种处方药

在本研究中被考虑。作为运行 DeepDDI2 的结果,1,628 种处方药  
预计会与 nirmatrelvir 和/或 ritonavir 发生 DDI (补充数据 3)。

开发环境和数据可用性。所有的计算都已实现

在 Linux Ubuntu 环境中使用 Python 3.6。 Python 包 Keras (版本 2.0.6)

(<https://keras.io/>)和 TensorFlow 后端 7 (版本 1.2.1)用于开发 DNN

DeepDDI2 的。更新后的 DeepDDI2 的源代码可在

<https://bitbucket.org/kaistsystemsbiology/deepddi2>。

补充表

补充表1. 为 DeepDDI2 中的 DNN 检查的超参数。

超参数	测试值	DDI 类型预测的最优值
隐藏层数	3, 4, 5, 6, 7, 8	3个
每层节点数	256, 512, 1024	1024
辍学率	0.3、0.4、0.5	0.3
学习率	log0.005 - log0.2	0.0069

## 补充参考资料

1. Wishart, DS,等人。 DrugBank 5.0:2018 年 DrugBank 数据库的重大更新。核酸研究。 46, D1074-D1082 (2018)。
2. Ryu, JY, Kim, HU & Lee, SY 深度学习改进了药物-药物和药物-食物相互作用的预测。过程。国家队。学院。科学。美国115, E4304-E4311 (2018)。
3. Kingma, DP & Ba, J. Adam:一种随机优化方法。 arXiv (2014)。
4. Gal, Y. & Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning.在诠释。会议。马赫。学习。 1050-1059 (2016 年)。
5. Snoek, J., Larochelle, H. & Adams, RP 机器学习算法的实用贝叶斯优化。在进阶神经信息过程。系统。 2951-2959 (2012)。
6. Vilar, S.,等人。基于相似性的药物相互作用大规模预测建模。纳特。协议书。 9, 2147-2163 (2014)。
7. 阿巴迪,M.,等。 TensorFlow:一个用于大规模机器学习的系统。 Osd 16 论文集:第 12 届 Usenix 操作系统设计与实现研讨会, 265-283 (2016)。