

利用多图卷积网络来感知综合症的 草药推荐方法

Syndrome-aware Herb Recommendation with Multi-Graph Convolution Network

Yuanyuan Jin*, Wei Zhang*, Xiangnan He[†], Xinyu Wang* and Xiaoling Wang*[‡]

* Shanghai Key Laboratory of Trustworthy Computing, East China Normal University
Shanghai, China

{yyj, xinyuwang}@stu.ecnu.edu.cn, zhangwei.thu2011@gmail.com, xlwang@cs.ecnu.edu.cn

[†]School of Data Science, University of Science and Technology of China
Hefei, China, xiangnanhe@gmail.com

[‡]Shanghai Institute of Intelligent Science and Technology, Tongji University
Shanghai, China

Published in: 2020 IEEE 36th International Conference on Data Engineering (ICDE)

Date of Conference: 20-24 April 2020

INSPEC Accession Number: 19760493

Date Added to IEEE Xplore: 27 May 2020

DOI: 10.1109/ICDE48307.2020.00020

► ISBN Information:

Publisher: IEEE

► ISSN Information:

Conference Location: Dallas, TX, USA, USA

1. 论文简介

草药推荐问题，创新点受到传统中医治疗过程的启发，将由多个症状推导出综合症的过程引入进来，研究进而变为：

症状和草药的单一交互关系----->综合症与草药的交互关系

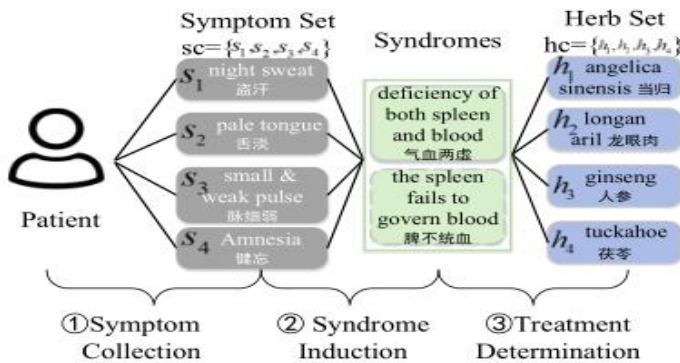


Fig. 1. An example of the therapeutic process in TCM.

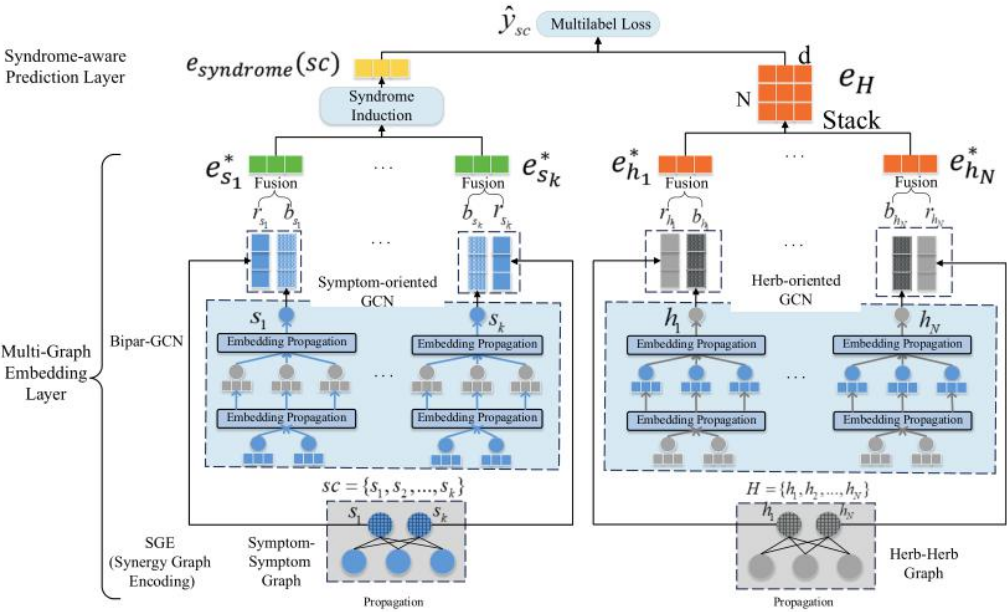
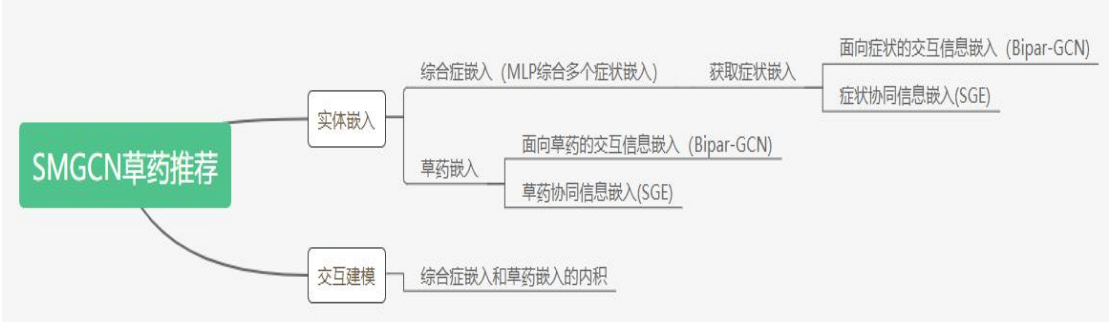
2. 论文具体思路

推荐任务可以概括为：

步骤	含义	领域范畴	具体技术
----	----	------	------

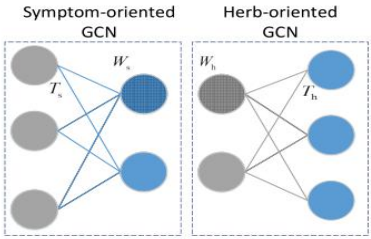
获取实体嵌入	获取能反映实体属性和实体间交互关系的嵌入表示	图表示学习	同质图的结点表示 异质图的结点表示
对交互关系建模	即构建交互函数		简单内积、 结合潜在因子分解和 MLP (论文 Neural Collaborative Filtering)

本论文的实验思路：



(1) Bipar-GCN

构建此组件的原因：建模草药和症状两种实体间的关系，



组件：

说明：a.图卷积通常有两类，一类是谱域图卷积，即利用图上的傅里叶变化将空间域的卷积转化到谱域上，如 SCNN、ChebNet、GCN；另一类是空域图卷积，即利用结点在空间中的

邻居结点聚集信息，与矩阵的卷积很类似，如 GraphSAGE、GAT、PGC 方法；这里使用的是后一类

b. 具体的卷积过程：选择邻居结点、构造消息、聚合消息、高阶传播

Neural Graph Collaborative Filtering*

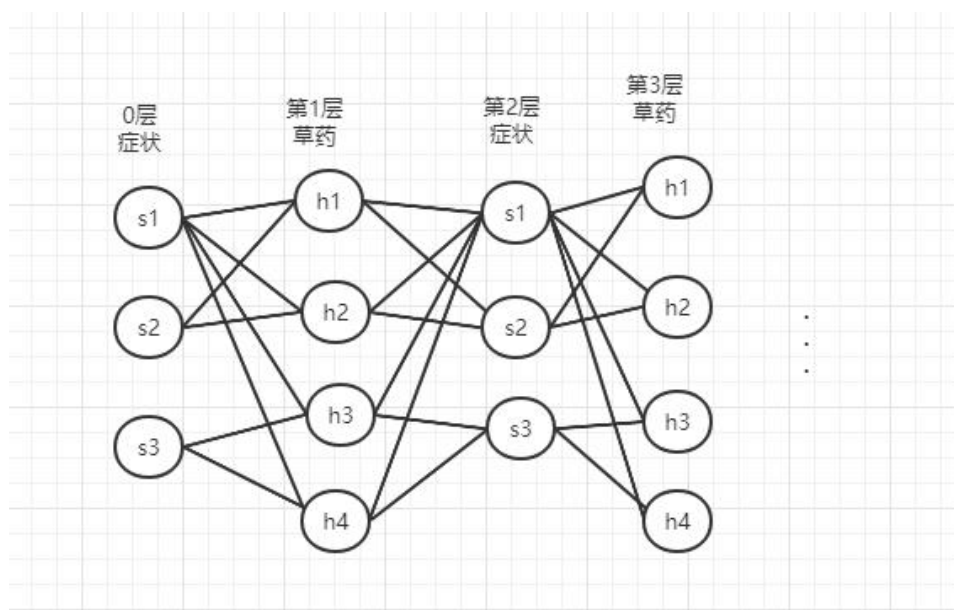
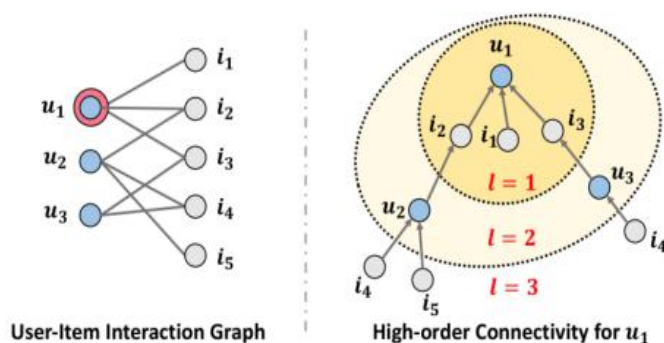
Xiang Wang
National University of Singapore
xiangwang@u.nus.edu

Xiangnan He[†]
University of Science and Technology
of China
xiangnanhe@gmail.com

Meng Wang
Hefei University of Technology
eric.mengwang@gmail.com

Fuli Feng
National University of Singapore
fulifeng93@gmail.com

Tat-Seng Chua
National University of Singapore
dcscts@nus.edu.sg



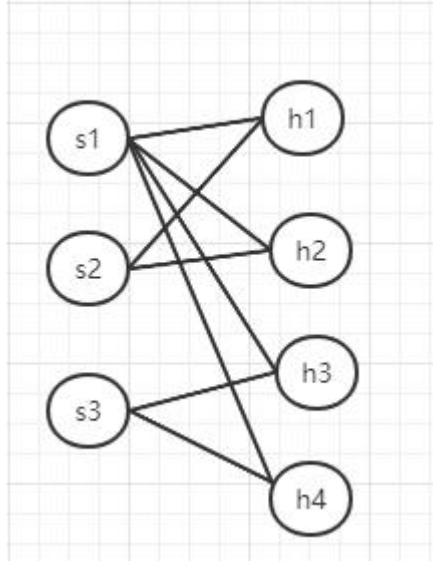
c. 草药嵌入和症状嵌入的方法略有不同，区别在于第 0 层即初始嵌入层的对象不同；

(2) SGE 协同图编码

构建此组件的原因：考虑到推荐的草药集在药理上应该是相互兼容的，症状间存在继发、相互影响等联系，我们需要建模同种实体间的关系；

同时引入更多信息有助于缓解中药处方的数据稀疏性问题

不使用“二部图的 2 阶近邻”建模同质关系的原因：



$p1=\langle\{s1,s2\},\{h1,h2\}\rangle$ 和 $p2=\langle\{s1,s3\},\{h3,h4\}\rangle$,
 $\{h2,h3,h4\}$ 通过连接 $s1$ 成为 $h1$ 的 2 阶邻居。因此在协同图中, $h1$ 和 $h3$ 、 $h1$ 和 $h4$ 没有边。

组件： 草药-草药图，症状-症状图

说明： a. 草药-草药图、症状-症状图实际上是同质图，不过在逻辑上可以看成是同类实体构成的特殊二部图；

b.将处方中草药间和症状间同时出现的次数作为衡量协同效果的强度，设置相应的阈值，超过阈值才被视作具有协同效果；

c.论文中使用邻接矩阵来表示图，其中具有协同效应记作 1，而不是使用其次数 n ，我觉得使用次数更好；

$$HH_{h_m, h_n} = \begin{cases} 1, & \text{if frequency}(h_m, h_n) > x_h; \\ 0, & \text{otherwise.} \end{cases}$$

其中HH表示为草药-草药图， x_h 是草药-草药对的阈值。

d.对协同图进行一层图卷积以期获得嵌入：

$$\begin{aligned} r_s &= \tanh(\sum_{k \in N^{ss}_s} e_k \cdot V_s) \\ r_h &= \tanh(\sum_{q \in N^{hh}_h} e_q \cdot V_h) \end{aligned} \quad (10)$$

其中 e_k 和 e_q 分别是症状 k 和草药 q 的初始嵌入， N^{ss}_s 表示SS中 s 的邻居集， N^{hh}_h 表示HH中 h 的邻居集。 V_s 和 V_h 权重参数分别用于SS和HH。

(3) 信息融合：

$$\begin{aligned} e_s^* &= r_s + b_s \\ e_h^* &= r_h + b_h \end{aligned} \quad (11)$$

其中 e_s^* and e_h^* 分别是症状 s 和药草 h 的合并嵌入。

(4) 综合症感知

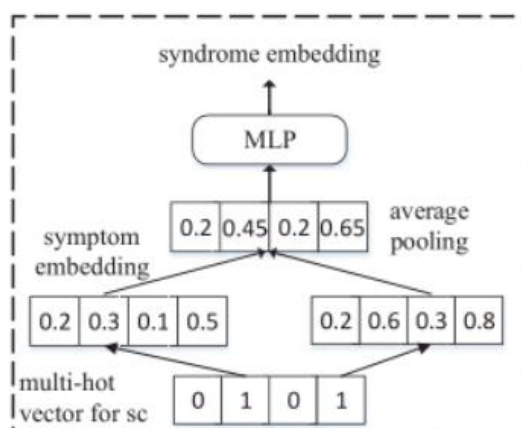


Fig. 4. The MLP-based method for syndrome induction.

输入：所有症状的合并嵌入

处理过程：一层平均池化层、一层 MLP

$$e_{syndrome}(sc) = \text{ReLU}(W^{mip} \bullet \text{Mean}(e_{sc}) + b^{mip}), \quad (12)$$

其中 $e_{syndrome}(sc)$ 是针对 sc 的诱导综合症嵌入。

(5) 模型训练：多标签损失函数

$$Loss = \arg \min_{\theta} \sum_{(sc, hc') \in P} WMSE(hc', g(sc, H)) + \lambda_{\theta} \|\Theta\|_2^2$$

$$g(sc, H) = e_{syndrome}(sc) \cdot e_H^T \quad (13)$$

3. 论文实验部分

基线：

基线种类	包含	含义
主题模型	HC-KGETM	不仅要考虑中医处方中的共现信息，还要考虑知识图谱中症状和草药的全面语义相关性
基于图神经网络的模型	GC-MC PinSage NGCF	分别利用 GCN 获得用户和项目的表示、GraphSAGE 在项目-项目图上的工业应用、显式构造了二部用户-项目图以对高阶连通性进行建模
作者提出的基线	HeteGNN	基于内容的异质图表示学习问题，异质包含异质图结构信息和异质节点内容（比如文本、图像）两部分信息； 主要过程包括重定位的随机游走采样邻居节点、聚合异质的节点内容嵌入、聚合同一类型的邻居节

		点、聚合不同类型的邻居节点
--	--	---------------

4. 思考和发现

(1) 论文中 Bipar-GCN 部分提到了 GraphSAGE 来聚合消息，公式中却只用了一个连接，但是 GraphSAGE 是对不同阶的邻居采用不同的聚合函数，并不用于嵌入连接，不知道其是否指的是 GraphSAGE 中使用的聚合方法；

3) 消息聚合：

在接受到合并的邻居表示后，下一步是更新目标结点的嵌入。在这里我们采用【9】中提取出来的GraphSAGE聚合器，它将两个表示形式连接起来，然后进行非线性激活操作。第一层的症状表示 b_s^1 和草药表示 b_h^1 可以定义如下：

$$b_s^1 = \tanh(w_s^1 \bullet (e_s \parallel b_{N_s}^0)) \quad (4)$$

$$b_h^1 = \tanh(w_h^1 \bullet (e_h \parallel b_{N_h}^0)) \quad (5)$$

其中 \parallel 表示两个向量的连接操作， W_s 和 W_h 分别表示症状和草药的聚合权重矩阵。

(2) 还没有推导 Bipar-GCN 的矩阵表示形式，论文中只给出的是单一结点的，代码中需要使用矩阵

(3) 本文中 Bipar-GCN 所述方法引用了何向南团队 2019 年的一篇论文 “Neural Graph collaborative filtering”，后找到今年的一篇论文也引用了并对这个框架做了一点泛化，后面可以试试使用它改进的 Bipar-GCN:

Multi-Graph Convolution Collaborative Filtering

Jianing Sun*, Yingxue Zhang*, Chen Ma^{†*}, Mark Coates[†], Huifeng Guo[‡], Ruiming Tang[‡], Xiuqiang He[‡]

^{*}Montreal Research Center

Huawei Noah's Ark Lab, Montreal, QC, Canada

{jianing.sun, yingxue.zhang}@huawei.com

[†] McGill University, Montreal, QC, Canada

chen.ma2@mail.mcgill.ca, mark.coates@mcgill.ca

[‡]Huawei Noah's Ark Lab, Shenzhen, China

{huifeng.guo, tangruiming, hexiuqiang1}@huawei.com