

## [2020.3.5] 思路v1.0

在这篇论文（智能医疗诊断系统的研究与实现）中一共涉及**7836**个疾病，**6617**种症状，**54**个科室，疾病和症状之间的关系有**77658**个。实现的功能是根据用户输入的症状信息，得到患有疾病的概率。计算了一些统计学的规律，比如疾病与症状，症状与症状，疾病与性别等之间的关系。

### 一、信息采集模块

1. 数据爬取：在线医疗网站：好大夫在线，39健康，寻医问药等等，诊断结果也需要。知乎：抑郁症的症状有哪些。抑郁症一般的网站描述。
2. 数据清洗，人工滤除不合适的数据。
3. 分词：基于词典的分词，语义相等，参考同义词林等方式找到同义词，通过随机向量场+马尔科夫等形式发现词。
4. 词向量的训练（为了寻找同义词）。
5. 数据挖掘，症状与症状之间的关系，疾病与症状，统计词频，求概率等。目的：对抑郁症患者的数据进行一个统计学方面的分析，并且为了寻找症状与症状间的相关关系。

### 二、问题理解模块

1. 根据用户症状表述，存在表述不清楚的问题，进一步询问（对话的生成，考虑基于模式规则的方式（优先），深度学习方式等）
2. 对用户的询问给出回答（比如对症状不清楚的地方，或者对通识知识的回答，对于这些知识会爬取）通过闲聊接入对话。量表的问题，直接通过语音合成问题。
3. 对不足以支撑诊断，症状再加以询问。根据数据采集模块的分析，判断出疾病与症状，症状与症状之间的依赖关系，以及诊断疾病所必须的条件。对程度的鉴别。**结合量表诊断规则标准的给出选项让病人进行选择**
4. 自杀预警（量表）
5. 对相应的症状采取一些干预措施（若干预措施明确的情况下，涉及对话的生成）
6. 合适的时候对用户进行量表的测试（什么时候给做量表？根据什么判断）
7. 难以应对时接入人工

### 三、疾病诊断模块

1. 根据相应的症状给出患病概率
2. 根据症状，疾病给出治疗建议

疾病诊断可以通过推理规则实现，基于症状和疾病间的关系，结合朴素贝叶斯给出概率。、

症状+限定词+程度词

根据回答再给出相应的问题，给出abcd

诊断分为两个模块：

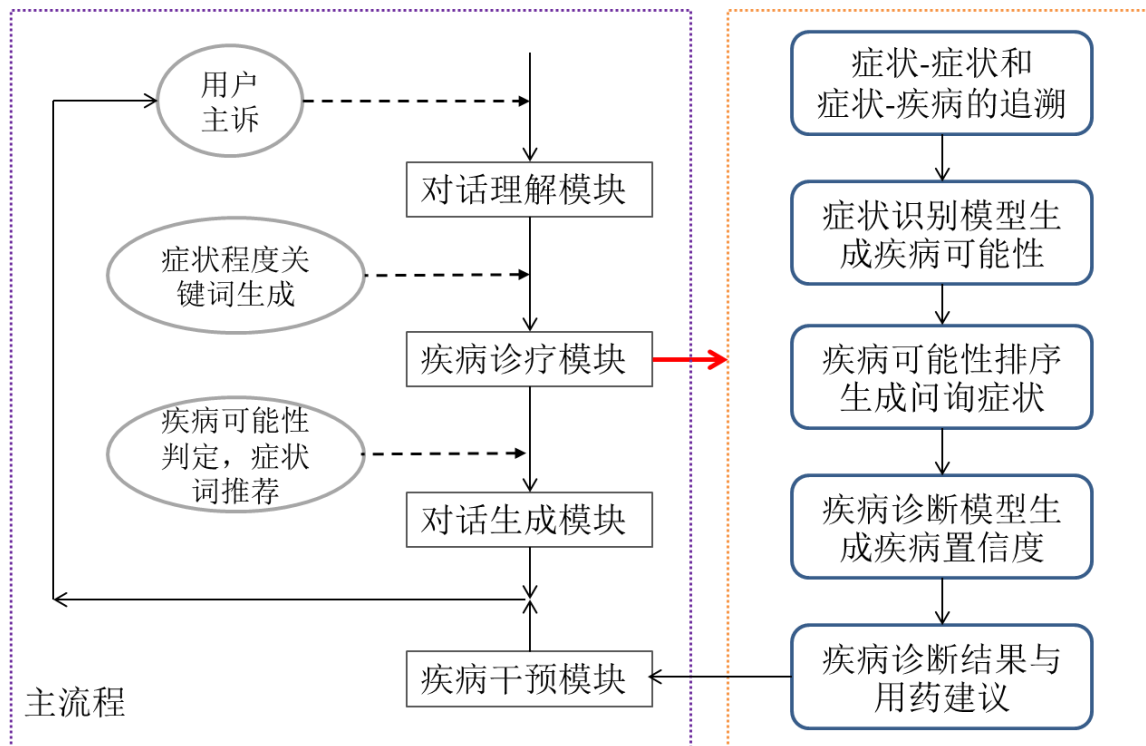
1. 根据初步叙述，得到基本疾病的诊断
2. 然后更新问题。

知识图谱的作用：

1. 对疾病的可视化方式。
2. 对病人信息知识图谱呈现方式

## [2020.3.27] 思路v1.1

现已基本完成信息采集模块等功能的基础上，对思路进行新一步细致的梳理：



### 一、对话理解模块

基于词典、词向量模型，同义词词表以及实体关系识别模型对实体进行识别，思路如下：

用户输入一句话，按，。分隔开分别对每一小句进行处理，先基于词典提取出时间，程度，频率（**ac**自动机匹配加速），因为这些词都是比较容易匹配到词典的，就是说这些词的变化不大，词典的内容大概可以包含。比较难处理的是症状，我目前的想法是，对于已给词典的症状如果能匹配到词典中是最好的，如果匹配不到，则利用实体关系识别，找到标注为症状的词，然后利用词向量模型找到最相近的几个症状词，由于这些症状词不一定出现在我们的词典中，所以还需要利用同义词词典来计算症状词的相似度，进而匹配到我们的词典中。

——from Guoguang

——from Miaomiao

result.txt

```
1 吃不下饭 不想吃 0.7313481603889138
2 晚上睡不着觉 睡不着 0.9431022951007094
3 容易暴躁 容易醒来 0.7159076517731544
4 每天心情压抑烦躁不安 烦躁不安 0.7900946308959323
5 想割腕 割腕 0.9292731288492808
6 对很多事情没兴趣 没有意义 0.784378742064146
7 喜欢独处 独处 0.9298011494772463
8 心情压抑 压抑 0.8910361724931872
9 经常头脑发热 有过想自杀 0.5556703540907528
10 神情恍惚 担心失控 0.42198934056482673
11 晚上总偷偷地哭 开心不起来 0.7001518181034025
12 没有希望 希望自己已经死了 0.8669156540117382
13 喜欢胡思乱想 胡思乱想 0.8787734169552018
14 意志消沉 生活疏懒 0.546361387677778
15 目光呆滞 视线模糊 0.558705149380255
16 身体无力 头痛痛 0.7102193535493578
17 不爱交流 不想说话 0.7864915203283931
18 容易发火 发脾气 0.6863460055452628
19 睡眠多梦 失眠 0.7873521818195992
20 情绪不稳定 不开心 0.7626135254017424
```

——from Liujia

(a) BIOES 序列标注结果			(b) 中间结果
263	检 0		我[@始终#adfrequencyhigh*]还是[@没有#denydes*]办法[@喜欢#nsym*][@自己
264	查 0		#nsym*], 我[@觉得#vsym*]我[@不#denydes*][@可能#affirmdes*][@会
265	之 B-adfrequencyhigh		#denydes*]被人[@喜欢#nsym*], 怎么[@可能#affirmdes*][@会#denydes*]有人[@
266	后 E-adfrequencyhigh		喜欢#nsym*]我, 当我[@有#affirmdes*]了[@很#addegreedeeep*][@好
267	, 0		#adescription *]的[@朋友#nsym*][@之后#adfrequencyhigh*], 我就[@每天
268	确 0		#nproptime*][@都#addegreedeeep*][@很#addegreedeeep*][@焦虑#nsym*], 我[@
269	诊 0		每天#nproptime*][@都#addegreedeeep*][@很#addegreedeeep*][@担心#vsym*],
270	的 0		他们[@会#denydes*][@不会#denydes*][@不#denydes*][@喜欢#nsym*]我, [@会
271	重 B-ndegreedeeep		#denydes*][@不会#denydes*][@讨厌#adescription*]我, [@会#denydes*][@不会
272	度 E-ndegreedeeep		#denydes*][@觉得#vsym*]我很烦, 我跟他们[@讲#vsym*]这个话他们[@会
273	抑 B-asm		#denydes*][@不会#denydes*][@生气#asm*], 他们[@对#affirmdes*]我[@好
274	郁 E-asm		#adescription *]是不是只是因为同情我敷衍我, 我什么[@都
275	重 B-ndegreedeeep		#addegreedeeep*][@很#addegreedeeep*][@害怕#nsym*]
276	度 E-ndegreedeeep		
277	焦 B-nsym		
278	虑 E-nsym		
279	, 0		
280	今 B-nproptime		
281	年 E-nproptime		
282	三 0		
283	月 0		

注：难点在于症状，问最终是否可以拿到可以对应到知识图谱中的症状？

## 二、疾病诊疗模块

1. 将模块一中得到的症状实体在知识图谱中查询，若搜索到存在这个症状，判断症状的层级（up\_symptom, down\_symptom等），往上依次追溯到最大类可用于诊断的症状，并且得到具备该症状可能所患的疾病。
2. 若追溯到多个疾病，则记录多个疾病的可能性（通过症状的多少计算），症状程度、时间、肯定否定的判定，在这里构建一个数学模型进行计算，例如躁狂发作30%，抑郁发作50%（这主要是根据符合疾病症状的标准给出）。
3. 针对以上分析，涉及对应多个疾病时的症状询问问题，设置一个公式进行询问，根据所患疾病的可能性进行排序，询问可能性最高疾病的症状,直到拿到所有症状可以诊断出疾病。这里存在的一个

问题是只询问大致症状还是还需询问具备大症状下的其他小症状，那么小症状如何确定。目前先手动建立大症状下级的小症状中的必问症状集合和机动集合。

4. 疾病的诊断，设置一个诊断模型，疾病的诊断结果以及置信度的设置，例如轻度抑郁发作置信度90%（深度考虑到疾病的程度等其他方面要求，已经诊断为抑郁发作，可信度为90%）。
5. 疾病治疗，根据疾病的诊断结果结合症状表现给出用药建议。

诊断中会涉及的数据：

1. 疾病：大类疾病 $\{Dis_1, Dis_2, \dots, Dis_n\}$ ，大类疾病下对应的小类疾病 $Dis_k = \{dis_1, dis_2, \dots, dis_m\}$ ，即一共有 $n$ 种大类疾病，每种大类疾病下对应 $m$ 个小类疾病，小类疾病一般指疾病的轻中重程度。
2. 症状：一般诊断标准以大类疾病进行划分， $DisSym_k = \{Sym_1, Sym_2, \dots, Sym_q\}$ 。但小类疾病中可能存在一些特殊的症状表现， $disSym_k = \{Sym_1, Sym_2, \dots, Sym_p\}$ 。一般疾病直接以大症状的标准为主，大症状可划分为小症状， $Sym_k = \{sym_1, sym_2, \dots, sym_e\}$ 。小症状又可映射到很多同义症状，在此不进行表示。
3. 修饰词：修饰词包括程度、时间、频率以及肯定否定。
  - 程度词包含 $Degree = \{DeepDegree, ShallowDegree, OtherDegree\} = \{90\%, 65\%, 50\%\}$ 。
  - 时间词包含 $Time = \{ntime, nproptertime, atime\}$ 。
  - 频率词包含 $Frequency = \{HighFre, LowFre, NumOther, OterFre\} = \{90\%, 65\%, \text{待定}, 50\%\}$ 。
  - 肯定否定包含 $Description = \{DenyDes, AffirmDes\}$ 。

诊断流程，对于对话理解模块生成的一个症状列表：

$$[\{s : s1, deg : d1, time : t1, fre : f1, des : des1\}, \{s : s2, deg : d2, time : t2, fre : f2, des : des2\}, \dots, \{s : su, deg : du, time : tu, fre : fu, des : desu\}]$$

1. 将这 $u$ 个大症状或者小症状追溯到最大症状，可以得到 $l$ 个大症状，可能符合不同疾病中的症状， $l \in \{0, 1, \dots, l\}$ 。则对可能所患疾病计算（这里计算只需要考虑症状和肯定词否定词以及满足的时间关系，不考虑程度和频率）可能性： $P(Dis_k | s1, s2, su) = \frac{x}{q}$ ，其中， $q$ 为该大疾病的大症状数量， $x$ 为与症状列表匹配的症状数量。对概率不为0的疾病的可能性进行排序，下面即可询问可能患有的最大可能性疾病的症状进行询问。
2. 同时检测 $P(Dis_k) = \frac{x}{q} > \frac{\theta}{q}$ ，其中 $\theta$ 为诊断标准的阈值。若还未达到大类疾病的最严重诊断结果且还有症状没有问完，则继续询问得到症状列表返回第1步，否则得到疾病诊断结果进行第3步。
3. 求具体小疾病 $dis_k$ 的置信度：

$$P(dis_k | s \in \{0, 1, \dots, l\}) = \frac{l}{\theta} \cdot \frac{\sum_{i=1}^l \frac{\sum_{j=1}^{e_i} deg_{ij}}{e_i}}{l}$$

其中， $\theta$ 为诊断标准的阈值， $l$ 为满足大症状的数量， $e_i$ 为大症状 $i$ 中小症状的数量， $deg_{ij}$ 为大症状 $i$ 中小症状 $j$ 下的频率，程度数值。因为程度频率一般不会同时存在，则近似为一个，若同时都存在，则 $\frac{fre+degr}{2}$ 。以上数值可能为 $> 1$ ，目前先按照这个式子，之后可调整为更合适的。

4. 治疗方案给定，根据所患的疾病给出药物推荐，同样的根据病人的症状进行必要的推荐。

### 三、对话生成模块

目前这里的对话生成模块主要是针对病人最先开启的主诉症状开展可能所患疾病的症状的询问：

1. 比如针对抑郁症询问，“最近有情绪低落的情况出现吗”，“最近的体重怎么样呢？”等；
2. 根据用户症状表述，存在表述不清楚的问题或者病人对询问的症状不清楚的地方，进一步询问；
3. 对不足以支撑诊断，症状再加以询问。根据疾病与症状之间的依赖关系，以及诊断疾病所必须的条件。对程度的鉴别，给出深层一步的问题，问题涉及的症状需要有诊疗模块提供。功能类似于1。

```
model init finished .....
用户:我最近总是有睡眠障碍的问题,睡不着觉.
preques_type: new_symptom, question_types: new_symptom.
ROBOT: 了解了,然后你的爱好有没有明显的变化?
用户:什么意思,应该还好吧.
preques_type: new_symptom, question_types: spec_symptom.
ROBOT: 详细来说,有没有喜欢的事情失去愉悦感,等这样的症状?
用户:嗯,最近经常
preques_type: spec_symptom, question_types: new_symptom.
ROBOT: 好的那么你最近的体重怎么样呢?
用户:最近几周我的体重下降了很多,不怎么吃得下饭
preques_type: new_symptom, question_types: new_symptom.
ROBOT: 好的那么你的食欲有没有明显的变化?
用户:食欲减退了
preques_type: new_symptom, question_types: ques_degree.
ROBOT: 情况是持续性的吗,是几乎每天都这样吗?
用户:是的
preques_type: ques_degree, question_types: new_symptom.
ROBOT: 了解了,然后你最近的睡眠怎么样呢?
用户:还行,没有什么异常
preques_type: new_symptom, question_types: spec_symptom.
ROBOT: 也就是,有没有入睡难,易醒,早醒,嗜睡,等这样的症状?
用户:没有
preques_type: spec_symptom, question_types: new_symptom.
ROBOT: 嗯嗯,明白了.你的运动状态还好吗?
用户:嗯,应该还不错吧
preques_type: new_symptom, question_types: spec_symptom.
ROBOT: 就像是,有没有说话、行动缓慢,等这样的症状?
用户:没有的
preques_type: spec_symptom, question_types: new_symptom.
ROBOT: 了解了,然后你的行为动作状态还好吗?
用户:嗯,感觉没有奇怪的地方
preques_type: new_symptom, question_types: spec_symptom.
ROBOT: 也就是,有没有行为易怒易烦躁,等这样的症状?
用户:|
```

## 四、疾病干预模块（后续）

疾病干预进行的方式主要是对话语音交互，必要的话播放视频音频以及展示图片等。之前一直在说加入灵活的对话，而对话目的一直不清。在此可以明确，作为对疾病完成诊断后进行一些干预手段，因为都是精神疾病，通过此种方式可以有一些效果。那么，加入灵活对话的目的以及干预的方式为：

- 分析用户的情绪状态，缓解病人的情绪，进行情感方面的干预，例如平复情绪，开展幽默对话等；
- 深入的问询症状与病人情况，根据实际情况延伸，主要是俏皮聊天。可直接接入开源对话系统的相关内容，例如：

请输入：*我最近对什么事都提不起兴趣*  
青云客\_回答： 那个那个什么事来着  
思知\_回答： 我最近对什么事都提不起兴趣  
请输入：*那怎么办*  
青云客\_回答： 我陪你好么  
思知\_回答： 小思也不知道了，唉  
请输入：*失眠是什么原因*  
青云客\_回答： 你在我心里  
思知\_回答： 大脑睡眠功能紊乱，躯体疾病，精神障碍，药物滥用等  
请输入：*你可以讲个笑话吗*  
青云客\_回答： ★ 我早就了解你了{br}一个小伙于向姑娘求婚，姑娘说：“不过，  
思知\_回答： 你怎么不讲  
请输入：*我不会*  
青云客\_回答： 不会就得学呀~  
思知\_回答： 《我不会》是连载在连城读书的同人类小说，作者是桦彦。  
请输入：*比较容易自卑没有自信*  
青云客\_回答： 问世间情为何物，只教人生死相随  
思知\_回答： 做任何事情都要拿出你十足的勇气，才会成功哦

思知机器人可添加个人知识库，创建问题和回答，更符合项目的需求。[思知AI机器人开放平台](#)

- 医学领域的专业干预手段，利用医学心理干预方式，形成完成的课程、方法等干预体系（需要咨询医院）。

## 可构建的知识图谱

- 问答知识图谱（医学百科知识）
- 抑郁症诊断治疗知识图谱（专业诊疗知识）
- 干预陪伴知识图谱（闲聊类知识以及干预手段知识）
- 病人知识图谱