

中醫診治系統

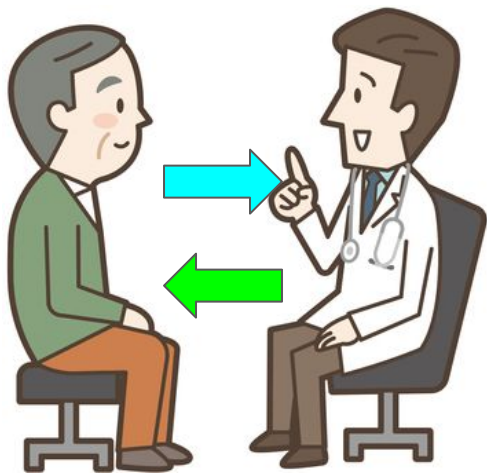
- 以桂枝湯為例 -

林依蓁

2018/12/6

系統架構

使用情境(人機互動)



Patient : 我頭痛、發燒

Doctor : 還有其他症狀嗎？

Patient : 覺得好冷

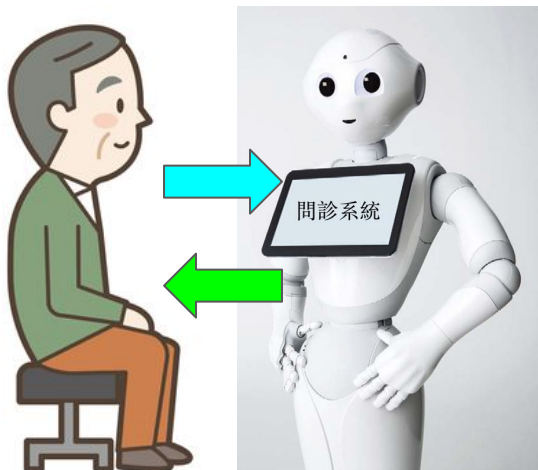
Doctor : 有流汗嗎？

Patient : 有

Doctor : 我開個桂枝湯證給你吃。

系統架構

使用情境(人機互動)



User : 我頭痛、發燒

Robot : 是體溫38.5度C嗎？

User : 是

Robot : 還有其他症狀嗎？

User : 覺得好冷

Robot : 是指惡寒嗎？

User : 是

Robot : 有出汗嗎？

User : 有

Robot : 你同時有頭痛、發熱、惡寒、出汗的症狀, 我判斷你得"桂枝湯證"的機率為95%。

系統架構

問題界定

User : 我頭痛、發燒

Robot : 是體溫38.5度C嗎？

User : 是

Robot : 還有其他症狀嗎？

User : 覺得好冷

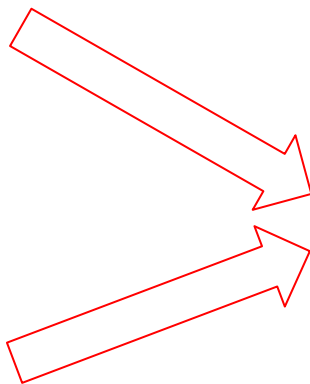
Robot : 是指惡寒嗎？

User : 是

Robot : 有出汗嗎？

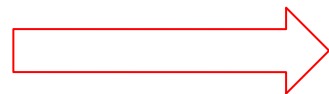
User : 有

Robot : 你同時有頭痛、發熱、惡寒、出汗的症狀，我判斷你得“桂枝湯證”的機率為95%。



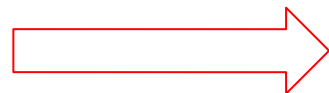
A

症狀詞的確認(同義詞)



B

症狀的蒐集
(1.問診表 2.以可能方劑為核心)

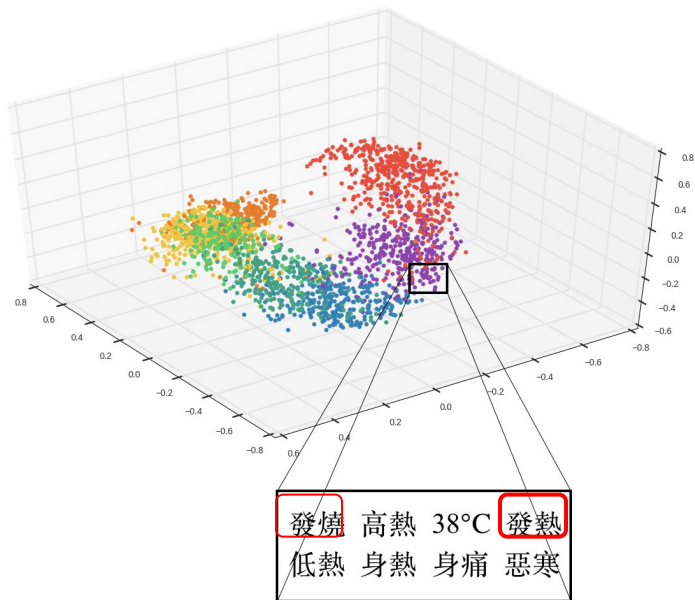
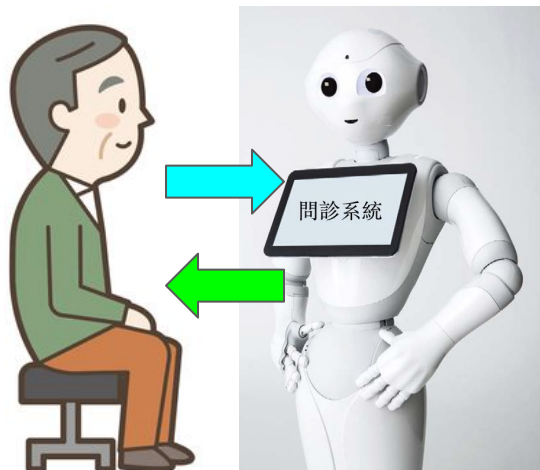


C

症狀與方劑的對應

系統架構

問題界定



問題(任務) A:

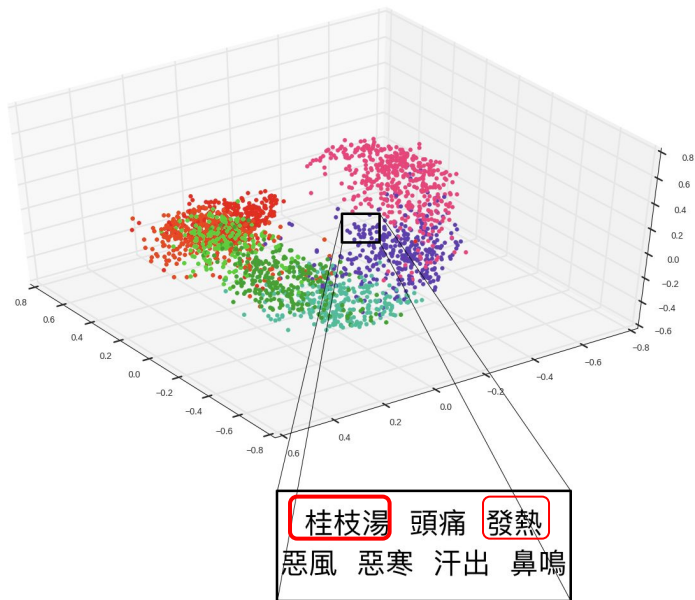
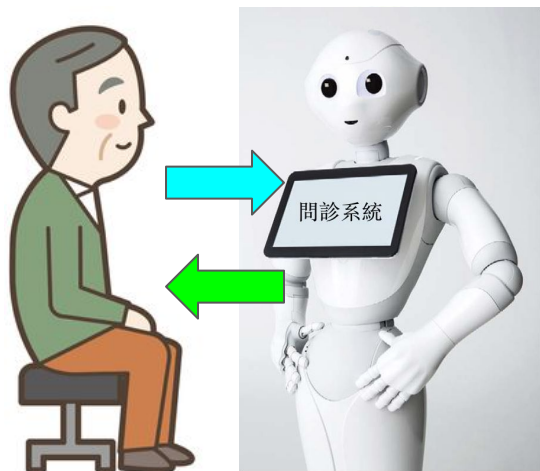
以使用者輸入的症狀詞，尋找最接近的核心詞。

例如：

以"發燒"找到"發熱"

系統架構

問題界定



問題(任務) B:

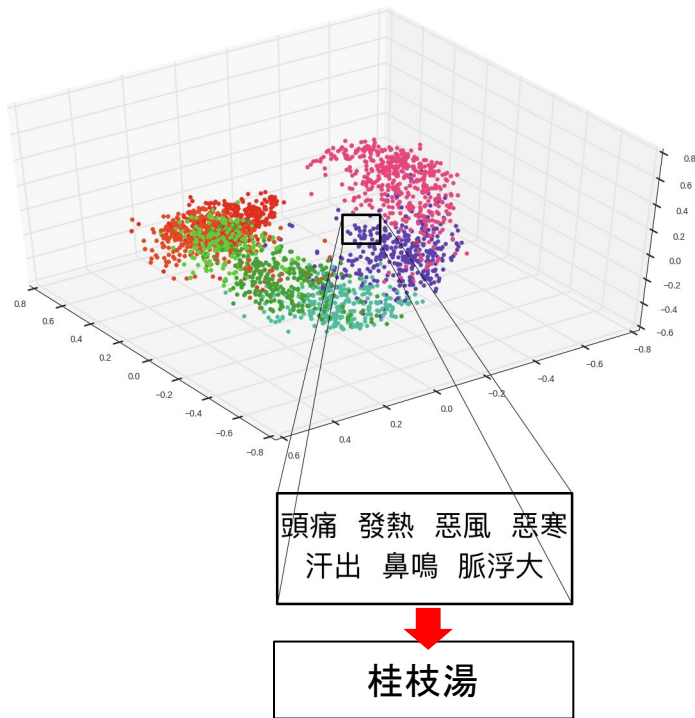
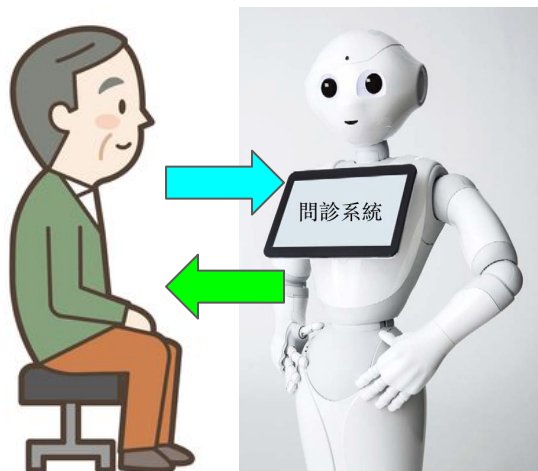
以問題A找到的核心詞，尋找最接近的方劑，再以此抓出相關的症狀。

例如：

以“發熱”找到“桂枝湯”，再以桂枝湯抓出“汗出”等症狀詞。

系統架構

問題界定



問題(任務) C:

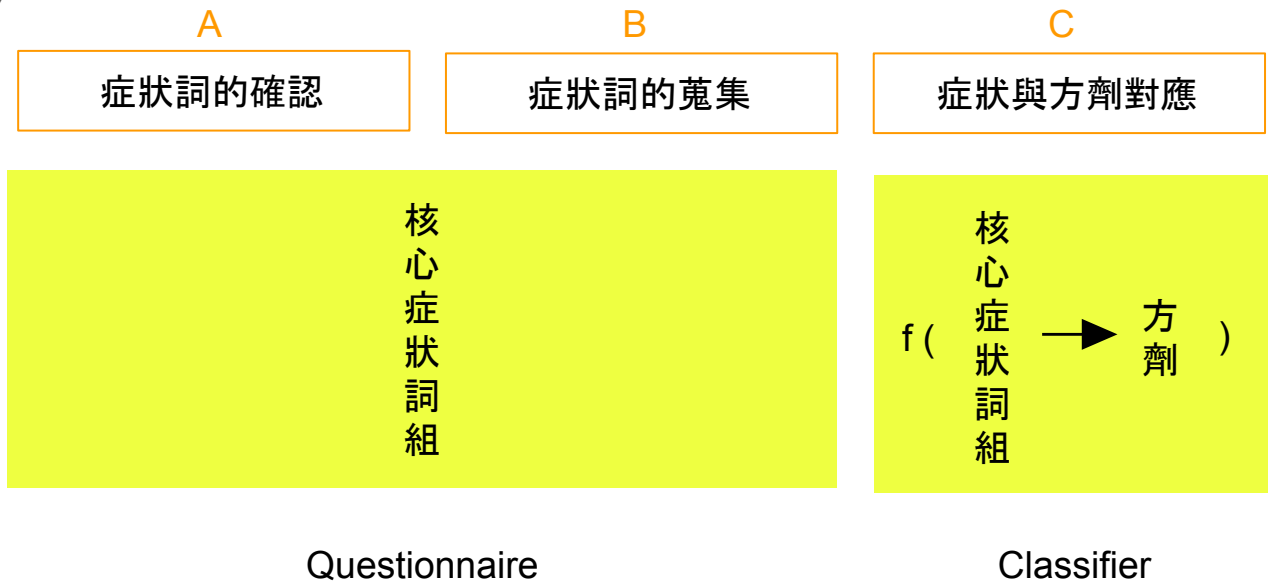
以多個核心詞或症狀詞找出最符合的方劑。

例如:

以 頭痛、發熱、惡寒、出汗
4個症狀詞 找到“桂枝湯”

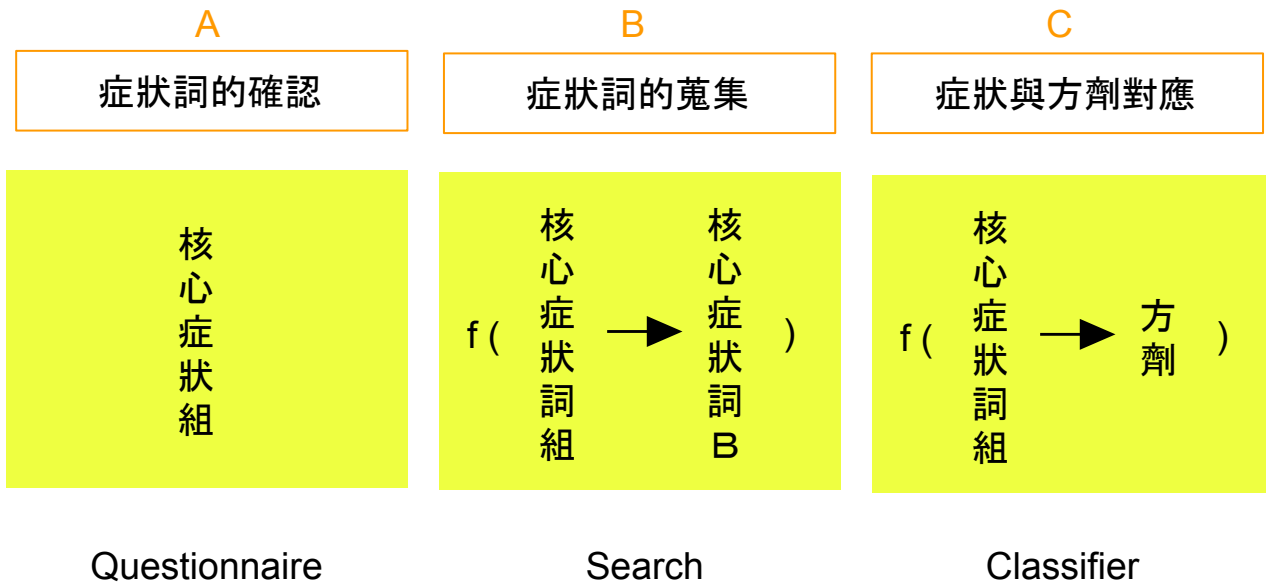
系統架構

解決方式 (I)



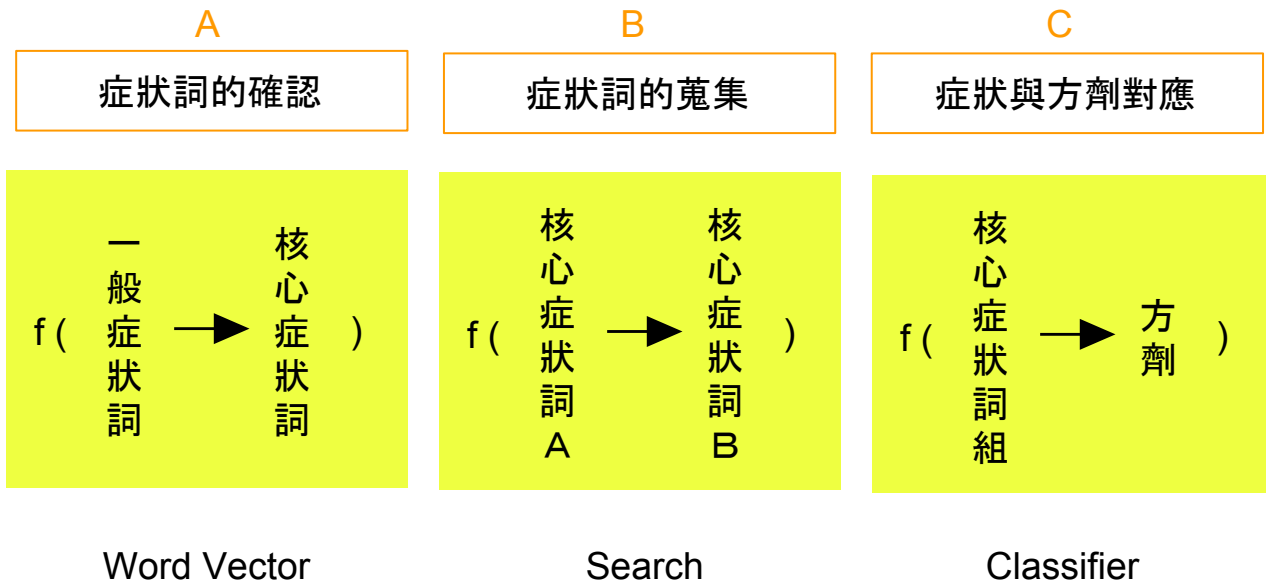
系統架構

解決方式 (二)



系統架構

解決方式 (二)



實驗一：同義詞與相關詞擷取問題

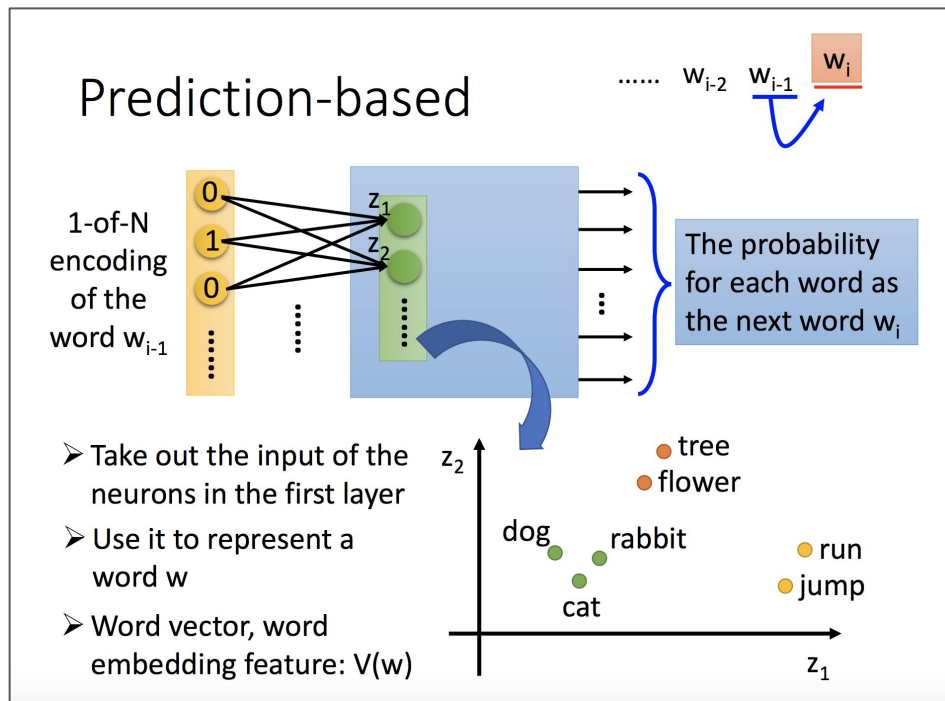
資料集

- 中醫醫案，共5,428筆，共4,231,852字。
- 純文字檔，非結構化資料。
- 字數：
 - min : 48字
 - max : 29,373字
 - mean : 780字
- 例一：
 - 李某，女，67岁。每于夏日洗澡后即寒热不适，自感发热而又需披衣裹护。并于不知不觉中汗透衣服，急又换衣，如此三四次方渐趋平复。年年如此，以至夏日视洗澡为畏途。曾住院治疗不效。中医鉴于盛夏炎热，每以涤暑益气，敛汗固表为治，总不见效。三年前来我处就诊。症如前述，脉虚大而迟，苔薄黄。乃处桂枝汤加黄芪、炮附片，1剂效，2剂愈。后每年夏日发作，均来服药2剂即愈。

實驗一：同義詞與相關詞擷取問題

模型工具

- 演算法: Word Embedding
- 模型: [Google Word2vec](#)
- 套件工具: [gensim](#)
- 實驗
 - 斷詞: 以 jieba 進行斷詞 (1,728,353個詞)
 - 參數:
 - dimension size : 300
 - window : {5,10,20,30,40,50,100}
 - 輸入詞: 发热
 - 輸出詞: 前30名的相似詞



投影片出處

: [http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2017/Lecture/word2vec%20\(v2\).pdf](http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2017/Lecture/word2vec%20(v2).pdf)

實驗一：同義詞與相關詞擷取問題

實驗結果

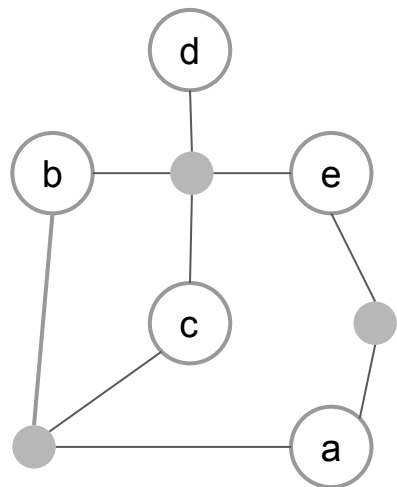
- 以“發熱”預測前三十名之相近詞的結果
- 評測
 - 輸入詞：发热
 - 輸出詞：前30名的相似詞，及其similarity
 - 人工評價：可視為同義詞者標為綠色
- 結論
 - window number 設為20時，擷取**同義詞**的效果最好，比例為23%。
 - window number 設為5時，擷取**相關症狀詞**的效果最好，比例為83%。

10	20	30
高热,0.9314804673194885	高热,0.9218583106994629	高热,0.8631914258003235
恶寒,0.8657320141792297	恶寒,0.8630971908569336	咽痛,0.8139976263046265
寒战,0.8538491725921631	喷嚏,0.8620761632919312	流涕,0.8114215731620789
鼻塞,0.853233277797699	身热,0.8567236065864563	鼻塞,0.809853732585907
流涕,0.8417866230010986	鼻塞,0.8507486581802368	寒战,0.8079014420509338
喷嚏,0.8386059403419495	流涕,0.8480745553970337	疹,0.797214925289154
咽痛,0.8373203873634338	高烧,0.8436099886894226	清涕,0.7952760457992554
无汗,0.8225550651550293	咽痛,0.8378678560256958	身热,0.7887198328971863
身热,0.8210572600364685	肛温,0.8299147486686707	38.0.7833618521690369
感身,0.8054195046424866	无汗,0.825305700302124	青霉素,0.7796722650527954
洗澡,0.7867088913917542	寒战,0.8163182139396667	微咳,0.7761843800544739
高烧,0.7853121161460876	37.3.0.7886863350868225	39.5.0.7758045196533203
身痛,0.7850538492202759	发烧,0.7806981205940247	痒,0.773948609828949
无畏,0.781367301940918	午后,0.7798621654510498	C,0.7733820676803589
脓痰,0.7792244553565979	低热,0.7681023478507996	无汗,0.7729420065879822
今早,0.7780599594116211	37.4.0.7673729062080383	恶寒,0.7678341865539551
天来,0.7641262412071228	疹,0.7636924982070923	体温,0.7667778730392456
汗出,0.7598701119422913	咯血,0.756317138671875	热退,0.7611746788024902
昨日,0.7587429881095886	惊惕,0.7546350955963135	秋燥,0.758543074131012
午后,0.7577895522117615	身痛,0.7519845366477966	低热,0.7567633986473083
发黄,0.7569559216499329	泡沫状,0.7516618967056274	测体温,0.7549898624420166
突然,0.7500507235527039	带血,0.7504773736000061	未退,0.7519897818565369
40.5.0.7489706873893738	C,0.7461422681808472	38.5.0.751026451587677
当风,0.7485036849975586	℃,0.7452738285064697	°,0.7457417249679565
发烧,0.7477428317070007	体温,0.7438117861747742	39.0.7453436255455017
后身,0.7476031184196472	疹,0.7428701519966125	疹色,0.7398219704627991
鼻流,0.7475764751434326	°,0.7418293356895447	多言,0.7355985045433044
低热,0.7470089793205261	今早,0.7410593628883362	喷嚏,0.7350829243659973
腥臭,0.7458845973014832	昨日,0.7408146858215332	疹,0.7306127548217773
阵,0.7439358830451965	咳血,0.7396745681762695	抗生素,0.7303558588027954
20%	23%	23%

實驗二

想法

- 核心症狀詞： $S : \{a,b,c,d,e\}$
- 給了一些條件(同一組): (a,b,c) , (b,c,d,e) , (a,e)
- input是一個或多個 x , $x \in S$
- output是一個 y , $y \in S$
- 且經過的灰點愈少愈好



實驗三：症狀對應方劑之二元分類問題

資料集

- 資料格式

- 中醫醫案(經人工處理), 共2,107筆。
- JSON檔, 半結構化資料。平均字數約 70字。

- 資料預處理

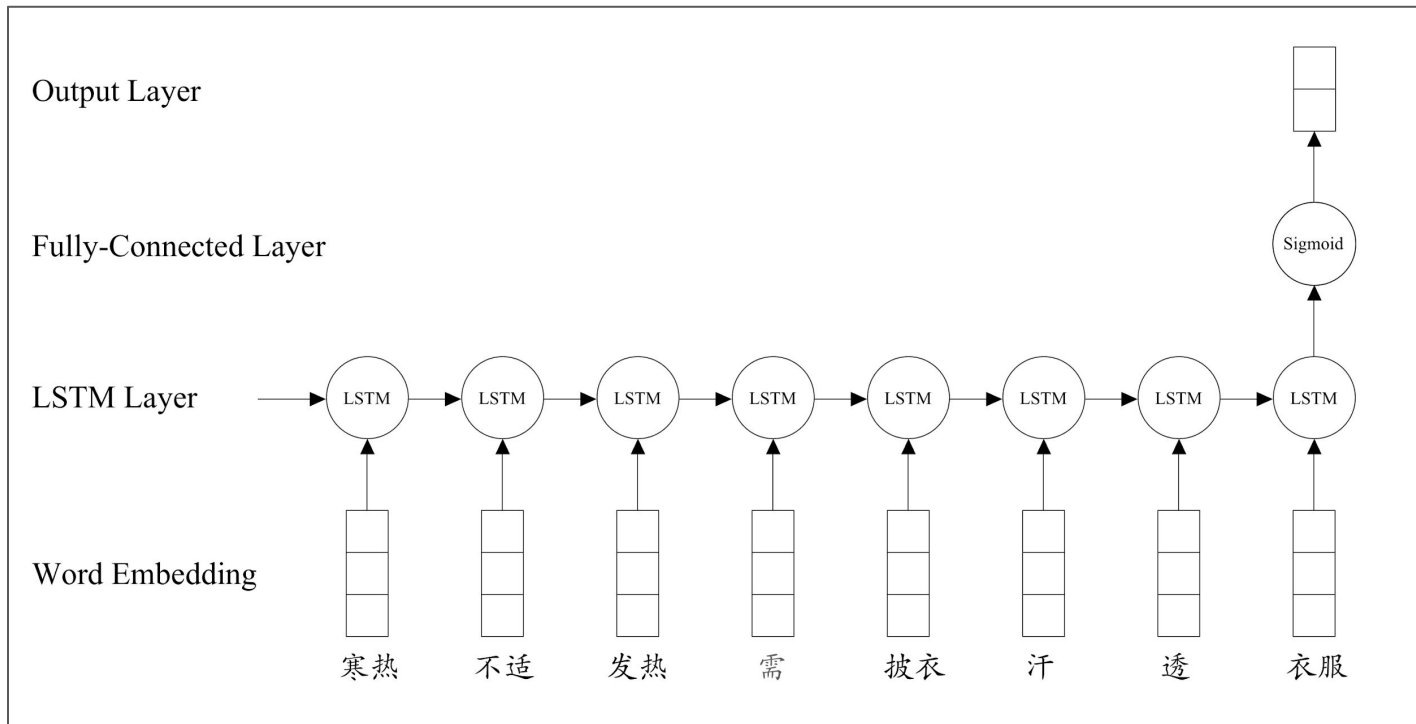
- 合併主證(main)與兼證(second), 以 jieba 斷詞。
- 以實驗一訓練之詞嵌入模型, 取出每個詞的詞向量。
- 以標題(title)比對是否有"桂枝湯"字串, 有則標註為1, 反之為0。
- 取資料100筆(標註為1有39筆, 標註為0有61筆), 約 2 : 3

- 資料實例

- `{'main': '寒热不适, 发热又需披衣, 汗透衣服', 'pulse': '脉虚大而迟', 'recipe': {'raw': '', 'memo': '', 'title': '桂枝汤'}, 'second': '', 'tongue': '苔薄黄', 'chinese': '发热', 'therapy': '', 'western': '', 'dialectical': ''}`

實驗三：症狀對應方劑之二元分類問題

模型架構



實驗三：症狀對應方劑之二元分類問題

模型參數

- neurons : 32
- epochs : 100

實驗結果

- training dataset accuracy : 98%
- testing dataset accuracy : 98%
- 分類結果
 - 20筆測試資料
 - 16筆答對, 正確率 80%

未來工作

詞嵌入模型

- 使用字典訓練出更多同義詞
- 僅使用症狀詞(JSON)訓練同義詞

分類模型

- 使用 bi-RNN
- 使用 Attention Mechanism

簡報完畢

敬請指教