



A method of structured conversion for Chinese medicine prescription data by using Microsoft Excel Macro

ConghuaJi^{1,2*}, PengleiZhou¹, ShanLiu², SijiaHuang¹

¹Department of Information, Zhejiang Provincial Hospital of Traditional Chinese Medicine, Hangzhou, Zhejiang, China

²Center of Clinical Evaluation and Analysis, Zhejiang Provincial Hospital of Traditional Chinese Medicine, Hangzhou, Zhejiang, China

*jchi2005@126.com

Abstract

Background: There are a lot of the research is about the relation of Chinese medicine prescription and individual diagnosis and treatment of disease. In the hospital information system, the Chinese medicine prescription data usually exists in two forms: structured longitudinal data and unstructured text data. In the doctor's advice system, the Chinese medicine prescription exists in structured longitudinal data, and in the electronic medical record system, the Chinese medicine prescription exists in unstructured text data. These two forms are not meet the needs of data analysis.

Objective: This paper introduces a method for translating the Chinese medicine prescription

data into structured horizontal data for data analysis. This method can be used by a general researcher.

Method: Different Microsoft excel macros are compiled according to different data characteristics. Prepare data in EXCEL, and run the corresponding macros, then can achieve the data conversion.

Results: longitudinal structured data and text data can be quickly converted to horizontal structured data.

Conclusion: Microsoft Excel is a commonly office software tool for general researchers, who have low barriers to use it. This method is worth popularizing.

Keywords: Traditional Chinese medicine; data mine; Microsoft Excel Macro; data conversion

1.Introduction

Traditional Chinese medicine (TCM) has been gradually developed from long-term clinical practices. As an alternative to modern western medicine, TCM is receiving increasingly attention worldwide. The treatment methods of TCM need to be adjusted according to the individual characteristics of patients. These treatment methods include Chinese herbal medicine, massage, acupuncture and other non-drug treatment[1-28]。 Among them, most of the research is about the relation of Chinese medicine prescription and individual diagnosis and treatment of disease, it involves many diseases such as Coronary Heart Disease[1], pediatric atopic dermatitis[2], cough[3], tic disorder [4], acne[5], cancer[6], melancholia[8], hyperlipidemia[10], knee osteoarthritis[12], primary dysmenorrhea[13], urticaria[14-15], acquired immunodeficiency syndrome[16], breast cancer[18], respiratory disease[19], menopausal syndrome[21], diabetes[22], influenza[23], climacteric women[24], and etc. At the same time, there are many researches are about ideas and techniques for Chinese herbal medicine [7,9,11,17,20,25].

We can use some software tools, such as SPSS Clementine [26] and weka [27], to data mining the rule of using Chinese medicine prescription. Data mining methods include decision tree,

association rule, cluster analysis, discriminant analysis, neural network analysis, and etc. However, if we are to use these methods for analysis, the raw data must be horizontal structured data. Obviously, in general, this condition is not available. If our raw data comes from the hospital information system, the data format should be in vertical structured format, and if our raw data comes from the doctor's writing, the data format should be in text format, so data format conversion is necessary. How to make a general researcher can easily carry out data conversion, without mastering professional information technology? We look at the Microsoft Excel Macro.

A macro is a series of commands and functions stored in the visual basic module, and can be run at any time when you need to perform the task. The function is prewritten formula, to a value of one or more operations are performed, and returns a value of one or more. Functions simplify and shorten formulas in worksheets, especially when they are executed by formulas that are long or complex. The module is stored together as a named unit statement, and statement collection process. There are two types of modules: standard modules and class modules. Visual basic developed by Microsoft company, for the preparation of windows based applications. "Visual basic editor" is a design for beginners to write and edit macro code, but also provides a lot of help. Don't have to learn how to program or how to use Visual Basic language to carry out a simple modification of the macro. Using "visual basic editor", you can edit macros, copy macros in the module copy, macros in a different workbook, rename stored macro modules or rename macros.

This paper introduces a method of structured conversion for Chinese medicine prescription data by using Microsoft Excel Macro, which can be used by a general researcher.

2. Materials and methods

In the hospital information system, the Chinese medicine prescription data usually exists in two ways. In the doctor's advice system, the Chinese medicine prescription exists in structured longitudinal data, and in the electronic medical record system, the Chinese medicine prescription exists in unstructured text data. These two ways are not meet the needs of data analysis. So now we are going to explore how to translate them into horizontal structured data.

2.1 The longitudinal data into horizontal data

2.1.1 Materials

If the data comes from doctor's advice system, the Chinese medicine prescription exists in structured longitudinal data, include "ID", "herb", "dosage(g)"(shown as table1). Data with the same ID represents the same patient. So, the information that table 1 shows is about the Chinese medicine prescription for three patients with ID of 1001, 1002 and 1003.

Table1 raw data of Chinese medicine prescription (longitudinal data)

ID	Herb	dosage(g)
1001	chaihu	10
1001	xiangfu	10
1001	biejia	15
1001	sanleng	10
1001	Eshu	10
1001	shengdi	15
1001	shudihuang	15
1001	shanyao	20
1001	longdancao	8
1001	haijinsha	15
1001	jinqiancao	30
1001	huangqi	30
1001	muli	24
1001	danshen	15
1002	chaihu	10
1002	xiangfu	10
1002	yujin	10
1002	qingpi	10

ID	Herb	dosage(g)
1002	cangshu	12
1002	zhizhi	10
1002	jineijin	15
1002	liushenqu	12
1002	yuanzhi	10
1002	changpu	15
1002	longdancao	9
1002	yiwiren	20
1002	huanglian	10
1002	hehuanpi	10
1002	suanzaoren	20
1003	shengdi	30
1003	shanyurou	10
1003	shanyao	20
1003	danpi	12
1003	yunling	20
1003	xudanpian	15
1003	chaihu	10
1003	longdancao	10
1003	huangqi	30
1003	baimaogen	24
1003	tufuling	10
1003	huangbai	12
1003	danggui	10

2.1.2 Methods

Here, the purpose of the Excel Macro is to translate structured vertical data into structured horizontal data. For detailed methods and principles, please see the macros and annotations (table 2).

Table2 the macros and annotations for longitudinal data into horizontal data

macros	*annotations
<pre> Sub convert() Dim rng As Range On Error Resume Next Set rng = Application.InputBox("Please choose the convert area", "Identify the convert area", ActiveWindow.RangeSelection.Address(0, 0), , , , 8) If Err <> 0 Then Exit Sub If rng.Columns.Count <> 3 Then MsgBox "Only 3 columns of data area are supported.", vbOKOnly, "Friendly reminder": Exit Sub Set rng = Intersect(rng, ActiveSheet.UsedRange) If rng Is Nothing Then MsgBox "Do not select a blank area!", vbOKOnly, "Friendly reminder": Exit Sub If WorksheetFunction.CountA(rng.Columns(1)) = 0 Then MsgBox "Program can't continue because the first column of selected area is blank", vbOKOnly, "Friendly reminder": Exit Sub If WorksheetFunction.CountA(rng.Columns(2)) = 0 Then MsgBox "Program can't continue because the second column of selected area is blank", vbOKOnly, "Friendly reminder": Exit Sub If WorksheetFunction.CountA(rng.Columns(3)) = 0 Then MsgBox </pre>	<pre> *annotations *Select the data to be converted </pre>

"Program can't continue because the third column of selected area is blank", vbOKOnly, "Friendly reminder": Exit Sub

Dim targetrng As Range

Set targetrng = Application.InputBox("Select the storage area of the two-dimensional table,select a single cell", "target area", , , , , 8)

targetrng(1) = rng(1, 1).Value

Dim arr, i, j As Integer, dic1, dic2, rowheader As Range, columnheader As Range

arr = Intersect(rng, rng.Offset(1, 0)).Value

Set dic1 = CreateObject("scripting.dictionary")

With dic1

For i = 1 To UBound(arr, 1)

If Len(arr(i, 1)) > 0 Then .Item(arr(i, 1)) = arr(i, 1)

Next I

Set rowheader = targetrng(2, 1).Resize(.Count, 1)

rowheader = WorksheetFunction.Transpose(dic1.keys)

End With

Set dic2 = CreateObject("scripting.dictionary")

With dic2

For j = 1 To UBound(arr, 1)

If Len(arr(j, 2)) > 0 Then .Item(arr(j, 2)) = arr(j, 2)

*Pop up the input box to allow the user to specify the storage area for the converted worksheet

*Using dictionary objects to determine line labels and column labels

<pre> Next j Set columnheader = targetrng(1, 2).Resize(1, .Count) columnheader = dic2.keys End With With targetrng(2, 2).Resize(dic1.Count, dic2.Count) .FormulaArray = "=INDEX(" & rng.Columns(3).Address & ",MATCH(" & rowheader.Address & _ "&""@""&" & columnheader.Address & ", " & rng.Columns(1).Address & "&""@""&" & rng.Columns(2).Address & ",0))" .Value = .Value .CurrentRegion.Borders.LineStyle = xlContinuous .CurrentRegion.EntireColumn.AutoFit End With Cells.Replace What:="#N/A", Replacement:="" End Sub </pre>	<p>*Referencing an area other than the column heading and entering an array formula in the data region, referencing the values of the third column in the array, with a reference to the corresponding relation to the row header and column heading</p> <p>*add border</p> <p>*automatically adjust the column width</p>
--	---

2.2 Text data into structured data

2.2.1 Materials

If the data comes from an electronic medical record system, the data type is text prescription. If the test samples (same as table 1) come from our hospital electronic medical record system, and we extracted 3 Chinese medicine prescriptions data as the data format conversion samples(table3). Now we can find the Chinese medicine prescription data includes the names of herbs, quantities, and units of quantity. All the information about herbs are stored in a same field, and distinguish by commas.

Table3 raw data of Chinese medicine prescription (text data)

ID	The detail of Chinese medicine prescription
1001	chaihu10g,xiangfu10g,biejia15g,sanleng10g,eshu10g,shengdi15g,shudihuang15g,shanyao20g,longdancao8g,haijinsha15g,jinqiancao30g,huangqi30g,muli24g,danshen15g,
1002	chaihu10g,xiangfu10g,yujin10g,qingpi10g,cangshu12g,zhizi10g,jineijin15g,liushenqu12g,yuanzhi10g,changpu15g,longdancao9g,yi'yi'ren20g,huanglian10g,hehuanpi10g,suanzaoren20g,
1003	shengdi30g,shanyurou10g,shanyao20g,danpi12g,yunling20g,xudianpian15g,chaihu10g,longdancao10g,huangqi30g,baimaogen24g,tufuling10g,huangbai12g,danggui10g,

2.2.2 methods

Here, our main objective is to split text data into different grids and to guarantee horizontal data formats. For detailed methods and principles, please see the macros and annotations (table 4).

Table4 the macros and annotations for Text data into structured data

Macros	*annotations
Public Sub Queue_conversion ()	
	*The result is stored in the processing result

<pre> Selection.Replace What:="g", Replacement:="," Selection.TextToColumns Destination:=Range("B1"), DataType:=xlDelimited, _ TextQualifier:=xlDoubleQuote, Other:=True, OtherChar:="," Dim arr, arrjg, i, j, m, n, k As Integer, dic2, columnheader As Range Dim cell As Range, str As String, zm As String, sz As String Dim rng As Range Set rng = ActiveSheet.UsedRange arr = rng.Value Set dic2 = CreateObject("scripting.dictionary") Set rng = Intersect(rng, rng.Offset(1, 1)) m = rng.Rows.Count </pre>	<p>same time</p> <p>*Extract data from sheet and assign the data in sheet to RNG</p> <p>*After removing the header line from the area represented by the variable RNG, assign it to the variable ARR, where the ARR becomes an array</p> <p>*Set 'dic2' as dictionary object</p> <p>*removes the title row and header column from the area represented by the variable RNG</p> <p>*m and n store the number of rows and columns in the RNG region</p>
--	---

<pre> n = rng.Columns.Count ReDim arrzm(1 To m, 1 To n) ReDim arrsz(1 To m, 1 To n) k = 1 For Each cell In rng If Len(cell) > 0 Then For j = 1 To Len(cell.Text) str = Mid\$(cell.Text, j, 1) If str Like "[a-z,A-Z]" Then zm = zm & str If str Like "[0-9.]" Then sz = sz & str Next j End If arrzm(Int((k - 1) / n) + 1, (k - 1) Mod n + 1) = zm arrsz(Int((k - 1) / n) + 1, (k - 1) Mod n + 1) = Replace(WorksheetFunction.Trim(Replace(sz, ".", "")), "", ".") zm = "": sz = "" If Len(arrzm(Int((k - 1) / n) + 1, (k - 1) Mod n + 1)) > 0 Then dic2.Item(arrzm(Int((k - 1) / n) + 1, (k - 1) Mod </pre>	<p>*Resets an array based on row number and column number</p> <p>*k is used to record the number of times traversed. The parameter RNG represents the region of the data. The extraction of the word and numbers in the data table is the name of the drug and the number is the dose</p>
--	---

<pre> n + 1)) = arrzm(Int((k - 1) / n) + 1, (k - 1) Mod n + 1) End If k = k + 1 Next cell ReDim col(1 To 1, 1 To dic2.Count) col = dic2.keys ReDim arrjg(1 To m, 1 To dic2.Count) For i = 1 To m For j = 1 To n For k = 1 To dic2.Count If arr(i + 1, 1) & arrzm(i, j) = arr(i + 1, 1) & col(k - 1) Then arrjg(i, k) = arrsz(i, j) Next k Next j Next i ActiveSheet.UsedRange.Select Selection.ClearContents </pre>	<p>*Use dictionary objects to take unique column labels and store the non empty word as column labels in dic2.</p> <p>*reset array</p> <p>*store the data in the original data table to the corresponding location</p> <p>*clear the old data</p>
--	---

<pre> Dim targetrng As Range Set targetrng = Range("a1") Range("a1").Resize(UBound(arr, 1), 1) = Application.Index(arr, , 1) Set columnheader = targetrng(1, 2).Resize(1, dic2.Count) columnheader = dic2.keys With targetrng(2, 2).Resize(m, dic2.Count) .Value = arrjg .CurrentRegion.EntireColumn.AutoFit End With End Sub </pre>	<pre> *The location where the settings result is stored *fill in the destination cell with the row tag *Fill the column tag into the target cell *fill data into the target cell </pre>
---	--

3. Results

3.1 The result of longitudinal data into horizontal data

Through the macro program operation (shown in table 2), the structured but longitudinal data of Chinese medicine prescription (shown in Figure 1 as a sample), is converted into completely structured and horizontal data of Chinese medicine prescription (shown in Figure 2-1, Figure 2-2, Figure 2-3).

	A	B	C
1	ID	herb	dosage (g)
2	1001	chaihu	10
3	1001	xiangfu	10
4	1001	biejia	15
5	1001	sanleng	10
6	1001	eshu	10
7	1001	shengdi	15
8	1001	shudihuang	15
9	1001	shanyao	20
10	1001	longdancao	8
11	1001	haijinsha	15
12	1001	jinqiancao	30
13	1001	huangqi	30
14	1001	muli	24
15	1001	danshen	15
16	1002	chaihu	10
17	1002	xiangfu	10
18	1002	yujin	10
19	1002	qingpi	10
20	1002	cangshu	12
21	1002	zhizhi	10
22	1002	jineijin	15
23	1002	liushenqu	12
24	1002	yuanzhi	10
25	1002	changpu	15
26	1002	longdancao	9
27	1002	yiwiren	20
28	1002	huanglian	10
29	1002	hehuanpi	10
30	1002	suanzaoren	20
31	1003	shengdi	30
32	1003	shanyurou	10
33	1003	shanyao	20
34	1003	danpi	12
35	1003	yunling	20
36	1003	xudanpian	15
37	1003	chaihu	10
38	1003	longdancao	10
39	1003	huangqi	30
40	1003	baimaogen	24
41	1003	tufuling	10
42	1003	huangbai	12
43	1003	danggui	10
44			

Figure1 The raw data of Chinese Medicine prescription(longitudinal data)

E	F	G	H	I	J	K	L	M	N	O	P
ID	chaihu	xiangfu	biejia	sanleng	eshu	shengdi	shudihuang	shanyao	longdancao	haijinsha	jinqiancao
1001	10	10	15	10	10	15	15	20	8	15	30
1002	10	10							9		
1003	10					30		20	10		

Figure 2-1 The data results after running the macro program

Q	R	S	T	U	V	W	X	Y	Z	AA	AB
huangqi	muli	danshen	yujin	qingpi	cangshu	zhizhi	jineijin	liushenqu	yuanzhi	changpu	yiwiren
30	24	15	10	10	12	10	15	12	10	15	20
30											

Figure 2-2 The data results after running the macro program (continue)

AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
changpu	yiwiren	huanglian	hehuanpi	suanzaoren	shanyurou	danpi	yunling	xudanpian	baimaogen	tufuling	huangbai	danggui
15	20	10	10	20		10	12	20	15	24	10	12

Figure 2-3 The data results after running the macro program (continue)

3.2 the result of text data into structured data

Through the macro program operation (shown in table 4), the unstructured text data of Chinese medicine prescription (shown in Figure 3 as a sample, there are 41 Chinese medicine prescriptions), is converted into completely structured and horizontal data of Chinese medicine prescription (shown in Figure 4).

File Home Insert Draw Page Layout Formulas Data Review View Tell me what you want to do

B1 The prescription of traditional Chinese medicine

ID	The prescription of traditional Chinese medicine															
1	chahu10g	xiangfu10g	biejia15g	sanleng10g	eshu10g	shengdi15g	shudihuang15g	shanyao20g	longdancao8g	haijinsha15g	jinjiancao30g	huangqi30g	muli24g	danshen15g		
2	chahu10g	xiangfu10g	yujin10g	qingli10g	cangshu12g	zhizi10g	jinejin15g	liushenqu12g	yuanzhi10g	changpu15g	longdancao9g	yi'yiren20g	huanglian10g	hehuanni10g	suanzaoren20g	
3	shengdi30g	shanyurou10g	shanyao20g	danni12g	yunling20g	xuduopian15g	chahu10g	longdancao10g	huangqi30g	baimaogen24g	tufuling10g	huangbai12g	danggui10g			
4	chahu10g	xiangfu10g	huangqi36g	danggui10g	chuanxiong15g	zhen12g	xicao12g	honghua8g	lujun15g	shao12g	taizishen10g	shanyao20g	baishu12g			
5	chahu10g	xiangfu10g	shanyao20g	gancacao3g	haizao3g	kunbu8g	jiangcan6g	huangqi12g	baishu6g	chuanjianzi4g	luxiaoma10g					
6	huangq36	shengdi15g	shudihuang15g	chahu10g	chuanjianzi10g	xiangfu10g	houpu6g	peilan15g	huomaren24g	fupenzi12g	banbianlian30g	baihuashe20g	sugeng10g	maidong10g	xiyangshen4g	
7	chahu10g	xiangfu10g	biejia10g	jixueteng15g	sanleng10g	eshu10g	baishu10g	huangqi30g	shengdi15g	shudihuang15g	yuanshen18g	yinchen10g	fuling15g	shanyao20g	jigucac30g	
8	chahu10g	xiangfu10g	biejia10g	jixueteng15g	sanleng10g	eshu10g	baishu10g	huangqi30g	shengdi15g	shudihuang15g	yuanshen18g	yinchen10g	fuling15g	shanyao20g		
9	chahu10g	chuanxiong15g	qinjui10g	baizhi10g	jiangcan6g	danshen15g	chishao12g	baishao12g	fuling15g	shanyao20g	baishu12g	guya12g	jinejin15g	quanxie5g		
10	chahu10g	xiangfu10g	yujin10g	baizhen10g	shudihuang30g	chenpi6g	buguzhi12g	gusuibu12g	huangqi36g	fangfeng10g	baishu12g	yi'yiren20g	cangshu10g	sangjisheng15g	chuanxiong15g	
11	chahu10g	xiangfu10g	chishao12g	baishao12g	jixueteng15g	jigucac30g	huangqi36g	shengdi15g	shudihuang15g	yuanshen18g	biejia10g	shanyao20g	baishu12g	fuling15g	longdancao8g	
12	qinjui12g	danggui10g	muli30g	longchi24g	taoren10g	xingren10g	yuanzhi10g	changpu15g	qingdi10g	chenpi6g	fabanxia6g	jixueteng15g	jigucac30g	xuduopian15g	suanzaoren20g	
13	chahu10g	danggui10g	chishao12g	chuanxiong15g	lulutong10g	lujun15g	zhenlan12g	huangqi36g	shudihuang30g	chenpi6g	shanyao20g	difuzi12g	digupi10g	jiangcan6g		
14	chahu10g	xiangfu10g	longdancao10g	huangqi36g	danggui10g	yunling15g	shanyao20g	baishu12g	danshen18g	jinejin15g	suanzaoren20g	jiangcan6g	difuzi12g	chuanxiong15g	baizhi10g	
15	huangqi30g	shengdi15g	shudihuang15g	yuanshen15g	gegen18g	shanyao20g	shihu24g	danshen18g	chahu10g	xiangfu12g	yunling15g	guizhi6g	chishao12g	baishao12g	danggui10g	
16	chahu10g	xiangfu10g	biejia15g	sanleng10g	eshu10g	danggui10g	shengdi15g	shudihuang15g	danshen15g	yunling15g	shanyao20g	yuanshen15g	huanglian10g	chuanxiong15g	baizhi10g	
17	chahu10g	xiangfu10g	chuanjianzi10g	danshen15g	jixueteng15g	chishao12g	baishao12g	shudihuang30g	chenpi6g	shanyao20g	xuduopian15g	zhishi6g	baishu12g	suanzaoren20g	fuzhi6g	
18	chahu10g	danggui10g	chuanxiong15g	qinjui10g	baizhi10g	yunling15g	cangshu12g	baishu12g	lianzi12g	fuzhi10g	zhenlan15g	rougui5g				
19	chahu10g	xiangfu10g	baishao15g	shudihuang30g	chenpi6g	yuanshen15g	nvzhenzi12g	changpu12g	yuanzhi10g	danshen18g	maidong10g	hehuanni10g	suanzaoren20g			
20	chahu10g	danggui10g	shengdi15g	shudihuang15g	shanyao20g	baishu12g	yunling15g	gualou10g	muxiang6g	danshen15g	qingjiao10g	qingxiang10g	chishao12g	guizhi5g		
21	shudihuang30g	changpu6g	shayuanzi15g	tusizi12g	gouqizi12g	wuweizi10g	roucongrong15g	yinyanghuo12g	xianmao10g	yangqishi15g	huangqin10g	xinyihua6g	xuduopian15g	baishu15g	suanzaoren20g	
22	chahu10g	danggui12g	chuanxiong15g	qinjui10g	baizhi10g	danshen18g	sugeng10g	muxiang6g	fupenzi12g	huomaren24g	zaojiaoci15g	jinjiancao30g	shihu20g	chejiancao15g	chejianzi15g	
23	chahu10g	xiangfu10g	shudihuang30g	chenpi6g	suoyang12g	roucongrong15g	shayuanzi15g	yinyanghuo12g	shanyao20g	baishu12g	huangqi36g	huainiuxi12g	suanzaoren20g	jinejin15g	gualou10g	muxiang6g
24	shengdi30g	shanyurou10g	shanyao20g	fuling20g	danni12g	chejianzi15g	chejiancao15g	huangqi36g	danggui10g	yuanshen15g	jindenglong10g	gancacao6g	huomaren20g	baimaogen24g	tufuling10g	
25	huangqi36g	danggui10g	baishao12g	baishu12g	huangqin10g	ajiao10g	shenaiye10g	zhongjitan10g	shanyao20g	yunling15g	shengdi24g	chenpi6g	gancacao6g			
26	chahu10g	xiangfu10g	jixueteng15g	jigucac30g	huangqi36g	shengdi15g	shudihuang15g	shanyao20g	gegen18g	danshen18g	sangjisheng15g	roucongrong15g	huanglian15g			
27	chahu10g	xiangfu10g	yuanshen15g	chuanjianzi10g	qingpi6g	kunbu10g	muxiang6g	rougui5g	shanyao20g	cangshu10g	baishu10g	huangqi30g	yi'yiren20g	luxiaoma20g	gualou10g	
28	chahu10g	danggui10g	xiangfu10g	shengdi30g	yuanshen15g	shayuanzi12g	xuduopian15g	roucongrong15g	gouqizi12g	huainiuxi10g	huanglian10g					
29	chahu10g	danggui10g	xiangfu10g	huangqin10g	huanglian10g	peilan15g	danshen18g	huangqi30g	fangfeng10g	baishu12g	shanyao20g	suanzaoren20g	sangjisheng15g	baihuashe24g		
30	chahu10g	xiangfu10g	shengdi15g	shudihuang15g	sanleng10g	eshu10g	cangshu12g	fuling20g	yi'yiren20g	chejiancao15g	chejianzi15g	suanzaoren20g	jixueteng15g	jinejin15g		
31	chahu10g	xiangfu10g	shengdi15g	jixueteng15g	jigucac30g	shanyao20g	yinchen10g	huangqi30g	shengdi15g	shudihuang15g	yuanshen15g	chuanjianzi10g	gouqizi10g	danshen15g		
32	xiyangshen4g	maidong10g	wuweizi10g	gancacao8g	baizhen10g	chahu10g	jixueteng15g	danshen18g	houpu5g	zhishi6g	huomaren20g	suanzaoren20g	huangqi24g			
33	chahu10g	xiangfu10g	yujin10g	longdancao8g	sugeng10g	maidong10g	muxiang6g	qinjui10g	huangqi30g	jixueteng15g	houpu6g	yi'yiren20g	cangshu10g			
34	chahu6g	xiangfu6g	huangqin10g	baishu12g	shanyao20g	danggui10g	baishao12g	jixueteng15g	shudihuang24g	chenpi6g	maidong10g					
35	huangq100g	dangui40g	shanyao60g	yunling50g	baishu40g	chahu36g	danshen60g	xiangfu40g	gancacao30g	jinejin50g						
36	huangq36g	shengdi15g	shudihuang15g	shanyao20g	cangshu10g	yuanshen15g	gegen18g	taizishen10g	maidong10g	danshen15g	wuweizi12g	baizhen10g	chahu10g	xiangfu10g	suanzaoren20g	
37	chahu10g	xiangfu10g	danggui10g	shengdi15g	shudihuang15g	yuanshen15g	nvzhenzi12g	qinjui10g	chuanxiong15g	baizhi10g	zhishi6g	baishu10g	jixueteng15g			
38	huangqi30g	shengdi15g	shudihuang15g	danggui10g	chahu10g	chuanjianzi10g	qingpi6g	shanyao20g	jixueteng15g	jigucac24g	xiyangshen4g	huangqin12g	banbianlian20g	baihuashe24g		
39	chahu10g	xiangfu10g	jixueteng15g	jigucac30g	shanyao20g	baishu12g	danshen18g	huangqi36g	shihu24g	zaojiaoci12g	shengdi15g	shudihuang15g	jiangcan6g	difuzi12g		
40	shengdi15g	shudihuang15g	shanyurou10g	danni12g	baimaogen30g	huangqi30g	tufuling10g	danshen18g	huanglian10g	chahu8g	xiangfu8g	shanyao20g				
41	chahu10g	xiangfu10g	longdancao10g	yuanshen15g	chuanjianzi10g	jiangcan6g	jieregeng10g	jindenglong10g	danggui10g	zaojiaoci10g	haizao12g					
42	41															

Figure 3 The raw data of Chinese Medicine prescription (text data)

File Home Insert Draw Page Layout Formulas											Data					
ID	A	B	C	D	E	F	G	H	sh	DF	DG	DH	DI	DJ	DK	
ID	chahu	xiangfu	biejia	sanleng	eshu	shengdi	shudihuang	sh		yang	huainiuxi	jindenglong	ajiao10g	shenaiye	zhongjitan	
1	10	10	15	10	10	15	15	15		0	0	0	0	0	0	
2	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
3	10	10	0	0	0	0	30	0		0	0	0	0	0	0	
4	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
5	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
6	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
7	10	10	0	0	0	0	15	15		0	0	0	0	0	0	
8	10	10	10	10	10	10	15	15		0	0	0	0	0	0	
9	10	10	10	10	10	10	15	15		0	0	0	0	0	0	
10	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
11	10	10	0	0	0	0	0	30		0	0	0	0	0	0	
12	10	10	10	10	0	0	15	15		0	0	0	0	0	0	
13	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
14	10	10	0	0	0	0	0	30		0	0	0	0	0	0	
15	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
16	10	10	12	0	0	0	15	15		0	0	0	0	0	0	
17	10	10	10	15	10	10	15	15		0	0	0	0	0	0	
18	10	10	10	0	0	0	0	30		0	0	0	0	0	0	
19	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
20	10	10	0	0	0	0	0	30		0	0	0	0	0	0	
21	10	10	0	0	0	0	15	15		0	0	0	0	0	0	
22	10	10	0	0	0	0	0	30		0	0	0	0	0	0	
23	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
24	10	10	10	0	0	0	0	30		12	12	0	0	0	0	
25	10	10	0	0	0	0	0	30		0	0	10	0	0	0	
26	10	10	0	0	0	0	24	0		0	0	0	10	10	10	
27	10	10	0	0	0	0	15	15		0	0	0	0	0	0	
28	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
29	10	10	0	0	0	0	30	0		0	10	0	0	0	0	
30	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
31	10	10	0	0	10	12	15	15		0	0	0	0	0	0	
32	10	10	0	0	0	0	15	15		0	0	0	0	0	0	
33	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
34	10	10	0	0	0	0	0	0		0	0	0	0	0	0	
35	6	6	0	0	0	0	0	24		0	0	0	0	0	0	
36	36	40	0	0	0	0	0	0		0	0	0	0	0	0	
37	10	10	0	0	0	0	15	15		0	0	0	0	0	0	
38	10	10	0	0	0	0	15	15		0	0	0	0	0	0	
39	10	10	0	0	0	0	15	15		0	0	0	0	0	0	
40	8	8	0	0	0	0	15	15		0	0	0	0	0	0	
41	10	10	0	0	0	0	0	0		0	0	10	0	0	0	
42	10	10	0	0	0	0	0	0		0	0	0	0	0	0	

Figure 4 The data results after running the macro program

***Because of space reasons, here omitted the results of the I to DE column**

4. Discussion

The data mining of TCM has a very good prospect, such as name normalization [28], establish research database[29-30], aid decision[31], analysis on acupoint selection rule[32-36] and some other method and application[37-42]. Overall, the research of TCM data mining mainly focus on two aspects: the clinical interpretation and analysis methods. Data preparation and cleanup usually comes from large databases, which are done by professional technicians. Data preprocessing is the beginning of data mining. The preprocessing should include the format transformation and data content transformation on the basis of information preserving. Few data preprocessing methods for general researchers are introduced. Because there are a lot of Chinese medical records, these valuable resources need to be used effectively, the appropriate method for using these resources is important particularly.

There are more advanced tool such as Python with pandas, or Weka to replace Excel for the data format transfer. Microsoft Excel is a commonly office software tool for general researchers, and they have low barriers to use. The researchers copy the macro code wholly intact to their Excel, and run, can achieve data conversion. When the researcher have mastered this self-service technology, the research space of TCM data mining will be greater, include unstructured prescription data, symptoms, acupuncture, acupuncture and other data analysis. Therefore, this method is worth popularizing.

Conflicts of interest

All authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by grants from the State Administration of Traditional Chinese Medicine of the People's Republic of China (No.JDZX2012172), the Administration of Traditional Chinese Medicine of Zhejiang Province (No.2014ZZ003).

REFERENCES

- [1] Ren Y, Chen ZQ, Zhang MZ, Guo LH, He DY. [Cluster Analysis of Medication Laws for Treating Coronary Heart Disease by Distinguished Veteran Doctors of Traditional Chinese Medicine]. *Zhongguo Zhong Xi Yi Jie He Za Zhi*. 2016 Apr;36(4):411-4. Chinese.
- [2] Chen YC, Lin YH, Hu S, Chen HY. Characteristics of traditional Chinese medicine users and prescription analysis for pediatric atopic dermatitis: a population-based study. *BMC Complement Altern Med*. 2016 Jun 8;16:173. doi:10.1186/s12906-016-1158-1.
- [3] Xu XJ, Xu LM, Shen CF, Wang CH, Shen CT. [Analysis on medication principles for cough based on experience of Xu Di-hua, descendant of Meng He Medical School]. *Zhongguo Zhong Yao Za Zhi*. 2015 Nov;40(21):4301-5. Chinese.
- [4] Sun LY, Li QP, Zhao LL, Ding YQ. [Traditional Chinese medicine inheritance system analysis of professor Ding Yuanqing in treating tic disorder medication based on experience]. *Zhongguo Zhong Yao Za Zhi*. 2015 Aug;40(16):3314-8. Chinese.
- [5] Chen HY, Lin YH, Chen YC. Identifying Chinese herbal medicine network for treating acne: Implications from a nationwide database. *J Ethnopharmacol*. 2016 Feb 17;179:1-8. doi: 10.1016/j.jep.2015.12.032. Epub 2015 Dec 22.
- [6] Yu M, Cao QC, Su YX, Sui X, Yang HJ, Huang LQ, Wang WP. [Analysis on traditional Chinese medicine prescriptions treating cancer based on traditional Chinese medicine inheritance assistance system and discovery of new prescriptions]. *Zhongguo Zhong Yao Za Zhi*. 2015 Aug;40(15):3110-4. Chinese.
- [7] Yang HC, Iqbal U, Nguyen PA, Lin SH, Huang CW, Jian WS, Li YC. An automated technique to identify potential inappropriate traditional Chinese medicine (TCM)

- prescriptions. *Pharmacoepidemiol Drug Saf.* 2016 Apr;25(4):422-30. doi:10.1002/pds.3976. Epub 2016 Feb 23.
- [8] Zhao YQ, Teng J, Yang HJ. [Analysis on medication regularity of modern traditional Chinese medicines in treating melancholia based on data mining technology]. *Zhongguo Zhong Yao Za Zhi.* 2015 May;40(10):2042-6. Chinese.
- [9] Zhang JH, Zhu Y, Fan XH, Zhang BL. Efficacy-oriented compatibility for component-based Chinese medicine. *Acta Pharmacol Sin.* 2015 Jun;36(6):654-8. doi: 10.1038/aps.2015.8. Epub 2015 Apr 13. Review.
- [10] Chu SM, Shih WT, Yang YH, Chen PC, Chu YH. Use of traditional Chinese medicine in patients with hyperlipidemia: A population-based study in Taiwan. *J Ethnopharmacol.* 2015 Jun 20;168:129-35. doi: 10.1016/j.jep.2015.03.047. Epub 2015 Mar 28.
- [11] Li YB, Zhou XZ, Zhang RS, Wang YH, Peng Y, Hu JQ, Xie Q, Xue YX, Xu LL, Liu XF, Liu BY. Detection of herb-symptom associations from traditional Chinese medicine clinical data. *Evid Based Complement Alternat Med.* 2015;2015:270450. doi: 10.1155/2015/270450. Epub 2015 Jan 11.
- [12] Liu J, Pan J, Wang Y, Lin D, Shen D, Yang H, Li X, Luo M, Cao X. Component analysis of Chinese medicine and advances in fuming-washing therapy for knee osteoarthritis via unsupervised data mining methods. *J Tradit Chin Med.* 2013 Oct;33(5):686-91.
- [13] Chen HY, Lin YH, Su IH, Chen YC, Yang SH, Chen JL. Investigation on Chinese herbal medicine for primary dysmenorrhea: implication from a nationwide prescription database in Taiwan. *Complement Ther Med.* 2014 Feb;22(1):116-25. doi:10.1016/j.ctim.2013.11.012. Epub 2013 Dec 6.
- [14] Chien PS, Tseng YF, Hsu YC, Lai YK, Weng SF. Frequency and pattern of Chinese herbal medicine prescriptions for urticaria in Taiwan during 2009: analysis of the national health insurance database. *BMC Complement Altern Med.* 2013 Aug 15;13:209. doi: 10.1186/1472-6882-13-209.
- [15] Lin YH, Chen YC, Hu S, Chen HY, Chen JL, Yang SH. Identifying core herbal treatments for urticaria using Taiwan's nationwide prescription database. *J Ethnopharmacol.* 2013 Jul 9;148(2):556-62. doi: 10.1016/j.jep.2013.04.052. Epub 2013 May 14.
- [16] Cui M, Li J, Li H, Song C. Herbal compatibility of traditional Chinese medical formulas

- for acquired immunodeficiency syndrome. *J Tradit Chin Med*. 2012 Sep;32(3):329-34.
- [17] Xue R, Fang Z, Zhang M, Yi Z, Wen C, Shi T. TCMID: Traditional Chinese Medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Res*. 2013 Jan;41(Database issue):D1089-95. doi: 10.1093/nar/gks1100. Epub 2012 Nov 29.
- [18] He Y, Zheng X, Sit C, Loo WT, Wang Z, Xie T, Jia B, Ye Q, Tsui K, Chow LW, Chen J. Using association rules mining to explore pattern of Chinese medicinal formulae (prescription) in treating and preventing breast cancer recurrence and metastasis. *J Transl Med*. 2012 Sep 19;10 Suppl 1:S12. doi:10.1186/1479-5876-10-S1-S12. Epub 2012 Sep 19.
- [19] Fu XJ, Song XX, Wei LB, Wang ZG. Study of the distribution patterns of the constituent herbs in classical Chinese medicine prescriptions treating respiratory disease by data mining methods. *Chin J Integr Med*. 2013 Aug;19(8):621-8. doi: 10.1007/s11655-012-1090-2. Epub 2012 May 19.
- [20] Poon SK, Poon J, McGrane M, Zhou X, Kwan P, Zhang R, Liu B, Gao J, Loy C, Chan K, Sze DM. A novel approach in discovering significant interactions from TCM patient prescription data. *Int J Data Min Bioinform*. 2011;5(4):353-68.
- [21] Chen HY, Lin YH, Wu JC, Chen YC, Yang SH, Chen JL, Chen TJ. Prescription patterns of Chinese herbal products for menopausal syndrome: analysis of a nationwide prescription database. *J Ethnopharmacol*. 2011 Oct 11;137(3):1261-6. doi: 10.1016/j.jep.2011.07.053. Epub 2011 Aug 2.
- [22] Zhang XP, Zhou XZ, Huang HK, Feng Q, Chen SB, Liu BY. Topic model for Chinese medicine diagnosis and prescription regularities analysis: case on diabetes. *Chin J Integr Med*. 2011 Apr;17(4):307-13. doi: 10.1007/s11655-011-0699-x. Epub 2011 Apr 21.
- [23] Tang SH, Chen JX, Li G, Wu HW, Chen C, Zhang N, Gao N, Yang HJ, Huang LQ. Research on component law of Chinese patent medicine for anti-influenza and development of new recipes for anti-influenza by unsupervised data mining methods. *J Tradit Chin Med*. 2010 Dec;30(4):288-93.
- [24] Yang YH, Chen PC, Wang JD, Lee CH, Lai JN. Prescription pattern of traditional Chinese medicine for climacteric women in Taiwan. *Climacteric*. 2009 Dec;12(6):541-7. doi: 10.3109/13697130903060081.

- [25] Chen X, Zhou H, Liu YB, Wang JF, Li H, Ung CY, Han LY, Cao ZW, Chen YZ. Database of traditional Chinese medicine and its application to studies of mechanism and to prescription validation. *Br J Pharmacol*. 2006 Dec;149(8):1092-103. Epub 2006 Nov 6.
- [26] <https://www.ibm.com/us-en/marketplace/spss-modeler>
- [27] <http://weka.wikispaces.com/>
- [28] Wang Y, Yu Z, Jiang Y, Xu K, Chen X. Automatic symptom name normalization in clinical records of traditional Chinese medicine. *BMC Bioinformatics*. 2010 Jan 20;11:40. doi: 10.1186/1471-2105-11-40.
- [29] Fang YC, Huang HC, Chen HH, Juan HF. TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining. *BMC Complement Altern Med*. 2008 Oct 14;8:58. doi: 10.1186/1472-6882-8-58.
- [30] Cai YM, Chen LP, Li JS, Li QL, Sun SM, Li CW. [Database Establishing and Data Mining of Pulmonary Diseases Based on Clinical Works by Modern Famous Veteran Doctors of Chinese Medicine]. *Zhongguo Zhong Xi Yi Jie He Za Zhi*. 2015 Oct;35(10):1186-91. Chinese.
- [31] Feng S, Ren Y, Fan S, Wang M, Sun T, Zeng F, Li P, Liang F. Discovery of Acupoints and Combinations with Potential to Treat Vascular Dementia: A Data Mining Analysis. *Evid Based Complement Alternat Med*. 2015;2015:310591. doi:10.1155/2015/310591. Epub 2015 Jul 30.
- [32] Zhao Y, Liu B, He L, Bai W, Yu X, Cao X, Luo L, Rong P, Zhao Y, Li G, Liu B. A novel classification method for aid decision of traditional Chinese patent medicines for stroke treatment. *Front Med*. 2017 May 13. doi:10.1007/s11684-017-0511-1.
- [33] Liu K, Wang Jie, Wang YG. [Acupoint Selection Laws for Massage Therapy of Infantile Anorexia: an Analysis Based on Data Mining]. *Zhongguo Zhong Xi Yi Jie He Za Zhi*. 2016 Jun;36(6):753-6. Chinese.
- [34] Tao S, Xu W, Gao Z, Dong Q. [Anaysis on acupoint selection rule of acupuncture for trigeminal neuralgia]. *Zhongguo Zhen Jiu*. 2016 Feb;36(2):207-11. Chinese.
- [35] Han D, Liu C, Qie L, Wang F, Wang Zhaohui. [Acupoint selection and medication rules analysis for allergic rhinitis treated with acupoint application-based on data mining technology]. *Zhongguo Zhen Jiu*. 2015 Nov;35(11):1177-80. Chinese.

- [36] Cai L, Wu J, Ma T, Yang L. [Characteristics of acupoint application for the sub-healthy condition treated with ancient and modern acupuncture based on data mining exploration]. *Zhongguo Zhen Jiu*. 2015 Oct;35(10):1073-9. Chinese.
- [37] Yang M, Li JQ, Jiao LJ, Chen PQ, Xu L. [Effective core formulae for lung cancer based on complex network and survival analysis]. *Zhongguo Zhong Yao Za Zhi*. 2015 Nov;40(22):4482-90. Chinese.
- [38] Wu C, Zhang H, Wang C, Qin H, Zhu M, Zhang J. An Integrated Approach for Studying Exposure, Metabolism, and Disposition of Multiple Component Herbal Medicines Using High-Resolution Mass Spectrometry and Multiple Data Processing Tools. *Drug Metab Dispos*. 2016 Jun;44(6):800-8. doi: 10.1124/dmd.115.068189. Epub 2016 Mar 24.
- [39] Hou ZK, Liu FB, Li PW, Zhuang KH. [Mining analysis and experience summary for chronic atrophic gastritis cases treated by Professor LIU Feng-bin]. *Zhongguo Zhong Yao Za Zhi*. 2015 Jun;40(11):2227-34. Chinese.
- [40] Zhou X, Li Y, Peng Y, Hu J, Zhang R, He L, Wang Y, Jiang L, Yan S, Li P, Xie Q, Liu B. Clinical phenotype network: the underlying mechanism for personalized diagnosis and treatment of traditional Chinese medicine. *Front Med*. 2014 Sep;8(3):337-46. doi: 10.1007/s11684-014-0349-8. Epub 2014 Aug 12. Review.
- [41] Su K, Zhu F, Guo L, Zhu Y, Li W, Xiong X. Retrospective study on Professor Zhongying Zhou's experience in Traditional Chinese Medicine treatment on diabetic nephropathy. *J Tradit Chin Med*. 2013 Apr;33(2):262-7.
- [42] Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, Guo Y, Zhang H, Gao Z, Yan X. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. *Artif Intell Med*. 2010 Feb-Mar;48(2-3):139-52. doi: 10.1016/j.artmed.2009.07.012. Epub 2010 Feb 1.