ACM-2013-0178-ver9-Zhao_1P
Type: research-article

THE JOURNAL OF ALTERNATIVE AND COMPLEMENTARY MEDICINE
Volume 0, Number 0, 2014, pp. 1–7

DOI: 10.1089/acm.2013.0178

Original Article

Discovering Symptom Co-occurrence Patterns from 604 Cases of Depressive Patient Data Using Latent Tree Models

Yan Zhao, MD,[1] Nevin L. Zhang, PhD,[2] Tianfang Wang, MD,[1] and Qingguo Wang, MD[1]

Abstract

Objectives: In order to treat depressive patients using Traditional Chinese Medicine (TCM), it is necessary to classify them into subtypes from the TCM perspective. Those subtypes are called *Zheng* types. This article aims at providing evidence for the classification task by discovering symptom co-occurrence patterns from clinic data.
Methods: Six hundred four (604) cases of depressive patient data were collected. The subjects were selected using the Chinese classification of mental disorder clinic guideline CCMD-3. The symptoms were selected based on the TCM literature on depression. The data were analyzed using latent tree models (LTMs).
Results: An LTM with 29 latent variables was obtained. Each latent variable represents a partition of the subjects into 2 or more clusters. Some of the clusters capture probabilistic symptom co-occurrence patterns, while others capture symptom mutual-exclusion patterns. Most of the co-occurrence patterns have clear TCM *Zheng* connotations.
Conclusions: From clinic data about depression, probabilistic symptom co-occurrence patterns have been discovered that can be used as evidence for the task of classifying depressive patients into *Zheng* types.

Introduction

In China, doctors at Traditional Chinese Medicine (TCM) hospitals and clinics often classify depressive patients into several *Zheng* types from the TCM perspective, and treat different types differently. This article is concerned with the important issue of how to carry out the classification properly.

There are currently no widely accepted guidelines on the classification of depressive patients into *Zheng* types.[1] In clinic research, different researchers adopt different strategies. For example, Gao and Fang[1] divide depressive patients into three types: Stagnation of Liver *Qi*, Spirit Injured by Worry, and Heart–Spleen Dual Vacuity. You et al.[2] divide them into three types: Liver Depression and Spleen Vacuity, Heart–Spleen Dual Vacuity, and Deficiency of Liver-*Yin* and Kidney-*Yin*. Guo *et al*.[3] divide them into four types: Liver Depression and Spleen Vacuity, Liver Blood Stasis and Stagnation, Heart–Spleen Dual Vacuity, and Spleen and Kidney Dual Vacuity.

The objective of this work is, through the analysis of clinic symptom data, to provide evidence that can be used to answer the following questions: What *Zheng* types are present in the population of depressive patients? What are the characteristics of each type? How can one differentiate between the different types? In TCM, patient classification (also known as syndrome differentiation) is based on symptom co-occurrence patterns. Therefore, this study aimed to identify such patterns from clinic symptom data.

Six hundred and four (604) cases of depressive patient data were collected. The subjects were selected using the Chinese classification of mental disorders clinic guideline CCMD-3.[4] The symptoms were selected based on the TCM literature on depression. In other words, symptoms of interest related to TCM that were reported to have occurred in depressive patients were used. The data were analyzed using latent tree models (LTMs).[5,6] The analysis reveals a host of probabilistic symptom co-occurrence patterns with clear TCM *Zheng* connotations, as well as symptom mutual-exclusion patterns.

Methods

Data collection

The data were collected in 2005–2006. The subjects were inpatients or outpatients aged between 19 and 69 years from

[1]Zhao Y. Depression syndrome-element analysis based on multiple unsupervised data analysis methods. PhD thesis, Beijing University of Traditional Medicine, 2007.

[1]Department of TCM Diagnostics, Beijing University of Chinese Medicine, Beijing, China.
[2]Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.

9 hospitals from several regions of China. They were selected using the Chinese classification of mental disorders clinic guideline CCMD-3.[4] CCMD-3 is similar in structure and categorization to the International Classification of Diseases (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM), though it includes some variations on their main diagnoses and around 40 culturally related diagnoses. For example, CCMD-3 places more emphasis than ICD and DSM do on neurasthenia, which denotes a mental disorder marked by chronic weakness and easy fatigability. *Koro* and *qigong* deviation are examples of culture-specific symptoms included in CCMD-3.

Excluded from the study were subjects who took antidepression drugs within 2 weeks prior to the survey, women in the gestational and nursing periods, patients suffering from other mental disorders such as mania, and those suffering from other severe diseases or having had operations recently.

The symptoms (and signs) were extracted from the TCM literature on depression between 1994 and 2004. The searched was done with the terms "抑郁" and "证" (depression and *Zheng*) on the CNKI (China National Knowledge Infrastructure) database. Among the articles returned by CNKI, only those were kept on studies where patients were selected using the ICD-9, ICD-10, CCMD-2, or CCMD-3 guidelines. This resulted in 65 articles, which contained a total of 198 distinct symptoms. The symptoms that appeared only 1 time or 2 times were removed. This resulted in 143 symptoms.

An epidemiologic survey was conducted on the 143 symptoms. Six hundred and four (604) patient cases were collected. Each patient case contains information about which symptoms occurred in the patient and which ones did not. Various measures were taken to ensure data quality. Examples include staff training, site visit by principal investigators, and dual data entry.

In the 604 patient cases, 57 symptoms occurred fewer than 10 times. They were removed from the data set, and the remaining 86 symptoms were included in further analysis.

Data analysis

In data-driven medical research, latent class analysis (LCA)[7] is often used to identify subtypes in a group of patients. LCA is based on the latent class model (LCM), which consists of 1 latent variable and multiple symptom variables that are observed. The symptom variables are assumed to be mutually independent, given the latent variable. To perform LCA, one needs to determine the number of states for the latent variable (i.e., the number of clusters, and the probabilistic parameters). Several researchers have used LCA to study major depressive disorder from the Western medicine perspective. According to the recent systematic review by van Loo *et al.*,[8] those studies "mainly grouped patients on overall severity, but not in classes with qualitatively different symptom profiles."

To analyze the TCM depression data, a generalization of LCM called LTMs was used.[5] An LTM can be viewed as a collection of LCMs where each LCM involves 1 latent variable and a distinct subset of the symptom variables, and the latent variables are connected to form a tree structure. Each LCM partitions the patients into two or more clusters. Because different LCMs involve different symptom variables, the clusters given by different LCMs have qualitatively different symptom profiles and can capture symptom co-occurrence patterns.

Currently, the state-of-the-art algorithm for latent tree analysis (LTA) is the EAST algorithm, where EAST stands for Extension Adjustment Simplification until Termination.[6] A Java implementation[2] of EAST was used to analyze the TCM depression data.

Results

Model structure

The result of the analysis is an LTM. The structure of the model is shown in Figure 1. The nodes labeled with English phrases represent symptom variables. Each of them has two possible values, indicating the presence or absence of the symptoms. The symptom variables come from the data set. The nodes labeled with the capital letter "Y" and integer subscripts are the latent variables. They are not from the data set. Rather they were introduced during data analysis to explain patterns in the data. There is an integer next to each latent variable. It is the number of possible states of that latent variable. For example, the latent variable Y_1 has 3 possible states, while Y_{29} has 2.

Each latent variable and the symptom variables directly connected to it form an LCM. For example, Y_{29} forms an LCM with "fear of cold," "cold limbs," and "surging pulse." It will referred to as the LCM headed by Y_{29}, or simply the Y_{29} LCM. Numerical information includes the conditional probability distribution of each symptom variable given the latent variable. The strengths of the dependencies as measured by mutual information are visually depicted by the widths of the edges. For example, Y_{29} is strongly correlated with "cold limbs," moderately correlated with "fear of cold," and weakly correlated with "surging pulse."

The latent variables are connected up to form a tree structure. The correlatedness between two neighboring latent variables is characterized by a probability distribution. The strength of the correlation is depicted, shown by the width of the edges between the two variables. For example, and Y_{15} are Y_{16} are strongly correlated, while Y_{28} and Y_{29} are only marginally correlated.

As will be seen later, most of the connections among the variables are consistent with the TCM postulates on how the variables are related to each other. However, several symptom variables in the model seem to be out of place. They include "somnolence" under Y_{11}, "tinnitus" and "pain in limbs" under Y_{12}, "dry eyes" under Y_{15}, "yellow urine" under Y_{16}, "sloppy stool" under Y_{24}, and "surging pulse" under Y_{29}. The reason is that those symptoms occur rarely in the data and hence there is not sufficient information to determine their appropriate locations in the model. It is for this reason that those symptom variables are only weakly related to the latent variables to which they are connected. We will ignore those variables in subsequent discussions.

Symptom co-occurrence patterns

Each latent variable in the model represents a partition of the patients surveyed, and each state of the latent variable denotes

[2]Available at: http://www.cse.ust.hk/∼lzhang/ltm/index.htm.

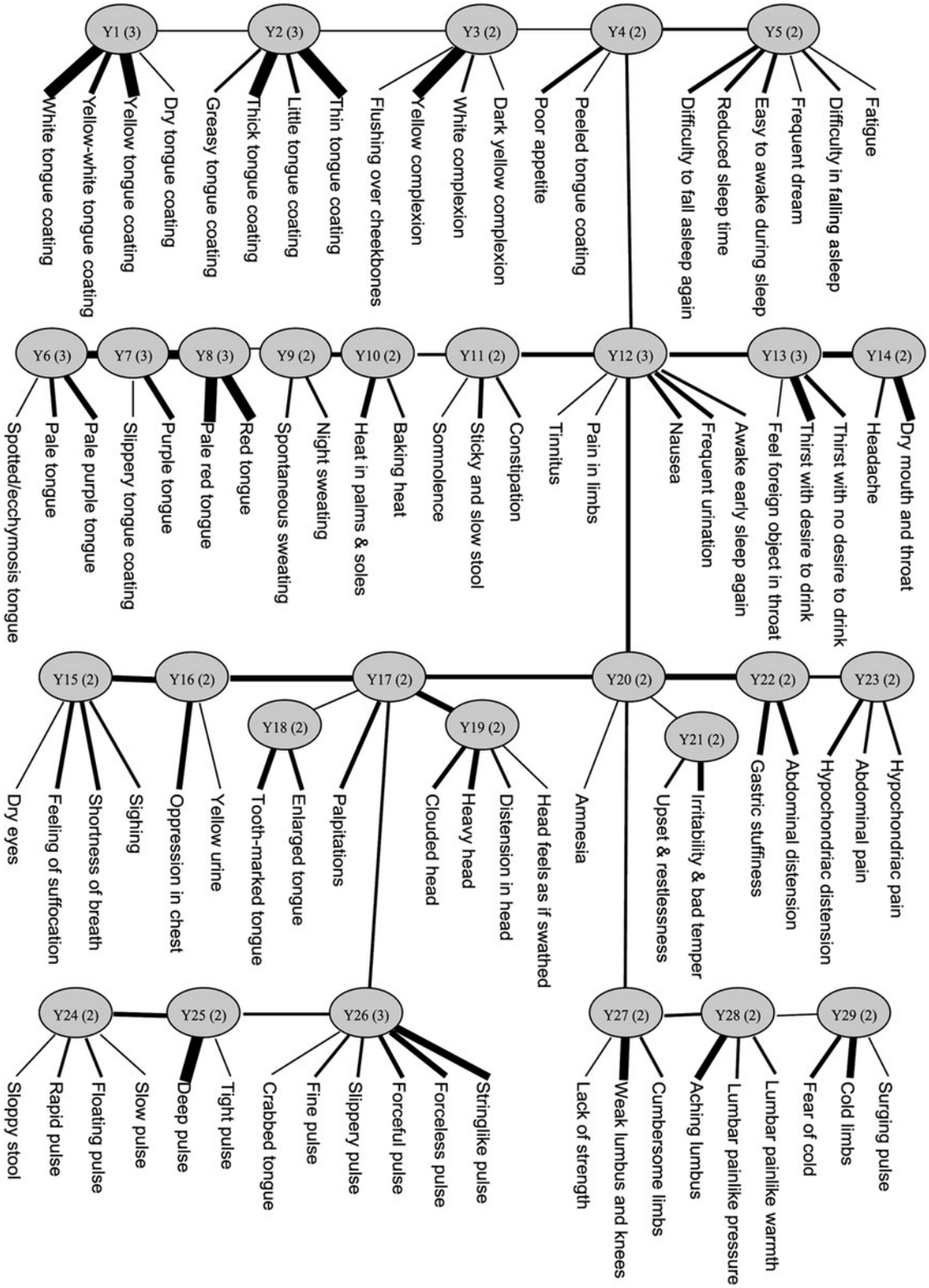

FIG. 1. The structure of the model obtained by latent tree analysis on the depression data set.

a cluster. For example, the latent variable Y_{29} has two states, which are denoted as $Y_{29}=s_0$ and $Y_{29}=s_1$, respectively. It represents a partition of the patients into two clusters.

In Figure 1, Y_{29} forms an LCM with three symptom variables. To appreciate the meaningfulness of the partition represented by Y_{29}, the probability distributions of the two symptom variables ''fear of cold'' and ''cold limbs'' in the two clusters $Y_{29}=s_0$ and $Y_{29}=s_1$ are examined. They are given in Table 1(a). It shows that the first cluster ($Y_{29}=s_0$) consists of 54% of the patients while the second cluster ◀T1

Table 1. Latent Variables That Reveal Probabilistic Symptom Co-occurrence Patterns: Each Latent Variable Represents a Partition of the Patients Surveyed; This Table Gives the Probability Distributions of Relevant Symptom Variables in the Clusters of the Partitions

(a) Partition given by $Y_{29}=s_1$	$Y_{29}=s_0$ *(0.54)*	$Y_{29}=s_1$ *(0.46)*
Cold limbs	0.02	0.80
Fear of cold	0.29	0.85
(b) Partition given by Y_{28}	$Y_{28}=s_0$ *(0.73)*	$Y_{28}=s_1$ *(0.27)*
Aching lumbus	0.01	0.81
Lumbar painlike pressure	0.01	0.39
Lumbar painlike warmth	0.00	0.33
(c) Partition given by Y_{27}	$Y_{27}=s_0$ *(0.56)*	$Y_{27}=s_1$ *(0.44)*
Weak lumbus and knees	0.15	1.0
Cumbersome limbs	0.43	0.82
Lack of strength	0.91	0.99
(d) Partition given by Y_{23}	$Y_{23}=s_0$ *(0.84)*	$Y_{23}=s_1$ *(0.16)*
Hypochondriac distension	0.06	0.71
Hypochondriac pain	0.01	0.38
Abdominal pain	0.04	0.39
(e) Partition given by Y_{22}	$Y_{22}=s_0$ *(0.72)*	$Y_{22}=s_1$ *(0.28)*
Abdominal distension	0.09	0.89
Gastric stuffiness	0.18	0.92
(f) Partition given by Y_{21}	$Y_{21}=s_0$ *(0.19)*	$Y_{21}=s_1$ *(0.81)*
Irritability and bad temper	0.24	0.98
Upset and restlessness	0.67	0.98
(g) Partition given by Y_{19}	$Y_{19}=s_0$ *(0.41)*	$Y_{19}=s_1$ *(0.59)*
Heavy head	0.18	0.88
Clouded head	0.48	0.96
Distension in head	0.24	0.78
Head feels as if swathed	0.08	0.32
(h) Partition given by Y_{18}	$Y_{18}=s_0$ *(0.61)*	$Y_{18}=s_1$ *(0.39)*
Tooth-marked tongue	0.03	0.74
Enlarged tongue	0.03	0.49
(i) Partition given by Y_{16}	$Y_{16}=s_0$ *(0.48)*	$Y_{16}=s_1$ *(0.52)*
Oppression in chest	0.31	1.0
Shortness of breath	0.27	0.75
Feeling of suffocation	0.26	0.74
Palpitations	0.60	0.88
Sighing	0.63	0.87
(j) Partition given by Y_{15}	$Y_{15}=s_0$ *(0.52)*	$Y_{15}=s_1$ *(0.48)*
Shortness of breath	0.20	0.86
Feeling of suffocation	0.19	0.85
Oppression in chest	0.42	0.92
Sighing	0.60	0.92
(k) Partition given by Y_{11}	$Y_{11}=s_0$ *(0.86)*	$Y_{11}=s_1$ *(0.14)*
Sticky and slow stool	0.07	0.95
Constipation	0.23	0.72
(l) Partition given by Y_{10}	$Y_{10}=s_0$ *(0.65)*	$Y_{10}=s_1$ *(0.35)*
Heat in palms and soles	0.05	0.75
Baking heat	0.13	0.50
(m) Partition given by Y_9	$Y_9=s_0$ *(0.34)*	$Y_9=s_1$ *(0.66)*
Spontaneous sweating	0.10	0.61
Night sweating	0.09	0.43

($Y_{29}=s_1$) consists of 46% of the patients. The two symptoms ''fear of cold'' and ''cold limbs'' do not occur often in the first cluster, while they both tend to occur with high probabilities (0.8 and 0.85) in the second cluster.

The probability distribution indicates that the two symptoms ''fear of cold'' and ''cold limbs'' tend to co-occur in the cluster $Y_{29}=s_1$. The co-occurrence is not a certain event. Rather it is probabilistic in nature. Thus, it is called a probabilistic symptom co-occurrence pattern. It turns out that the pattern is meaningful from the TCM perspective. As a matter of fact, TCM asserts that *Yang* Deficiency can lead to, among other symptoms, ''fear of cold'' and ''cold limbs''[9, p. 192] So, the presence of the probabilistic pattern $Y_{29}=s_1$ suggests the *Zheng*-type *Yang* Deficiency.

Note that there are three notions concerning $Y_{29}=s_1$: First, it is a state of the latent variable Y_{29}; second, it denotes a probabilistic symptom co-occurrence pattern; and third, it represents the cluster of patients with the pattern.

The LTA has revealed a host of probabilistic symptom co-occurrence patterns that are meaningful from the TCM perspective. As shown in Table 1(b), the latent state $Y_{28}=s_1$ captures the probabilistic co-occurrence of ''aching lumbus,'' ''lumbar painlike pressure'' and ''lumbar painlike warmth.'' This pattern is present in 27% of the patients and it suggests the *Zheng*-type Kidney Deprived of Nourishment.[9, p. 192] The latent state $Y_{27}=s_1$ (Table 1[c]) captures the probabilistic co-occurrence of ''weak lumbus and knees'' and ''cumbersome limbs.'' This pattern is present in 44% of the patients and it suggests the *Zheng*-type Kidney Deficiency.[9, p. 192]

The discussions, which so far have focused on the bottom right corner of the model structure, now move to the latent variables depicted on the second-last level. The latent state $Y_{23}=s_1$ (Table 1[d]) captures the probabilistic co-occurrence of ''hypochondriac distension,'' ''hypochondriac pain,'' and ''abdominal pain.'' This pattern is present in 16% of the

patients and it suggests the *Zheng*-type Liver *Qi* Stagnation.[9, p. 252] The latent state $Y_{22}=s_1$ (Table 1[e]) captures the probabilistic co-occurrence of ''gastric stuffiness,'' and ''abdominal distension.'' This pattern is present in 28% of the patients and it also suggests the *Zheng*-type Liver *Qi* Stagnation.[9, p. 252]

The latent state $Y_{21}=s_1$ (Table 1[f]) captures the probabilistic co-occurrence of ''upset and restlessness'' and ''irritability and bad temper.'' This pattern is present in 81% of the patients and it suggests the *Zheng*-type Stagnant *Qi* Turning into Fire (also known as Liver Fire Flaming Up).[9, p. 253] The latent state $Y_{19}=s_1$ (Table 1 [g]) captures the probabilistic co-occurrence of ''clouded head,'' ''heavy head,'' and ''distention in head.'' This pattern is present in 59% of the patients and it suggests the *Zheng*-type *Qi* Stagnation in Head.[9, p. 253] The latent state $Y_{15}=s_1$ (Table 1[j]) captures the probabilistic co-occurrence of ''feeling of suffocation,'' ''shortness of breath,'' and ''sighing.'' This pattern is present in 48% of the patients and it suggests the *Zheng*-type *Qi* Deficiency.[9, p. 234]

The latent variable Y_{17} is directly connected with only one symptom variable ''palpitations,'' which suggests Heart *Qi* Deficiency in TCM.[9, p. 264] The latent variable Y_{16} is directly connected with two symptom variables. However, the dependence of ''yellow urine'' on Y_{16} is only marginal. The other symptom variable is ''oppression in chest,'' which suggests *Qi* Deficiency in TCM.[9, p. 240] Those two latent variables do not reveal symptom co-occurrence patterns themselves. However, their relationships with neighboring latent variables do. Using those relationships, one can calculate, for instance, the probability distributions of relevant symptom variables in the Y_{16} clusters. They are shown in Table 1(i). The distributions for the cluster $Y_{16}=s_1$ indicate that ''oppression in chest'' tends to co-occur with ''shortness of breath,'' ''feeling of suffocation,'' ''palpitation,'' and ''sighing.''

It is noted that all the latent variables depicted on the second-last level except 1 are related to *Qi* disorders. The exception is Y_{18}. The latent state $Y_{18}=s_1$ (Table 1[h]) captures the probabilistic co-occurrence of ''enlarged tongue'' and ''tooth-marked tongue.'' This pattern is present in 39% of the patients and it suggests the *Zheng*-type Internal Accumulation of Excessive Dampness.[9, p. 59] This is not related to *Qi* disorders, which explains why the connection between Y_{18} and Y_{17} is weak.

The latent state $Y_{11}=s_1$ (Table 1[k]) captures the probabilistic co-occurrence of ''sticky and slow stool'' and ''constipation.'' This pattern is present in 14% of the patients and it also suggests the *Zheng*-type Deficiency of Stomach/Spleen *Yin*.[9, p. 280] The latent state $Y_{10}=s_1$ (Table 1[l]) captures the probabilistic co-occurrence of ''heat in palms and soles'' and ''baking heat.'' This pattern is present in 35% of the patients and it suggests the *Zheng*-type *Yin* Deficiency.[9, p. 293] The latent state $Y_9=s_1$ (Table 1[m]) captures the probabilistic co-occurrence of ''spontaneous sweating'' and ''night sweating.'' This pattern is present in 66% of the patients and it suggests the *Zheng*-type Deficiency of Both *Qi* and *Yin*.[9, pp. 239, 293]

T2 ▶ Table 2 shows information about two other latent variables, Y_5 and Y_{12}. The latent state $Y_5=s_1$ captures the probabilistic co-occurrence of ''difficulty in falling asleep,'' ''easy to awake during sleep,'' ''difficulty in falling asleep again,'' and ''reduced sleep time.'' This pattern is present in 68% of the patients and it clearly suggests sleep disorders. The latent state $Y_{12}=s_2$ captures the probabilistic co-occurrence of ''nausea,'' ''frequent urination,'' and ''awake early sleep again.'' This pattern is present in 8% of the patients and it is clearly meaningful.

TABLE 2. MORE LATENT VARIABLES THAT REVEAL PROBABILISTIC SYMPTOM CO-OCCURRENCE PATTERNS

(a) Partition given by Y_5	*$Y_5=s_0$ (0.32)*	*$Y_5=s_1$ (0.68)*
Difficulty in falling asleep again	0.13	0.81
Easy to awake during sleep	0.33	0.92
Reduced sleep time	0.35	0.93
Difficulty in falling asleep	0.53	0.89
(b) Partition given by Y_{12}	*$Y_{12}=s_0$ (0.47)*	*$Y_{12}=s_1$ (0.45)*	*$Y_{12}=s_2$ (0.08)*
Nausea	0.12	0.50	0.98
Frequent urination	0.02	0.32	0.83
Awake early sleep again	0.38	0.28	1.0

Symptom mutual-exclusion patterns

Table 3 shows information about latent variables that ◀T3 reveal symptom mutual-exclusion patterns. The latent variable Y_1 (Table 3[a]) reveals the mutual exclusion of ''white tongue coating,'' ''yellow tongue coating,'' and ''yellow-white tongue coating.'' Indeed, Y_1 divides the patients into three clusters. Each of the three symptoms occurs only in 1 of the clusters. No two symptoms occur in the same cluster. So the symptoms are mutually exclusive.

Similarly, Y_2 (Table 3[b]) reveals the mutual exclusion of ''thin tongue coating,'' ''thick tongue coating,'' and ''little tongue coating.'' Y_3 (Table 3[c]) reveals the mutual exclusion of ''yellow complexion'' and ''white/dark-yellow complexion.'' Y_6 (Table 3[d]) reveals the mutual exclusion of ''pale purple tongue,'' ''pale tongue,'' and ''purple tongue.'' Y_8 (Table 3[e]) reveals the mutual exclusion of ''pale red tongue'' and ''red tongue.''

Y_{13} (Table 3[f]) indicates that ''thirst with desire to drink'' and ''thirst with no desire to drink'' are mostly mutually exclusive. However, they do co-occur in a small fraction of the patients. The reason is that those patients were not sure which of the two alternatives to check off in the survey and thus checked off both. Note that Table 3(f) also indicates that those two symptoms tend to co-occur with ''dry mouth and throat.''

Y_{24} (Table 3[g]) reveals the mutual exclusion of ''deep pulse'' and ''floating pulse.'' It also uncovers the fact that ''rapid pulse'' co-occurs with only ''floating pulse,'' but not ''deep pulse.'' On the other hand, ''slow pulse'' co-occurs with only ''deep pulse,'' but not ''floating pulse.'' Y_{25} (Table 3[h]) reveals the mutual exclusion of ''forceless pulse'' and ''forceful pulse.'' ''String-like pulse'' tends to be mutual exclusive with both of them, while ''slippery pulse'' tends to co-occur with ''forceful pulse.''

In summary, this analysis has identified from survey data a host of probabilistic symptom co-occurrence and mutual-exclusion patterns. Some of the symptom co-occurrence

Table 3. Latent Variables That Reveal Probabilistic Symptom Mutual-Exclusion Patterns

(a) Partition given by Y_1	*$Y_1 = s_0$ (0.48)*	*$Y_1 = s_1$ (0.40)*	*$Y_1 = s_2$ (0.12)*
White tongue coating	1.0	0.00	0.00
Yellow tongue coating	0.00	1.0	0.00
Yellow–white tongue coating	0.00	0.00	0.88
(b) Partition given by Y_2	*$Y_2 = s_0$ (0.50)*	*$Y_2 = s_1$ (0.43)*	*$Y_2 = s_2$ (0.07)*
Thin tongue coating	1.0	0.00	0.00
Thick tongue coating	0.00	1.0	0.00
Little tongue coating	0.00	0.00	0.91
(c) Partition given by Y_3	*$Y_3 = s_0$ (0.42)*		*$Y_3 = s_1$ (0.58)*
Yellow complexion	1.0		.00
White complexion	.01		0.25
Dark yellow complexion	0.00		0.09
(d) Partition given by Y_6	*$Y_6 = s_0$ (0.08)*	*$Y_6 = s_1$ (0.05)*	*$Y_6 = s_2$ (0.87)*
Pale purple tongue	1.0	0.00	0.00
Pale tongue	0.00	1.0	0.00
Purple tongue	0.00	0.00	0.13
(e) Partition given by Y_8	*$Y_8 = s_0$ (0.47)*	*$Y_8 = s_1$ (0.30)*	*$Y_8 = s_2$ (0.23)*
Pale red tongue	1.0	0.00	0.00
Red tongue	0.00	1.0	0.00
(f) Partition given by Y_{13}	*$Y_{13} = s_0$ (0.39)*	*$Y_{13} = s_1$ (0.33)*	*$Y_{13} = s_2$ (0.28)*
Thirst with desire to drink	0.09	1.0	0.01
Thirst with no desire to drink	0.05	0.00	0.59
Dry mouth and throat	0.43	1.0	1.0
(g) Partition given by Y_{24}	*$Y_{24} = s_0$ (0.73)*		*$Y_{24} = s_1$ (0.27)*
Deep pulse	0.57		0.00
Floating pulse	0.00		0.22
Rapid pulse	0.08		0.53
Slow pulse	0.03		0.00
(h) Partition given by Y_{26}	*$Y_{26} = s_0$ (0.36)*	*$Y_{26} = s_1$ (0.31)*	*$Y_{26} = s_2$ (0.33)*
Forceless pulse	0.59	0.00	0.00
Forceful pulse	0.00	0.52	0.10
Stringlike pulse	0.23	0.15	1.0
Slippery pulse	0.04	0.43	0.17

patterns have clear TCM *Zheng* connotations and hence are especially interesting. Those patterns are summarized in Table 4. T4 ▶

Discussion

Previous data-driven investigations of depression were based on the 12 symptoms disaggregated from the nine DSM-III-R criteria for major depression.[8] The objectives were either to identify symptom dimensions using factor analysis or to determine subtypes of depression using LCA.

This study is based on 86 symptoms that are of interest from the TCM perspective. The data were analyzed using a new method called LTA. The analysis reveals both symptom dimensions and interesting subclasses of patients. Specifically, the output of the analysis is a LTM that consists of 29 latent variables. Each latent variable represents a symptom dimension, and a partition of the patients along that dimension. Some of the clusters in the partitions capture probabilistic symptom co-occurrence patterns, while others capture symptom mutual-exclusion patterns.

In China, depressive patients are often treated using TCM. To do so, doctors need to first classify those patients into subtypes (called *Zheng* types) from the TCM perspective, and then come up with treatment plans for each subtype. Three questions arise: (1) What *Zheng* types are present in the population of depressive patients? (2) What are the characteristics of each type? and (3) How can one differentiate between the different types? The results of this analysis can be used as evidence for answering those questions.

For example, the authors' analysis has revealed the probabilistic co-occurrence of ''hypochondriac distention,''

Table 4. Summary of Probabilistic Symptom Co-occurrence Patterns That Have Clear Traditional Chinese Medicine *ZHENG* Connotations

Latent state	*Symptom co-occurrence pattern*	Zheng *type*
$Y_9 = s_1$	Spontaneous sweating, night sweating	Deficiency of both *Qi* and *Yin*
$Y_{10} = s_1$	Heat in palms and soles, baking heat	*Yin* Deficiency
$Y_{11} = s_1$	Sticky and slow stool, constipation	Deficiency of Stomach/Spleen *Yin*
$Y_{15} = s_1$	Shortness of breath, feeling of suffocation, sighing	*Qi* Deficiency
$Y_{16} = s_1$	Oppression in chest, shortness of breath, feeling of suffocation, palpitation, sighing	*Qi* Stagnation
$Y_{18} = s_1$	Enlarged tongue, tooth-marked tongue	Internal Accumulation of Excessive Dampness
$Y_{19} = s_1$	Clouded head, heavy head, distension in head	Qi Stagnation in Head
$Y_{21} = s_1$	Upset and restlessness, irritability and bad temper	Stagnant *Qi* Turning into Fire
$Y_{22} = s_1$	Gastric stuffiness, abdominal distension	Liver *Qi* Stagnation
$Y_{23} = s_1$	Hypochondriac distension, hypochondriac pain, abdominal pain	Liver *Qi* Stagnation
$Y_{27} = s_1$	Weak lumbus and knees, cumbersome limbs	Kidney Deficiency
$Y_{28} = s_1$	Aching lumbus, lumbar painlike pressure, lumbar painlike warmth	Kidney Deprived of Nourishment
$Y_{29} = s_1$	Fear of cold, cold limbs	*Yang* Deficiency

''hypochondriac pain,'' and ''abdominal pain'' and that the pattern is present in 16% of the depressive patients (see $Y_{23} = s_1$ in Table 1[d]). From those, we can conclude the presence of the *Zheng*-type Liver *Qi* Stagnation in the population of depressive patients.

The subsequent questions are the following: (1) What are the characteristics of the *Zheng*-type Liver *Qi* Stagnation? and (2) How can one determine whether a particular patient belongs to the type? Y_{23} provides some evidence for answering those questions. However, the questions cannot be answered based solely on Y_{23}. The reason is that Y_{23} captures only one aspect of Liver *Qi* Stagnation. As shown in Table 4, the latent variables Y_{16}, Y_{21}, and Y_{22} are also related to Liver *Qi* Stagnation. They capture other aspects of the ZHENG type. Therefore, it will be necessary to jointly consider those latent variables (and potentially others) in order to obtain an appropriate overall characterization of Liver *Qi* Stagnation. Future research will determine how this can be done. In this article, some of the necessary building blocks have been provided.

Conclusions

By analyzing 605 cases of depressive patient data using LTMs, a host of probabilistic symptom co-occurrence patterns and symptom mutual-exclusion patterns have been discovered. Most of the co-occurrence patterns have clear TCM *Zheng* connotations, while the mutual-exclusion patterns are also reasonable and meaningful. The patterns can be used as evidence for the task of classifying depressive patients into *Zheng* types.

Acknowledgments

Research on this article was supported by China National Basic Research 973 Program under Project No. 2004CB517106, 2011CB505101, 2011CB505105, Guangzhou HKUST Fok Ying Tung Research Institute, Innovative Team Project of Beijing University of Chinese Medicine (2011-CXTD-08), and Research Base Development Project of Beijing University of Chinese Medicine (2011-JDJS-09).

Disclosure Statement

No competing financial interests exist.

References

1. Gao XZ, Fang YM. Dialectical nursing of 30 cases of depressive patients. Hebei J Trad Chin Med 1995;2:41–41.
2. You K, Zhang QP, Ding CL. Clinic observations of 120 cases of depressive patients treated with the TCM principle of syndrome differentiation. Heilongjiang Med Pharm 2001;4:107–108.
3. Guo YM, Liu CF, Yang JJ. TCM treatment of 38 cases of depression. China's Naturopath 2001;10:56–57.
4. Chen YF. Chinese classification of mental disorders (CCMD-3): Towards integration in international classification. Psychopathology 2002;35:171–175.
5. Zhang NL, Yuan SH, Chen T, Wang Y. Latent tree models and diagnosis in traditional Chinese medicine. Artif Intell Med 2008;42:229–245.
6. Chen T, Zhang NL, Liu TF, Wang Y, Poon LKM. Model-based multidimensional clustering of categorical data. Artif Intell 2011;176:2246–2269.
7. Bartholomew DJ, Knott M. Latent Variable Models and Factor Analysis, 2nd ed. London: Arnold, 1999.
8. van Loo HM, de Jonge P, Romeijn JW, et al. Data-driven subtypes of major depressive disorder: A systematic review. BMC Med 2012;10:156.
9. Yang WY, Meng FY, Jiang YN, et al. Diagnostics of Traditional Chinese Medicine. Beijing: Xueyuan Press, 1998.

Address correspondence to:
Nevin L. Zhang, PhD
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Clear Water Bay Road
Kowloon
Hong Kong, China

E-mail: lzhang@cse.ust.hk

AUTHOR QUERY FOR ACM-2013-0178-VER9-ZHAO_1P

AU1: Note former ref. 1 has been deleted and made into a footnote; unpublished PhD theses are not allowed in the reference list

AU2: Indicate which edition of DSM

AU3: Can ‘‘integer subscripts’’ be changed here to ‘‘integers in parentheses’’? Fig. 1 does not use subscripts, but parentheses instead