基于 Apriori 算法的中药气-味-效三维数据关联规则挖掘研究

金锐1, 林茜1, 张冰1, 刘欣1, 刘森茂2, 赵蕊1, 刘秀兰1
1. 北京中医药大学中药学院，北京 100102
2. 北京大学数学科学学院，北京 101871

摘要：中药四气五味理论是在阴阳五行哲学思想指导下，根据药物作用于机体后的生物学表现建立起来的药性核心理论。中药理论与药物的具体功效密不可分。本研究以经典中医学《神农本草经》中的365味中药性味及药效记载为数据源，在建立气-味-效三维数据立方体的基础上，运用关联规则挖掘中的 Apriori 算法，寻找气-味-效三者之间的频繁模式和强关联规则并进行分析，尝试着理清中药气、味、效之间的关系，为中药气、味、效的相关性研究提供新方法和新思路。

关键词：中药性味，四气五味，《神农本草经》；数学；Apriori 算法；中药药理学

中药四气五味药性理论是中医药学理论的重要组成部分，是中医临床经验与中国传统哲学思维相结合的科学药学理论，具有重要的理论价值和实践价值。一般认为，四气五味药性理论是在阴阳五行哲学思想指导下，根据药物作用于机体后的生物学表现（包括寒热反应、视觉味觉感性认识、作用趋向性等）归纳命题并初步建立起来的，其具体内容是对医疗实践经验的提炼，是经得起反复验证的，是中医药理论中占核心地位。随着中医药的现代化、四气五味理论的现代研究越来越成为国内外学者关注的热点，而现代药物化学同药性理论的联系尚缺乏可以遵循的规律[1]。在此背景下，以传统四气四性确定方法中占主导地位的药效反推法为切入点[2]，基于本草文献中大量有关中药药性与功效的记载，探讨四气五味与具体功效之间的关联关系就成为重要的突破口，而运用数学思想和方法进行数据挖掘对于解决该问题大有裨益[3]。本研究选取经典医学著作《神农本草经》（以下简称《本经》）中的365味中
药，在建立气、味、效三维数据立方体的基础上，运用关联规则挖掘中的 Apriori 算法，寻找气味效之间
的频繁模式和高关联规则并进行初步分析，旨在为中药四气五味药性理论研究提供新思路和新方法。

1 研究方法

1.1 气味效三维数据立方体的建立 数据立方体是对多位数据模型进行观察和建模的数据处理方
式[6]，它可以通过有效计算多个维集合上的聚集达到数据分析的目的。针对本草文献中药物记载的文
字特点，诗将其合成为气、味、效三维数据立方体形式。具体过程如下：以 Q 代表四气，W 代表五味
维，E 代表功效维，各维数据的特征属性项（变量）数目可以不同，依实际需要确定。建立三维数据表，
列代表气、味、效各维的特征属性项，行代表药物，每味中药视为一条记录。同时，运用布尔变量 0 和
1 进行药物定位，0 代表“是”，1 代表“非”。定位完成后，统计四气相同，或五味相同，或四气相同
的药物信息，将三维数据表转化为三维数据立方体形式。

根据以上原则，参考学苑出版社 2010 年版的《本草经》辑本（公元 1844 年，武陵山人遗
书第四册）为底本，并与《本草纲目》（公元 1799 年），《本草经集注》、《神农本草经》、《证类
本草》、《千金方》等进行对校（《神农本草经》[7]，完成了 365 味中药气、味、效三维数据的录入，得到数据记录 365 条，并建立气、味、效三维数据立方体（图 1）。

需要指出的是，第一，《本经》中的部分中药已经不为临床所用，甚至不知为何物，如白兔粪[8]；第二，《本经》中许多功效记载主要为道家炼丹成仙所乐
道，如“轻身”、“延年神仙”；第三，《本经》中许多功效记载涉及“精魅邪鬼”、“蛊毒”等古代人们对精神或
奇异疾病的朴素认识。虽然存在以上情况，但鉴于其完整的四气五味及功效记载，仍然将其纳入数据
立方体中。

1.2 数据挖掘方法的准备

1.2.1 关联规则介绍 关联规则主要反映了事物之间的真实性[6]。在数据集中，若大量数据有具有特
征属性 A 的同时，也频繁出现特征属性 B，则称特征属性 A 和 B 构成频繁模式[6]，表示 A 和 B 之间的关
联性，而这些模式可以用来关联规则来观察和分析。

关联规则的质量由规则的支持度和置信度来度量。设 D 为数据记录的集合，I 为特征属性项的集
合，其中每个记录 T 拥有一个标识符，并且 T ⊆ I。对于特征属性项集 A 和 B，其中 A ⊆ I，B ⊆ I，关联
规则可以记为[6]：

\[ A \Rightarrow B \] （1）

其中，支持度 s 是 D 中记录包含 A ∪ B（即同时具有特征属性项 A 与 B）的百分比，它是概率 P
（A ∪ B），即：

\[ \text{Support} (A \Rightarrow B) = P (A \cup B) \] （2）

依上式可知，

\[ \text{Support} (A \Rightarrow B) = \text{Support} (B \Rightarrow A) \] （3）

置信度 c 是记录包含 A 的同时也包含 B 的百分比，它是条件概率 P (B/A)，即：

\[ \text{Confidence} (A \Rightarrow B) = P(B/A) = \frac{\text{Support} (A \Rightarrow B)}{\text{Support} (A)} \] （4）

依上式可知，

\[ \text{Confidence} (A \Rightarrow B) \neq \text{Confidence} (B \Rightarrow A) \] （5）

同时满足最小支持度阈值和最小置信度阈值的规则称为强关联规则。为方便计算，采用最小支持度
计数代表最小支持度阈值，以自然数表示；采用最小置信度百分比代表最小置信度阈值，以百分比表示。

1.2.2 运用 Apriori 算法寻找强关联规则 Apriori 算法是 R. Agrawal 和 R. Srikant 于 1994 年提出
的，为布尔型关联规则挖掘模式模式的原创性算法[1]，因其有效的挖掘能力和优良的进化能力而被
广泛应用于医药卫生领域[8-11]。该算法利用一种逐层筛选的迭代方法，通过“连接”和“剪枝”在大量
数据挖掘结果中寻找令人感兴趣的关键规则。首先，通过全扫描数据库，得到特征属性项的总数为 k。
收集满足最小支持度计数的项，形成频繁 1-项
集，记作 L₁，将其视为第一次“剪枝”。接着，执行 L₁ 与自身的“连接”以产生候选 2-项集合，产生 L₂ 的超集，记作 C₂。然后，扫描 C₂，如果某候选 2-项（集）不在 L₂ 中，则该候选 2-项（集）也不可能是频繁的，将其从 C₂ 中删除，实现“剪枝”。接着以 L₂ 自身“连接”形成 C₃，如此下去，直到找到包含所有特征属性项的频繁 k-项集合。最后，根据最小置信度百分比由频繁项集产生强有力关联规则。

1.3 数据预处理
数据预处理是尽量减少数据不一致性和不完整性的干扰因素，提高数据质量，保证数据挖掘结果合理有效的关键[4]。本研究的数据预处理需要解决以下几个问题。

1.3.1 数据清洗
四气五味记载缺失的药物应予以排除。本研究选取的以清·顾观光草本为底本的《本经》中气味缺失的药物共有 11 种，包括柿子、文蛤、铁精、铁、锡、桑、大枣、鼠李、杨梅、黄芩、牛子。

部分药物记载了同一药用植物（动物）的不同部位，如核桃仁下还包括花纹、桃仁、桃毛、桃酸、白马茎下还包括眼、散步等，由于整条记载只描述了一次四气五味，故只纳入首药的数据。

个别药物记载有不同气味的属性，如紫参下“苦辛寒”，夏枯草下“苦温，辛寒”，均因与《本经》校对后取苦辛，舍辛味。

除去气味缺失的 11 条记录，完成数据清洗后三维数据表中记录条数为 354 条。

1.3.2 数据整合
1.3.2.1 合词
《本经》大约成书于东汉时期[5]，是东汉以前各地区医药学家经验性的总结，故具有早期本草的草创特征[7]，表现为记载有大量药物别名及功效的异名同实，以及由于文字俚俗、通假、缺失带来的语言理解困难。故本研究参考《本经》中总论的内容，在基于古汉语文字的医药词义分类的基础上，参考《中医内科学》[12]、《中医外科学》[13]、《诸病源候论》[14] 和《中医妇产科学》[15]，又尝试尽可能地对相似功效的不同表述方式进行归类，达到整合数据的目的[16]。整合后功效词共有特征属性项 134 条[16]。

1.3.2.2 分词 同时，对于《神农本草经》中部分出现频度很高，含义不确定的词汇，采用语言学的处理方法，如“寒热”和“邪气”两个词。由于四气药性中也使用寒、热、温、凉，故对其特别关注，但是准确理解《本经》中出现频率很高的“寒热”一词的含义就存在很大难度。其一，“寒热”一词本身由相反性质的两个字构成，直接反推药性会令人疑惑。其二，“寒热”一词具有多种含义。《中经辞海》记载其具有 3 种含义[17]，分为诊脉术语（根据脉象阴阳偏盛衰的两种不同性质的证候）、病症名（即寒热发热的统称）、《灵枢》篇名（论述瘀热），而《中医名医术语精华词典》中更是有 7 种含义[18]。其三，即使作为病症名，“寒热”的真正内涵也若隐若现。清代著名医家徐大椿曾这样评价，“此种寒热，既非感寒，亦非伤寒。其浅深有骨髓皮肤之殊，其久暂有岁月之异。轻者似有感冒，重者即变骨蒸。所以《内经》之后，诸书寒热，均有方论……寒热之因，千变万状。有属外感，有属内伤，外伤内伤之中，又各不相同。其外有属痰饮，有属瘀血，有属积聚，有属虚损，不可胜举”[19]。

通过参考医学文献[14, 19, 20]，借鉴汉语语法中对于定语概念及其作用的认识[21]，本研究将“寒热”和“邪气”二词作分词处理。见表 1。

1.3.3 数据编辑
在气味-效四气三维数据立方体中，四气维包含特征属性项 7 项，分别为寒、微寒、小寒、平、微温、温和大热。其中，“小寒”仅出现于曾青下 1 次，“大寒”仅出现于附石下 1 次，并因关联规则算法中最小支持度计数的需要，将“小寒”归入“微寒”项下，舍去“大热”项，同时减去“附石”的记录。因此，四气维共有 5 个特征属性项，分别为寒、微寒、平、微温、温和而数据立方体中记录条数也成为 353 条。

本研究重点在于分析四气与功效、五味与功效及四气合五味与功效之间的关联关系，故对于式 (1)，虽然理论上 A 和 B 均可以是若干特征属性项的组合，但本研究中 A 表示为独立四气维，独立五味维，四气维合五味维的特征属性项组合，而 B 只取一个简单项，以达到简化运算提高效率的目的[3]。因此，研究中选择性地分析气味-效三维数据立方体中的 QX、WX、QWX 三个数据单元，而关联规则的一般形式也可以表示为：

Q ⇒ X 或 Q ⇒ X
W ⇒ X 或 W ⇒ X
(Q ∧ W) ⇒ X 或 X ⇒ (Q ∧ W)

数据预处理后得到的三维数据表中总记录数为 353 条，三维数据立方体中，四气维具有特征属性项（变量）5 项，五味维具有特征属性项（变量）5 项，功

<table>
<thead>
<tr>
<th>名称</th>
<th>类型 1</th>
<th>类型 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>寒热</td>
<td>“寒热”单独出现，表证</td>
<td>“寒热”表征（如脉细，舌干，腹中热，皮肤）</td>
</tr>
<tr>
<td>邪气</td>
<td>“邪气”单独出现成为“寒热邪气”，致病因素</td>
<td>“邪气”致病（如心腹冷，心下痛）</td>
</tr>
</tbody>
</table>

表 1 特殊词汇处理表
效用具有特征属性项（变量）134 项。同时根据实际需要，本研究在 QX、WX 数据单元分析中，设定最小支持度计数为 6，最小置信度百分比为 45%；而在 QWX 数据单元分析中，由于约束条件更为严格，为了获得足够数量的频繁模式，将最小支持度计数减小为 3，将最小置信度百分比减小为 40%。

1.3.4.1 频繁 1-项集产生  在 Apriori 算法中，不能满足最小支持度的项不能构成频繁模式，故分别对上述特征属性项进行最小支持度的检验。四治中的寒、平、温 3 个特征属性的支持度计数分别为 97.128、81，均满足条件；功效维中“轻身延年”、“除热”、“欲逆上气”、“金疮”、“黑子（痣）”5 个特征属性进行计算示例。

1.3.4.2 连接 根据研究实际需要，分别将频繁 1-项集中属于四治维的 3 项与属于功效维中的 4 项两两配对（而不需要属于四治维的 3 项之间或属于功效维的 4 项之间的配对），进行“连接”。得到容量为 12 的候选频繁 2-项集合。例如：寒-轻身延年、平-轻身延年等。

1.3.4.3 找到强关联规则 分别计算 12 对特征属性对之间的 Q->X 和 X->Q 形式的置信度 c，得到 24 个相应的置信度值。以“寒-轻身延年”为例。

在数据立方体的四治维定位“寒”特征属性（在算式中以 H 代表），功效维定位“轻身延年”特征属性（在算式中以 QSYN 代表），计算其对应的置信度。（五味维特征属性项全部计入），得到 26 条符合条件的记录（即有 26 个寒性中药具有轻身延年的功效）。故，在 Q->X 形式中，

\[
\text{Confidence (H} \Rightarrow \text{QSYN}) = \frac{\text{Support (H} \Rightarrow \text{QSYN})}{\text{Support (H)}} = \frac{26}{97} = 0.268
\]

在 X->Q 形式中，

\[
\text{Confidence (QSYN} \Rightarrow \text{H}) = \frac{\text{Support (QSYN} \Rightarrow \text{H})}{\text{Support (QSYN)}} = \frac{26}{133} = 0.196
\]

计算所有 12 个满足特征属性项对条件的记录数，将其列于三维数据立方体内（图 2）。由于示例中只有 2 个特征属性，故其只是呈现在三维平面中。

![图 2 示例数据立方体展示](image)

基于此，依次计算 24 个置信度值，转化成百分比，见表 2。

<table>
<thead>
<tr>
<th>功效</th>
<th>气→效</th>
<th>气→气</th>
</tr>
</thead>
<tbody>
<tr>
<td>寒</td>
<td>轻身延年</td>
<td>26.8%</td>
</tr>
<tr>
<td>除热</td>
<td>30.9%</td>
<td>11.7%</td>
</tr>
<tr>
<td>欲逆上气</td>
<td>6.2%</td>
<td>10.2%</td>
</tr>
<tr>
<td>金疮</td>
<td>1.0%</td>
<td>3.9%</td>
</tr>
</tbody>
</table>

表 2 算法演示结果表

数据挖掘 数据及算法确定后，运用 MATLAB2010a 软件编程进行计算。

2 结果

与 1.3.3 节相对应，数据挖掘结果也分为 QX、WX、QWX 这 3 部分进行说明与分析。由于篇幅原因，结果部分先列出置信度的频次统计，再逐列列出。

2.1 四气与功效的关联规则挖掘 表 3 中四气与功效的关联规则挖掘结果显示，以 45%为置信度阈值，Q->X 形式的强关联规则有 1 条，X->Q 形式的强关联规则有 38 条。表 4 中按照置信度从大到小排序将其逐一列出。

2.2 五味与功效的关联规则挖掘 表 5 中五味与功效的关联规则挖掘结果显示，以 45%为置信度阈值，Q->X 形式的强关联规则有 3 条，X->Q 形式的强关联规则有 46 条。表 6 中按照置信度从大到小排序将其逐一列出。
### 表 3 四气与功效的关联规则置信度统计

<table>
<thead>
<tr>
<th>四气</th>
<th>效→气</th>
<th>效→气</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;45%</td>
<td>45%～35%</td>
</tr>
<tr>
<td>温</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>微温</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>平</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>微寒</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>寒</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### 表 4 四气与功效的强关联规则表

<table>
<thead>
<tr>
<th>四气</th>
<th>效→气</th>
<th>气→效</th>
</tr>
</thead>
<tbody>
<tr>
<td>温</td>
<td>—</td>
<td>效→气</td>
</tr>
<tr>
<td>微温</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>平</td>
<td>平→轻身延年（48.4%）</td>
<td>效→气</td>
</tr>
<tr>
<td>微寒</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>寒</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

“—”表示没有挖掘出满足条件的强关联规则。

### 表 5 五味与功效的关联规则置信度统计

<table>
<thead>
<tr>
<th>五味</th>
<th>株→效</th>
<th>效→株</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;45%</td>
<td>45%～35%</td>
</tr>
<tr>
<td>酸</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>苦</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>甘</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>辛</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>咸</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### 表 6 五味与功效的强关联规则表

<table>
<thead>
<tr>
<th>五味</th>
<th>效→株</th>
</tr>
</thead>
<tbody>
<tr>
<td>酸</td>
<td>—</td>
</tr>
<tr>
<td>苦</td>
<td>—</td>
</tr>
<tr>
<td>甘</td>
<td>—</td>
</tr>
<tr>
<td>辛</td>
<td>—</td>
</tr>
<tr>
<td>咸</td>
<td>—</td>
</tr>
</tbody>
</table>

“—”表示没有挖掘出满足条件的强关联规则。
2.3 四气合五味与功效的关联规则挖掘 在QWX数据单元分析中，由于约束条件更为严格，产生了大量的低置信度关联规则（平均置信度在6%左右）。为了获得足够数量的频繁模式，故将最小支持度计数减小为3，将最小置信度百分比减小为40%。表7中按照置信度从大到小将所有强关联规则逐一列出。

表7 四气合五味与功效的强关联规则表

<table>
<thead>
<tr>
<th>四气</th>
<th>功效</th>
<th>(气 A味 B) (\Rightarrow) 效</th>
<th>效 (\Rightarrow) (气 A味 B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>温</td>
<td>甘温⇒生津延年（57.1%），甘温⇒益精气（50.0%）</td>
<td>温中⇒辛温（54.5%），风头底病⇒辛温（41.2%）</td>
<td></td>
</tr>
<tr>
<td>微热</td>
<td>甘微温⇒生津延年（60.0%），甘微温⇒益精气（60.0%），辛微温⇒生津延年（44.4%），酸微温⇒酸积积聚（68.7%）</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>平</td>
<td>酸平⇒生津延年（57.1%），酸平Rightarrow益精气（42.9%），酸平Rightarrow病病（42.9%），甘平⇒生津延年（84.2%），甘平Rightarrow益精气（50.0%）</td>
<td>不饥⇒甘平（51.9%），安神益气⇒甘平（42.9%）</td>
<td></td>
</tr>
<tr>
<td>微寒</td>
<td>苦微寒Rightarrow除热（52.9%），甘微寒⇒生津延年（100.0%）</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>寒</td>
<td>苦寒⇒生津延年（73.3%），苦寒⇒寒热（30.0%），苦寒Rightarrow益精气（40.0%），英寒Rightarrow除热（47.4%），英寒Rightarrow病病（42.1%）</td>
<td>干燥积积积气⇒苦寒（68.7%），消渴⇒苦寒（55.6%），黄连⇒苦寒（54.5%），酸积积聚积聚（45.5%），身面浮肿⇒苦寒（42.9%）</td>
<td></td>
</tr>
</tbody>
</table>

“—”表示没有挖掘出满足条件的强关联规则。

3 讨论

3.1 对本研究关联规则挖掘结果的解释说明

3.1.1 四气与功效的关联规则挖掘 对于特征属性A和B，如果A⇒B为强关联规则，那么事物是否具有A属性的同时，很有可能也具有B属性。但是反之则不成立。在本研究中，当A为四气之气或五味之味，B为功能时，关联规则A⇒B提示了特定气药的作用特点。符合A为功效B，B为四气之气五味之味，关联规则A⇒B提示了实现此类功效的药物具有的特定气味，即该类功效可以表达相应的药性，这种关联规则对于寻找判别药性的功效具有重要意义。

根据表3和表4可知，Q⇒X形式的关联规则明显少于X⇒Q形式，说明温性(或寒性)药物可以具有很多种功效，并无特定的作用倾向。只有平性药多具有生津延年的作用倾向。同时，不同的药物通过不同的功效与特定的寒热药性相关联，即不同的药物可能通过不同的功效表达不同的药性。例如，药物可以通过气温性，或气性，或治疗风头底病表达温性，可以通过或补中，或坚筋骨，或养精神表达平性，可以通过或除寒，或利水道，或治疗烦满表达寒性。

另外还可知，表达平性的药物功效多为补益作用，包括养精神，安神，和五脏，益精气，补中，坚筋骨，强志，肥健不饥和生津延年等，这些相类似的补益功效为《本经》中上品所具备，可以划归为一大类，作为与平性药密切相关的功效群。同时，表达寒性的药物也具备两大功效群，一类与热性疾病相关，包括除热，治疗消渴，目赤痛和烦满等；一类与水液代谢相关，包括利水道，消水，治疗消胀，身面浮肿，淋病和泄泻等。以上结论与现代中医临床认意识基本相符。

3.1.2 五味与功效的关联规则挖掘 根据表5和表6可知，W⇒X形式的关联规则明显少于X⇒W形式。同理，说明苦味(或辛味)的药物也可以具有很多种功效，无特定作用倾向。只有甘味药多具有生津延年、益精气的作用。同时，不同的药物通过不同的功效与特定的五味相关联。例如，药物可以通过或除寒，或消水，或治疗痰饮疮疡表现在苦味特性，可以通过或补中，或清热，或治疗虚劳赢瘦表现在甘味特性，可以通过或助消化，或治疗咳嗽上气，或治疗恶肉死肌表现辛味特性。

另外还可知，表达苦味特性的功效群有3大类，一类与热病相关，包括除热，治疗消渴，目赤痛和疮疡疮疡；一类与水液代谢相关，包括消水，治疗身面浮肿和胀气；一类与留滞结固的治疗相关，包括消痰塞结饮食、治疗结气和消肿瘤瘤等。表达辛味特性的功效群也有3大类，一类与气机不畅疾病相关，包括气滞，治疗胸闷和咳逆上气；一类与祛除风邪相关，包括治疗中风，身痒和头风脑动；一类与治疗精魅邪鬼所致疾病相关，包括杀虫，通神明，辟恶气，治疗疮疮等。而表达甘味特性的功效群则与表达平性的药物功效类似，基本为补益类功效群。以上结论与现代中医临床认意识基本相符。

3.1.3 四气合五味与功效的关联规则挖掘 根据表7可知，在(Q A W)⇒X形式的强关联规则中，
许多气味组合均表现出轻身延年的作用倾向，包括甘温、甘微温、辛微温、甘平、酸平、甘微寒、甘寒，这与独立的四气、五味与功效关联规则挖掘时产生的强关联规则（\(\text{平} \Rightarrow \text{轻身延年，甘} \Rightarrow \text{轻身延年} \)）相符。

另外，对于四气、五味中的三种常见搭配辛温、甘平和苦寒，\(X \Rightarrow \{Q \land W\}\)形式的强关联规则提供了合理的表达这些药性的几个功效。而这为以功效推断药性提供了依据，即如果一味中药具有或温中、或治疗风头头痛的功效，那么其药性很有可能为辛温；如果一味中药具有或入骨、或破气不消的功效，那么其药性很可能为甘平；如果一味中药具有或透脓、或散瘀的功效，那么其药性很有可能为苦寒。以上结论与现代中医临床认识基本相符。

3.2 数学思想方法的引入为中药药性理论深入研究提供支持

3.2.1 有利于合理认识中药药性与功效的关系

本研究结果显示，不同的药物可能通过不同的功效表达相同的药性，且在研究具有相同药性、多个不同药物的药性表达时，不宜直接认为它们会具有相似的药性，或会纠正相似的病理状态，或会引起某个生化指标相似的变化趋势。而不应进行充分的药物针对性考量。这与国内药性研究认识前沿相符。中药药性与功效表现出复杂的离合关系，在掌握药物共性的基础上，认识中药的药性尤为重要。

如表5所示，表达药性中的药性有温中、虚风头、虚痛、主散逆气等5种，查阅《本经》中记载可知，温性药吴茱萸、辛、辛、麻黄等是很相似的这种不同药性表达药性；同样，表达寒性的药性有煮水、主消渴、主调赤等15种，而《本经》中的寒性药滑石、知母、黄连也很可能是分别通过这三种不同功效表达寒性，另外，同样的现象也出现于五味的研究结果中。

3.2.2 有利于理解药性理论思维形成过程的特点

根据表4、表6和表7的结果，与不同的四气五味对应的强关联规则数量有较大不同，与微寒、微温、酸和咸相关的关联规则较少出现，表现了四气五味出现频率之间的不均衡性。这表明，虽然古代医家依据阴阳五行学说确定了气与味的名称和种类，即四气与五味，但其在药物的药性归属上却并不平均一致，而是特殊求是地“按需分配”，所以四气五味主要还是长期用药经验的反映，是可以经得起重复验证的性质，认识到这一点对于明确中药药性理论的科学价值具有重要意义。

同时，也要认识到药性理论形成过程中缺乏如西方形式逻辑学般严密的——对应，而是具备独特的非形式逻辑思维的特点，包括无限制演绎、不全归纳等，这体现在本研究结果上，例如。

辛微温→轻身延年；甘平→轻身延年

又如，在QX数据单元的关联规则分析中存在这样的置信度接近阈值（45%）的关联规则：

疟疾→平；疟疾→温（置信度均为42.9%）

疟疾→平；疟疾→温（置信度均为44.4%）

这些关联规则表明，某些时候一个功效可以无倾向性地推断得到不同的药性，即部分功效与药性相对独立，突显出中药药性理论形成过程的复杂性，这在药性研究中需要充分关注。国内学者主张在理解中国传统文化的“象思维”[24-25]，延续传统中药药性理论的同时，也应当学会理性剖析，剥离出直觉类的猜测和推理，在传统药性理论与现代化学研究之间保持必要的张力[26]。

3.2.3 有利于深入理解《本经》中病证的寒热属性

通过考察药性与药性之间的关系，可以深入理解《本经》中病证的寒热属性。例如，表4中显示“暑病”寒（50.0%），即治疗暑病的药物很大部分为寒性，而医学认为暑病是伤风热所致[27]，故用寒性药物对症治疗是合适的。又如，表7中显示“寒热”寒寒（31.0%）、寒寒（0.0%），即经过特殊饱和处理的“寒热”与寒热性药性关联性较强，根据这样的结果判断，寒寒（寒寒表）更多时候所表示的可能是一种热性疾病，这与《难经》中对寒病“毛发燥”、“口白如盐”等的描述相符。若过于重视表寒热病“入不寒伤等法，大段易营中之热为主”[28]，则在药用的寒属性提供依据的同时，也体现了本研究的合理性。

3.2.4 有利于正确认识中药药性理体系的科学价值

以上研究结果显示，中药的四气五味药性理论与具体功效存在复杂相关性。某些情况下，特定的气与味与特定的功效相联系，某些情况下，药性与功效也表现出相对独立的特点。而造成这种复杂性的作用可能是药用植物本身的多样性，以及其在植物体内的相互作用以及观察手段的局限。而且，四气五味药理理论发展到今天，其内容也受到历史上各时期的社会发展特点的影响。因此，对于中药药性理论体系的科学认识要注意两个方面，其一，明确中药药性理论体系的重要价值，其二是明确中药药性理论体系的复杂性。
论是在特定历史环境下产生的面向实践的开放性理论体系，具有朴素唯物主义和非形式逻辑思维特点，这些与现代药学不同的理论内涵，为中医辨证论治提供了合适的理论指导，中华民族的几千年的医疗活动实践证明其具有重要价值。同时，针对疾病复杂性问题，在现代医学运用逻辑性数学模式进行个体化治疗的方法以外，基于中国传统“象思维”和“术数学”的中药药性理论体系和中医辨证论治体系可能是有效的解决方案。有专家预测，如果以世界科学共享的“药物基因组学”，遗传药理学以及分子生态学等的理论成果和实验成果为现代依据，那么中医药学从理论上和实践上的有机结合，将会成为完整的复杂科学体系[27]。

3.3 本研究有待完善之处 本研究采取的经典Apriori算法符合布尔型文本数据的关联规则挖掘需求，数据挖掘结果可分析性强，较中医药药性理论认识相符，具有一定的现实意义。同时，尚有以下几个有待完善。例如，最小支持度阈值的设定使得许多出现次数较的少数并未纳入关联规则的遴选中，而是直接被“剪枝”处理。其中包括“出汗”、“主面赤热”、“主手干舌热”、“主四肢寒厥”等很有可能与寒热药性相关的有效词汇。又如，挖掘结果中一些有效的相伴出现使得数据整合和精简也可作进一步调整，包括“益子精”与“主绝子”、“轻身延年”和“益弱精”以及“胀胀”和“身面浮肿”等。

3.4 展望 本研究从《本经》365味中药气味效三维数据立方体的关联规则挖掘入手，探索性地分析了四气五味药性与药物具体功效之间的关联关系，并对数据挖掘结果给予初步解释。鉴于中药药性理论继承性研究的重要性，针对药性复杂问题，未来研究可以选取更多的中药、更广泛的功效群，进行高通量分析，同时也应兼顾出现频率较低的功效，为探讨中药药性实质提供思考。

REFERENCES

A study of association rules in three-dimensional property-taste-effect data of Chinese herbal medicines based on Apriori algorithm

Rui Jin¹, Qian Lin², Bing Zhang³, Xin Liu⁴, Sen-mao Liu⁵, Qian Zhao⁶, Xiu-lan Liu⁷
1. School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100102, China
2. School of Mathematical Sciences, Peking University, Beijing 101871, China

Abstract: The theory of four properties (Qi) and five tastes (Wei) is the core of the property theory of Chinese materia medica. It is known that Qi and Wei are associated with the pharmacological effects (Xiao) of herbs. This study took records of all 365 Chinese herbs in Shenong’s Classic of Materia Medica (Shennong Ben Cao Jing) as the data resource and established a three-dimensional data cube, in the purpose of finding out and analyzing the frequent patterns and valued association rules of Qi, Wei and Xiao based on Apriori algorithm. The results of this study may give rise to innovative ideas and methods in research of traditional Chinese materia medica.

Keywords: properties and flavours; 4 natures and 5 flavours; Bencaojing; mathematics; Apriori algorithms; Chinese medicine property theory