基于莱文斯坦距离的易混淆药品目录自动生成算法及软件实现Δ

陈　杨*，淡重辉，何　瑶，阮　一，陈　肖，郑晓媛#（重庆市急救医疗中心药剂科，重庆　400014）

中图分类号　R952　　**文献标志码**　A　　**文章编号**　1001-0408（2024）15-1899-06
DOI　10.6039/j.issn.1001-0408.2024.15.17

摘　要　目的　构建一种高效的易混淆药品目录自动生成算法，并在此基础上开发一种易混淆药品目录管理系统，以此提升对易混淆药品目录的管理效率。**方法**　本研究以莱文斯坦距离算法为理论基础，深入研究易混淆药品组的自动识别机制以及相似性阈值筛选方法，进而构建易混淆药品目录自动生成算法。在系统开发层面，本研究采用Visual Basic.NET作为编程语言，结合SQL Server 2008 R2 Express数据库管理平台，设计开发易混淆药品目录管理系统。**结果**　相似性阈值δ是易混淆药品目录自动生成算法的关键参数，随着δ的逐渐增大，易混淆药品的总数逐渐减少，而易混淆药品组数则呈现先上升后下降的变化规律。在实际应用中，可根据药品的通用名或品种名构建易混淆药品目录，对应的相似性阈值可取0.75和0.83。此外，本课题组开发的易混淆药品目录管理系统将原本耗时约1周的目录建立时间缩短至不到1 h，极大地提升了工作效率。**结论**　本研究构建的易混淆药品目录自动生成算法高效快速，为易混淆药品管理提供了强有力的技术支撑。开发的易混淆药品目录管理系统极大地减轻了目录建立和维护的时间成本，提高了对易混淆药品目录的管理效率。

关键词　易混淆药品；莱文斯坦距离；药品目录；相似性阈值

Software implementation of Levenshtein distance-based algorithm for automatically generating easily confused drug catalogs

CHEN Yang, DAN Chonghui, HE Yao, RUAN Yi, CHEN Xiao, ZHENG Xiaoyuan (Dept. of Pharmacy, Chongqing Emergency Medical Center, Chongqing 400014, China)

ABSTRACT　OBJECTIVE To create a highly effective algorithm for automatically generating easily confused drug catalogs (ECDC), as well as to develop a management system for ECDC based on this algorithm, in order to improve the management efficiency of ECDC. **METHODS** This study, based on Levenshtein distance algorithm, delved deeply into the automatic identification mechanism of easily confused drugs and the screening method for determining similarity thresholds, ultimately leading to the development of an algorithm for automatically generating ECDC. Besides a management system was designed and developed, using SQL Server 2008 R2 Express as the data storage platform and Visual Basic.NET as the programming language. **RESULTS** The similarity threshold δ played a crucial role in the algorithm for automatically generating ECDC. As the value of δ gradually increased, the total count of easily confused drugs decreased gradually, while the count of drug groups exhibited a pattern of initially increasing and then decreasing. Practically, ECDC could be created using either the generic or varietal names of drugs, with corresponding similarity thresholds of 0.75 and 0.83. Furthermore, ECDC management system had significantly reduced the time required to establish a catalog from about one week to less than one hour, resulting in a substantial enhancement in work efficiency. **CONCLUSIONS** The algorithm used to automatically generate ECDC is highly efficient and rapid, offering robust technical assistance for the management of easily confused drugs. Implementing the ECDC system can greatly reduce the time cost related to building and maintaining the catalogs, thus significantly improving the efficiency of managing ECDC.

KEYWORDS　easily confused drugs; Levenshtein distance; drug catalogs; similarity threshold

易混淆药品指的是在外观（看似）、读音（听似）上相似，或存在一品多规、多剂型、多厂家等情形，容易引发混淆的药品[1-2]。这类药品是引起调配差错的重要原因[3-4]。据估计，美国每年与用药错误相关的费用高达数百万美元，其中听似和看似药品导致的用药错误分别

Δ**基金项目**　重庆市卫生健康委医学科研项目（No.2024WSJK030）；重庆市临床药学重点专科建设项目（No.渝卫办发〔2023〕2号，No.渝卫办发〔2023〕69号）

＊**第一作者**　主管药师，硕士。研究方向：医院药学、化学计量学。E-mail：yangchen8786@sina.com

\# **通信作者**　主任药师，硕士。研究方向：药理学、医院药学。E-mail：thymolblue@163.com

占总用药错误的25%和33%[4-5]。建立易混淆药品目录是减少易混淆药品调配差错的重要措施[6-7]。然而，当前大多数医院的药品数已逾千种[2,8]，仅靠人力维护易混淆药品目录不仅效率低下，而且极易出现疏漏。因此，利用先进的技术手段来辅助管理和维护易混淆药品目录就变得尤为重要。

易混淆药品引起混淆的主要原因在于药品名称中包含的文字元素具有高度相似性。因此，为了衡量药品名称之间的相似性，需要借助一种可靠的度量工具。莱文斯坦距离（Levenshtein distance，LD）作为一种编辑距离，在计算机科学中常用于量化2个字符串之间的相似性[9]。具体而言，LD是指将一个字符串转变为另一个字符串所需的最少单字符编辑操作次数。这些操作包括插入、删除和替换单个字符。LD在多个领域都有重要应用价值，如DNA分析、拼写检查、自然语言处理以及抄袭侦测等领域。在药学领域，尽管LD已被应用于筛选听似药品[10]，但尚未建立起易混淆药品组的自动识别方法，同时也缺乏对易混淆药品相似性阈值筛选过程的深入分析与研究。

为了增强药品使用的安全性，降低因药品名称混淆而引发的潜在风险，本课题组拟以LD算法为理论基础，探索易混淆药品组的自动识别机制和相似性阈值筛选方法。通过该研究，期望构建出一种高效的算法，以自动且精准地生成易混淆药品目录。为验证算法的实用性，本课题组拟在Visual Basic 2010 Express和SQL Server 2008 R2 Express平台上开发易混淆药品目录管理系统，提升对易混淆药品目录的管理效率。

1 资料与方法

1.1 资料来源

门诊药房和住院药房的药品目录均来自医院信息系统（hospital information system，HIS）。药品目录内包含了药品的唯一识别码、名称、规格、生产厂家和计数单位等重要信息。从HIS导出的药品信息仅限于在库药品。对于无库存的药品，因其不存在引起药品混淆的风险，故不需要导出。

1.2 易混淆药品目录的自动生成算法

1.2.1 药品名称的规范化处理

从HIS导出的药品名称中时常包含备注信息。这些备注信息通常被标记在一对括号之中，且位于药品通用名的最前端或最后端。鉴于备注信息并非药品名称的组成部分，需将其剔除。

药品通用名中的剂型或给药途径等信息，在特定情境下，可能对易混淆药品的识别产生一定干扰。为了排除这些潜在的影响因素，在得到药品通用名后，可以进一步剔除剂型和给药途径等信息，从而得到更利于易混淆药品识别的药品品种名。需要特别指出的是，有少数药品的通用名并不包含明确的剂型信息，如开塞露、强力枇杷露和破伤风抗毒素等。对于这类药品，可直接使用通用名作为药品的品种名。

1.2.2 药品名称的相似性度量

不管是药品的通用名还是品种名，本质上都是字符串。在计算机领域，常采用LD来衡量2个字符串的相似性。基于LD计算药品名称相似性的算法如下。

给定2个药品的名称字符串，以a和b表示，其长度分别记为$|a|$和$|b|$。那么计算a和b之间编辑距离的LD模型可表述为：

$$\mathrm{Lev}_{a,b}(i,j)=\begin{cases}\max(i,j) & \text{当}\min(i,j)=0\\ \min\begin{cases}\mathrm{Lev}_{a,b}(i-1,j)+1\\ \mathrm{Lev}_{a,b}(i,j-1)+1\\ \mathrm{Lev}_{a,b}(i-1,j-1)+\varepsilon\end{cases} & \text{当}\min(i,j)\neq 0\end{cases}\cdots(1)$$

其中，$\mathrm{Lev}_{a,b}(i,j)$表示a的前i个字符与b的前j个字符的编辑距离；ε为逻辑函数，取值为：

$$\varepsilon=\begin{cases}0 & \text{当}a_i=b_j\\ 1 & \text{当}a_i\neq b_j\end{cases}\cdots\cdots(2)$$

其中，a_i和b_j分别表示a的第i个字符和b的第j个字符。LD模型的计算过程通常记录在一个$(|a|+1)\times(|b|+1)$的二维数组L中，即$L(i,j)=\mathrm{Lev}_{a,b}(i,j)$。值得注意的是，$L$的索引从0开始，而不是1。LD模型的计算，从$i=0$、$j=0$开始，直到$i=|a|$、$j=|b|$结束。最终得到的$\mathrm{Lev}_{a,b}(|a|,|b|)$或$L(|a|,|b|)$，即为药品名称$a$与药品名称$b$的编辑距离。

编辑距离$\mathrm{Lev}_{a,b}(|a|,|b|)$只反映了编辑所需的次数，未包含字符串的长度信息，因此不能较好反映药品名称之间的相似性。本课题组基于LD进一步定义了衡量2个字符串相似度的指标，即$R(a,b)=1-L(|a|,|b|)/\max(|a|,|b|)$。$R$不仅考虑了字符串长度，而且最大值被限定为1，因此便于对药品名称之间的相似性进行比较。基于LD的药品名称相似性算法代码见图1。

1.2.3 易混淆药品组的自动识别

本课题组采用阈值法来自动识别易混淆药品组。首先，设定1个相似性阈值$\delta\in[0,1]$，用于判断2个药品名称之间的相似程度是否达到了易混淆的标准。接着，将所有药品名称存入一维数组A。最后，从A中识别出所有易混淆药品组。

```
Function LevR(a As String,b As String) As Double
  Dim m,n As Integer
  m = Len(a) :n = Len(b)
  If m = 0 or n = 0 Then Return 0
  Dim L(m,n),i,j As Integer
  For i = 1 To m :L(i,0) = i :Next i
  For j = 1 To n :L(0,j) = j :Next j
  Dim Insert,Delete,Replace As Integer
  For i = 1 To m
    For j = 1 To n
      Insert = L(i,j-1) + 1
      Delete = L(i-1,j) + 1
      If Mid(a,i,1) = Mid(b,j,1) Then
        Replace = L(i-1,j-1)
      Else
        Replace = L(i-1,j-1) + 1
      End If
      L(i,j) = Math.Min(Math.Min(Insert,Delete),Replace)
    Next j
  Next i
  Return 1-L(m,n)/Math.Max(m,n)
End Function
```

图1　基于LD的药品名称相似性算法代码

从A中识别出易混淆药品组的方法如下：首先，选出A中第一个药品名称A_1。然后，将A_1与数组中余下的所有药品名称进行相似度计算。如果计算得到的R不小于设定的相似性阈值δ，则把当前药品名称标记为与A_1相似。遍历完所有药品名称后，如果被标记的药品数＞1，则把A_1与所有标记的药品归为1个易混淆药品组。上述过程仅识别出了1组易混淆药品，为筛选出更多易混淆药品组，需要对A进行放缩，即将A_1和所有被标记的药品名（若存在）从A中删除。然后对放缩后的A重复上述识别步骤，即可得到第2组易混淆药品。以此往复，直到A中仅包含0个或1个药品名称为止，即可筛选出所有易混淆药品组。

1.2.4　相似性阈值的筛选

相似性阈值δ是易混淆药品目录自动生成算法的重要参数。若设定的δ值偏低，可能会导致易混淆药品组的质量不尽如人意，进而影响算法的准确性和实用性；反之，若δ的设定值偏高，则有可能造成某些重要的易混淆的药品被遗漏，从而削弱算法在易混淆药品管理和风险控制方面的作用。因此，合理地设定δ值对于确保算法的有效性和可靠性至关重要。

相似性阈值δ与易混淆药品组数之间存在着紧密关联。为明确该关系，本课题组定义了一个能将相似性阈值δ映射到易混淆药品组数的函数g。通过观察这个函数的图形，可以从整体上把握相似性阈值δ对易混淆药品组数的影响。从实际操作中可以观察到，随着δ的逐渐增大，易混淆药品组数$g(\delta)$的变化趋势表现出先上升后下降的规律，且图形整体呈现出类似钟形曲线的特征。

具有钟形曲线特征的常见函数是正态分布的概率密度函数，该函数的图像具有对称性：从峰值点开始，曲线两侧都呈现下降趋势，且在拐点处的下降速率最快。将该性质类比到函数g上，意味着存在某个特定的相似性阈值δ_c，使得易混淆药品组数最大。当相似性阈值大于δ_c时，易混淆药品组数将逐渐减少，且在函数g的右侧拐点处，记为$\left(\delta_k, g\left(\delta_k\right)\right)$，具有最大的下降速率。

在遴选最佳相似性阈值时，δ_c和δ_k均为可行的选项，但使用δ_k更为合理。原因有如下2点：（1）由函数g的性质可知，$g\left(\delta_c\right)>g\left(\delta_k\right)>g(1)$，表明以$\delta_k$作为相似性阈值时，易混淆药品组数不会出现过多或过少；（2）当相似性阈值超过δ_k后，易混淆药品组数减少的速度放缓，表明各组内的药品名称都具有高度的相似性。因此本研究使用δ_k作为最佳的相似性阈值。

实践中，函数g的图形并非为一条光滑曲线，这不利于确定δ_k。但可以寻找一条拟合曲线来代替函数g，从而将寻找函数g右侧拐点的问题转换为寻求拟合曲线右侧拐点。由于函数g与正态分布的概率密度函数具有相似的图形特征，故设拟合曲线的公式为：

$$f(\delta)=\alpha e^{-\frac{(\delta-\mu)^2}{2\sigma^2}} \quad \cdots\cdots (3)$$

其中α、μ和σ均为待定的参数。公式（3）中的自变量δ位于指数中，因此需对该公式进行对数化处理：

$$h(\delta)=\ln\left(f(\delta)\right)=\ln(\alpha)-\frac{(\delta-\mu)^2}{2\sigma^2} \quad \cdots\cdots (4)$$

可以看到，公式（4）是关于δ的二次函数，故需要进行二阶多项式拟合。进行参数拟合的过程如下：首先，在区间[0.1,0.9]内，以0.01为步长等间隔地取值，得到相似性阈值δ的81个样本，即δ_1=0.10，δ_2=0.11，……，δ_{81}=0.90。随后计算各相似性阈值所对应的易混淆药品组数，并取自然对数，得到1组样本数据$\left(\delta_1, \ln\left(g\left(\delta_1\right)\right)\right)$，$\left(\delta_2, \ln\left(g\left(\delta_2\right)\right)\right)$，……，$\left(\delta_{81}, \ln\left(g\left(\delta_{81}\right)\right)\right)$。利用该组数据，在Matlab中使用polyfit函数进行二阶多项式拟合，即可确定参数α、μ和σ。上述相似性阈值δ的样本取值区间未设定为[0,1]的原因在于，靠近0和1的易混淆药品组数会对参数拟合的结果产生不利影响。

函数f的定义与正态分布的概率密度函数非常相似，仅存在常数项的差异。这点差异不会影响右侧拐点的横坐标位置。因此，根据后者的性质，可推断出函数f

的右侧拐点横坐标位于$\mu+\sigma$处。由于最佳相似性阈值就是右侧拐点的横坐标，因此最佳相似性阈值即是$\mu+\sigma$。然而，鉴于易混淆药品的判断往往受到个人主观因素的影响，在实际应用中，将$\mu+\sigma$的邻近值作为相似性阈值亦被视为一种合理的选择。

1.2.5　易混淆药品目录的制定

在明确设定最佳相似性阈值δ之后，可利用“1.2.3”项下所述方法，生成由易混淆药品组构成的易混淆药品目录。该目录为初步目录，包含的易混淆药品组是备选组，随后还需通过人工辨识和优化调整，形成最终的易混淆药品目录。值得注意的是，基于药品通用名和品种名所生成的易混淆药品目录存在一定的差异。因此，在制定易混淆药品目录时，建议同时参考基于这2种名称生成的目录。

1.3　软件实现

系统架构基于C/S（Client/Server）模式，利用Visual Basic 2010 Express开发客户端用户界面，而后台数据库则构建在SQL Server 2008 R2 Express平台上。通过使用Visual Basic.NET编程语言，将易混淆药品目录自动生成算法集成到客户端。此外，通过对高拍仪进行二次开发，将其拍照功能整合进系统，可以实现药师快速上传药品图像信息。

2　结果

2.1　通用名目录和品种名目录

从HIS导出我院门诊药房和住院药房的在库药品目录，分别包含737、889种药品。经过药品名称的规范化处理，形成门诊药房和住院药房各自的通用名目录和品种名目录。

2.2　相似性阈值

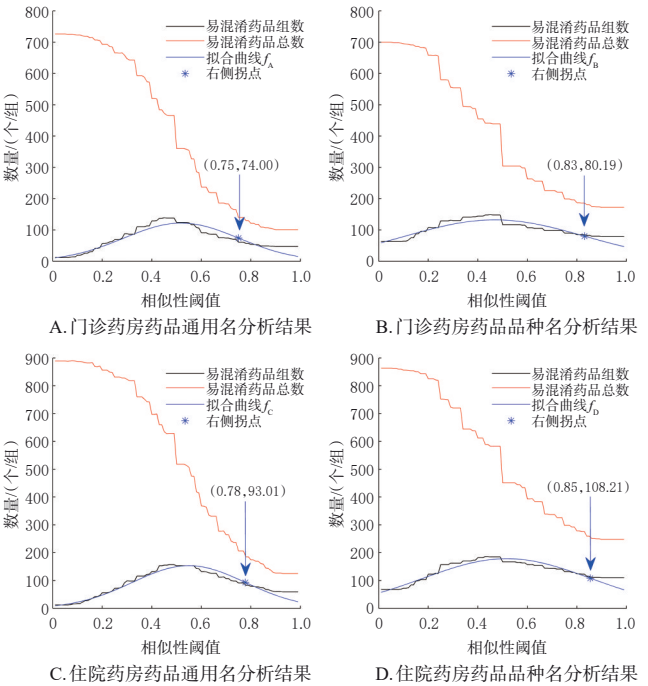
图2展示了相似性阈值δ对易混淆药品组数以及易混淆药品总数的影响，使用的药品名称数据来自门诊药房和住院药房的通用名目录和品种名目录。由图2可知，易混淆药品总数随相似性阈值δ的增大而逐渐减少。此结果与预期一致，即相似性阈值δ与易混淆药品总数呈负相关关系。此外，值得注意的是，随着相似性阈值δ的增大，易混淆药品组数呈现出先增后减的趋势，且曲线形态类似于钟形。

按“1.2.4”项下方法，对易混淆药品组数进行参数拟合，得到的各项参数值见表1。利用得到的参数值和公式（3），可进一步构建出各个拟合曲线的公式：

$$f_A(\delta)=122.01e^{-9.33(\delta-0.52)^2} \quad (5)$$

$$f_B(\delta)=132.21e^{-3.81(\delta-0.47)^2} \quad (6)$$

$$f_C(\delta)=153.35e^{-9.59(\delta-0.55)^2} \quad (7)$$



A.门诊药房药品通用名分析结果　B.门诊药房药品品种名分析结果

C.住院药房药品通用名分析结果　D.住院药房药品品种名分析结果

图2　相似性阈值δ与易混淆药品总数及易混淆药品组数的关系

$$f_D(\delta)=178.41e^{-4.40(\delta-0.52)^2} \quad (8)$$

图2展示了公式（5）至公式（8）的曲线，可以看到各拟合曲线与其各自对应的易混淆药品组数曲线均紧密贴合，表明拟合效果较好，因而可使用拟合曲线取代易混淆药品组数曲线来计算最佳相似性阈值。由“1.2.4”项下可知，函数f的右侧拐点横坐标位于$\mu+\sigma$处。利用该结论和表1的各项参数值，可以快速计算得到曲线f_A、f_B、f_C和f_D的右侧拐点横坐标，分别为0.75、0.83、0.78和0.85。换言之，当使用门诊药房和住院药房药品的通用名来建立易混淆药品目录时，最佳相似性阈值分别为0.75和0.78；当使用门诊药房和住院药房药品的品种名时，最佳相似性阈值分别为0.83和0.85。可以看到，使用药品通用名筛选得到的最佳相似性阈值比使用药品品种名得到的小，这是因为药品通用名中包含的剂型或给药途径等信息在特定情形下会降低药品名称之间的相似性。另外还可以发现，在使用药品通用名或药品品种名时，不同药房之间的最佳相似性阈值变化不大。因此，在实际应用时，可不考虑药房间的差异。在本研究中，使用药品通用名或药品品种名时，设置的相似性阈值分别为0.75和0.83，即$\delta_{通用名}=0.75$、$\delta_{品种名}=0.83$（选择较小的相似性阈值是为了纳入更多的易混淆药品）。

2.3　易混淆药品目录

运用易混淆药品目录自动生成算法处理门诊药房的药品通用名和品种名，得到2份初步的易混淆药品目录。随后经过人工辨识，排除一些不易引起混淆的备选

表1　对易混淆药品组数曲线进行参数拟合的结果

数据出处	数据集	参数		
		α	μ	σ
门诊药房通用名/药品品种目录	药品通用名	122.01	0.52	0.23
	药品品种名	132.21	0.47	0.36
住院药房通用名/药品品种目录	药品通用名	153.35	0.55	0.23
	药品品种名	178.41	0.52	0.33

药品组，例如不同规格的氯化钠注射液和葡萄糖注射液等。最后将上述2份优化调整后的易混淆药品目录合并，形成门诊药房的易混淆药品目录。采用同样的流程，也成功构建了住院药房的易混淆药品目录。整个过程中的关键统计数据详见表2。

表2　门诊药房和住院药房易混淆药品目录制定过程中的关键统计数据（个/种/组）

项目	门诊药房		住院药房	
	通用名	品种名	通用名	品种名
备选易混淆药品组数量	70	85	99	124
备选易混淆药品的总数	164	185	238	278
采纳的易混淆药品组数	64	81	88	122
采纳的易混淆药品总数	136	174	194	271
易混淆药品组数	95		149	
易混淆药品总数	209		330	

从表2可知，住院药房的易混淆药品组数和易混淆药品总数均多于门诊药房。这主要归因于住院药房具有更丰富的药品品种和剂型。从表2中还可观察到，无论是对于门诊药房还是住院药房，使用药品品种名筛选出的易混淆药品组数均大于使用药品通用名筛选出的结果。然而，这并不意味着使用2种药品名称生成的易混淆目录存在包含关系。为了明确这一点，本课题组精心挑选了7组典型实例，详见表3。在表3中，前4组易混淆药品是通过对比药品通用名之间的相似性得到；而后3组易混淆药品刚好相反，均通过对比药品品种名之间的相似性得到。可见使用药品品种名和通用名生成的易混淆药品目录不存在包含关系。因此，在制定易混淆药品目录时，为确保准确性和全面性，需要同时参考基于2种药品名称生成的目录。

2.4　易混淆药品目录管理软件

在Visual Basic 2010 Express平台上开发的客户端主界面见图3。图3左侧为药品的图片采集上传及下载显示区域，右侧为易混淆药品目录呈现区域。

图4展示的是易混淆药品目录的创建及维护界面。图4左侧为药品候选区，可按药品拼音码或关键字模糊查询药品，也可基于LD算法自动生成备选的易混淆药品组。获得备选药品组后，可根据实际需求建立易混淆药品组或将药品添加到已有的易混淆药品组中。图4右侧为易混淆药品组的呈现区域，可按药品拼音码或关键字模糊查询易混淆药品组。

表3　使用通用名和品种名筛选易混淆药品的7组典型实例比较

组号	通用名	品种名	$R_{通用名}$	$R_{品种名}$
1	盐酸多巴酚丁胺注射液	盐酸多巴酚丁胺	0.80	0.71
	盐酸多巴胺注射液	盐酸多巴胺		
2	注射用阿糖胞苷	阿糖胞苷	0.86	0.75
	注射用阿扎胞苷	阿扎胞苷		
3	盐酸氯丙嗪注射液	盐酸氯丙嗪	0.88	0.80
	盐酸异丙嗪注射液	盐酸异丙嗪		
4	注射用硫酸长春新碱	硫酸长春新碱	0.78	0.67
	注射用硫酸长春地辛	硫酸长春地辛		
5	聚维酮碘溶液	聚维酮碘	0.71	1.00
	聚维酮碘含漱液	聚维酮碘		
6	丙戊酸钠注射用浓溶液	丙戊酸钠	0.60	1.00
	丙戊酸钠口服溶液	丙戊酸钠		
7	人纤维蛋白原	人纤维蛋白原	0.63	0.83
	人纤维蛋白粘合剂	人纤维蛋白		

注：设定的相似性阈值为$\delta_{通用名}=0.75$、$\delta_{品种名}=0.83$。



图3　易混淆药品目录管理系统的主界面



图4　易混淆药品目录管理系统的目录创建及维护界面

3　讨论

易混淆药品管理是医院的一项重要工作。已有学者提出多种管理建议，例如制定易混淆药品目录、加强培训、优化药品摆放和设置警示牌等[7-8]。制定易混淆药品目录是进行易混淆药品管理的逻辑起点。然而，使用传统方式建立易混淆药品目录不仅效率低下、容易出现疏漏，而且后期维护工作繁重。由于手动维护目录困难，部分医院每年仅维护1次[2,11]。自药品集采以来，医院药品更替较频繁，迫切需要药师不定期维护易混淆药品目录。本课题组基于LD建立了易混淆药品目录自动生成算法，并据此开发了管理系统，以辅助药师创建及维护易混淆药品目录。

相较于传统的手工建立方法，本课题组所研发的易混淆药品目录管理系统在建立和维护易混淆药品目录的效率上得到了显著提升。传统方法构建此类目录往往需要耗时约1周，而采用本系统，仅需不到1 h便能高效完成目录的构建。即便考虑到药品图片的采集与录入工作，整个流程耗时亦能控制在1 d之内。值得强调的是，本系统成功实现了与HIS药品库存的实时同步功能。基于此，系统可在主界面仅展示在库的易混淆药品。这一设计极大地减轻了药师在更新和调整易混淆药品目录时的工作负担。使用该系统后，药师仅需专注于往目录内添加药品，不需要再额外关注药品的移除操作。

有学者使用LD算法识别易混淆药品时，仅考虑了药品两两之间的相似性[10]。这种做法可能导致一些相似药品被分配到多个组中。本课题组建立了易混淆药品组自动识别方法，将检测到的相似药品都归为一组，从而增强了算法的实用性。

相似性阈值δ是易混淆药品目录自动生成算法的重要参数。有文献报道，当基于药品通用名来建立易混淆药品目录时，可使用的相似性阈值δ为0.70[10]。本课题组发现使用通用名筛选易混淆药品时，δ应在0.75上下。此结论与报道的经验值相吻合，间接验证了本课题组建立的相似性阈值筛选方法有效、可行。

综上，本课题组以LD算法为研究的切入点，深入探讨了易混淆药品组的自动识别机制以及相似性阈值的筛选策略；在此基础上，本课题组提出了一种高效算法，旨在自动生成易混淆药品目录；此外，本课题组还简略介绍了所开发的易混淆药品目录管理系统，该系统显著提高了对易混淆药品目录的管理效率。本研究为药房的易混淆药品管理提供了新的思路和实践方法。

参考文献

[1] 刘翠文，宋再伟，杨毅恒，等. 基于人机料法环方法的住院药房易混淆药品管理与差错分析[J]. 中国药业，2022，31(1)：4-7.
LIU C W，SONG Z W，YANG Y H，et al. Management and error analysis of easily-confused drugs in inpatient pharmacy based on the 4M1E method[J]. China Pharm，2022，31(1)：4-7.

[2] 林小华，贺筱彬，欧阳浩仪. 医院易混淆药品管理模式探讨及应用[J]. 中国现代医生，2019，57(20)：1-3，7.
LIN X H，HE X B，OUYANG H Y. Discussion and application of hospital confusing drug management mode[J]. China Mod Dr，2019，57(20)：1-3，7.

[3] HOFFMAN J M，PROULX S M. Medication errors caused by confusion of drug names[J]. Drug Saf，2003，26(7)：445-452.

[4] CIOCIANO N，BAGNASCO L. Look alike/sound alike drugs：a literature review on causes and solutions[J]. Int J Clin Pharm，2014，36(2)：233-242.

[5] BERMAN A. Reducing medication errors through naming，labeling，and packaging[J]. J Med Syst，2004，28(1)：9-29.

[6] 杨满玲. 门诊西药房易混淆药品的分析[J]. 中国处方药，2014，12(3)：59-60.
YANG M L. Analysis of confusing drugs in outpatient western pharmacy[J]. J China Prescr Drug，2014，12(3)：59-60.

[7] 王慧. 门诊药房易混淆药品调查与对策分析[J]. 中国药事，2016，30(7)：719-721.
WANG H. Investigation and countermeasures on drug confusion in outpatient pharmacy[J]. Chin Pharm Aff，2016，30(7)：719-721.

[8] 李茜. 易混淆药品的调配差错分析及应对策略[J]. 基层医学论坛，2023，27(16)：120-125.
LI Q. Analysis of misallocation errors in easily confused drugs and corresponding strategies[J]. Med Forum，2023，27(16)：120-125.

[9] RISTAD E S，YIANILOS P N. Learning string-edit distance[J]. IEEE Trans Pattern Anal Mach Intell，1998，20(5)：522-532.

[10] 胡小刚，向露，李拓颖，等. 基于编辑距离算法的药房听似药品管理量化指标体系的建立[J]. 中国药业，2020，29(20)：19-22.
HU X G，XIANG L，LI T Y，et al. Establishment of quantitative index system for the sound-alike medications management in pharmacy based on edit distance algorithm [J]. China Pharm，2020，29(20)：19-22.

[11] 肖瑛，胡婷婷. 易混淆药品管理存在的问题与对策[J]. 抗感染药学，2016，13(1)：80-81.
XIAO Y，HU T T. Problems and countermeasures in managing easily confused drugs [J]. Anti Infect Pharm，2016，13(1)：80-81.

（收稿日期：2024-01-20　修回日期：2024-07-02）
（编辑：刘明伟）