1 对于全拼打字编码方式的改进分析

1.1 从汉语拼音的角度进行改进分析

在全拼输入法中，一共包含了23个声母：b、p、m、f、d、t、n、l、ɡ、k、h、j、q、x、zh、ch、sh、r、z、c、s、y、w；6个单韵母：a、o、e、i、u、v；18个复韵母：ai、ei、ui、ao、ou、iu、ie、ve、er、an、en、in、un、vn、ang、eng、ing、ong。

对给定文章的全拼拼音进行拆分统计，统计出该文章中声母、韵母的分布情况（其中将声母分为两类：单声母和双声母），如图-1所示：

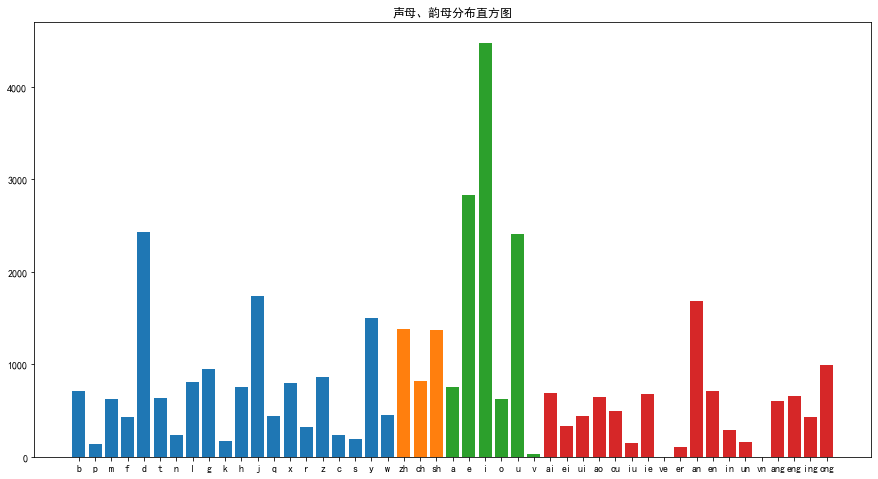


图- 1 声母、韵母分布直方图

通过对统计数据以及分布图的观察，不论是在声母还是韵母中，使用次数的分布都是不均匀的。同时对于双声母和复韵母来说，在输入时，它们的按键次数要比单声母和单韵母要多。

因此从输入效率的角度考虑，可以用单声母来代替复韵母，用单韵母来代替双声母，从而可以起到减少按键次数的效果，进而提示输入法的输入效率。

另外从按键使用均衡性的角度考虑，可以让使用次数较低的单声母来代替复韵母（单韵母替代同理），从而使不同按键的使用更加均匀，进而可以确保按键使用的均衡性。

1.2 从键盘打字方式的角度进行改进分析

考虑到目前主流的键盘打字方式，其指法大致如图-2所示：



图- 2 打字指法示意图

初始状态下，除去两个大拇指之外的其它手指是放在‘A’、‘S’、‘D’、‘F’、 ‘J’、‘K’、‘L’按键上，如果遇到这7个字母则直接敲击即可，而其他的字母则需要移动手指去进行敲击。

而针对给定文章的字母键使用热力图如图-3所示：

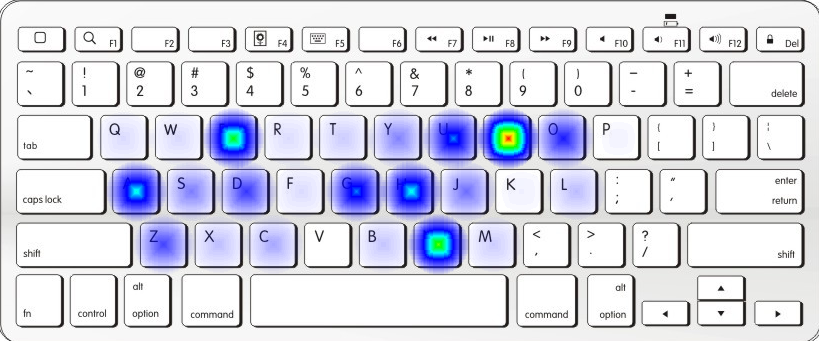


图- 3 26字母键使用热力图

从图-3中可以观察到部分字母键的使用频率较高，结合打字指法，可以考虑更改字母键在键盘上的位置分布，将使用频率较高的字母键移动到手指初始位置（原‘A’、‘S’、‘D’、‘F’、 ‘J’、‘K’、‘L’位置上），从而可以减少打字时的手指移动次数，进而可以提高打字的输入效率。

2 对于全拼打字编码方式的改进

针对1中的分析，这里对全拼打字编码方式进行改进，主要分为两步：首先是对双声母和复韵母的合理替换；然后便是对字母键在键盘上位置分布的合理更改。

2.1 声母、韵母的合理替换

假设表示单声母的使用统计数据，表示双声母的使用统计数据，表示单韵母的使用统计数据，表示复韵母的使用统计数据。

这里以单声母替换复韵母为例，进行说明计算，单韵母替换双声母同理。替换步骤如下：

1. 首先计算替换之后集合的平均值，计算公式如下：

(1)

1. 对集合进行升序排序，对进行降序排序，从集合中最小的元素开始，从中找出一个合适的子集，让与的和最接近，进而将中复韵母替换成单声母，同时将从中剔除；
2. 对集合中后面的元素进行同(2)的计算，直至将中的所有复韵母分配完毕，即完成替换。

计算过程使用python语言进行编程实现，运行结果如下：

1. 复韵母替换为单声母结果：

* an替换为p
* ong替换为k
* en替换为s
* ai替换为c
* ie替换为n
* eng替换为r
* ao替换为f
* ang替换为q
* ou替换为w
* ui替换为m
* ing替换为t
* ei替换为b
* in、un替换为h
* iu、er、ve、vn替换为x

1. 双声母替换为单韵母结果：

* zh替换为v
* sh替换为o
* ch替换为a

2.2 键盘位置更改

综合以上分析，这里主要考虑手指初始位置字母键的更换，更换步骤如下：

1. 使用2.1改进后的输入法对给定文章进行重新统计，统计出26个字母按键的使用数据：，并进行降序排序；
2. 从集合中最大的元素开始，将对应的字母移动到原‘J’的位置，如果对应的字母是‘J’、‘F’、‘K’、‘D’、‘L’、‘S’、‘A’其中的一个，则不进行移动（这里按照‘J’、‘F’、‘K’、‘D’、‘L’、‘S’、‘A’的顺序进行更换）；
3. 对后续‘F’、‘K’、‘D’、‘L’、‘S’、‘A’的位置进行同(2)的更换，直至结束，即完成键盘位置的更改。

使用python语言进行编程计算，得到的新的26个字母按键的使用次数数据如图-2所示：

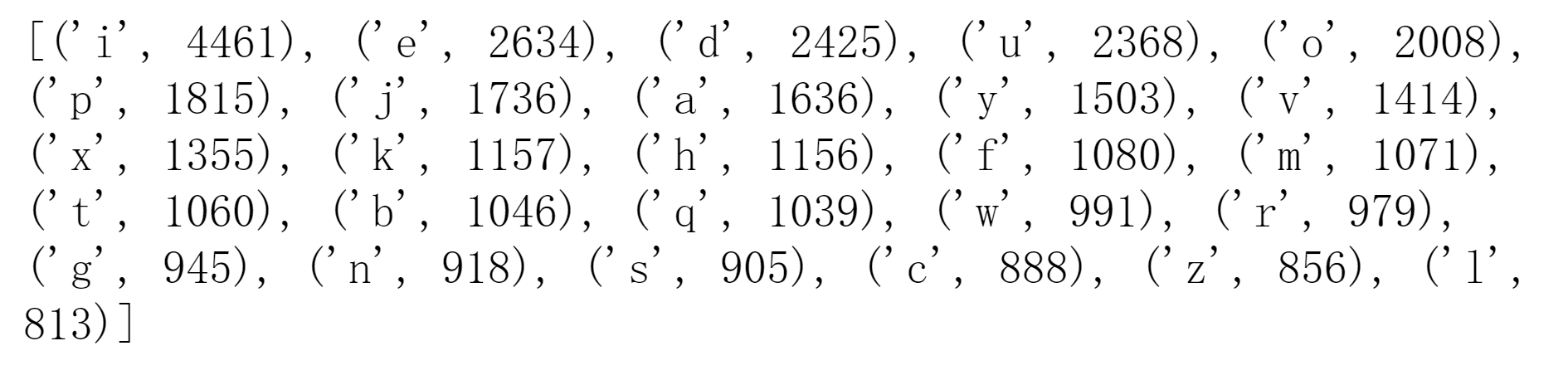


图- 4 更新后的26个字母按键的使用次数数据

进一步计算可得键盘更换结果如下：

1. 原始键盘字母键位置分布：

* 第一行：Q、W、E、R、T、Y、U、I、O、P
* 第二行：A、S、D、F、G、H、J、K、L
* 第三行：Z、X、C、V、B、N、M

1. 更改后键盘字母键位置分布：

* 第一行：Q、W、F、R、T、Y、K、A、L、S
* 第二行：J、P、D、E、G、H、I、U、O
* 第三行：Z、X、C、V、B、N、M

3 新旧方案在均衡性和输入效率方面的对比

使用问题2中制定的评价指标对新的打字编码方案进行计算，运算结果如图-5所示：

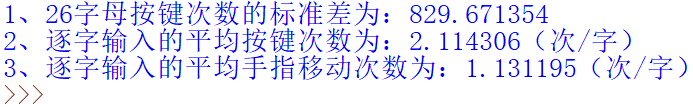


图- 5 新方案的运行结果

与就方案（图-6）相比，新方案的3个评价指标均有所下降，因此可以认为新的打字编码方案在均衡性和输入效率方面均有所提高。

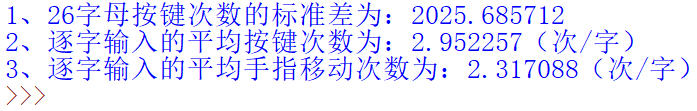


图- 6 就方案的运行结果