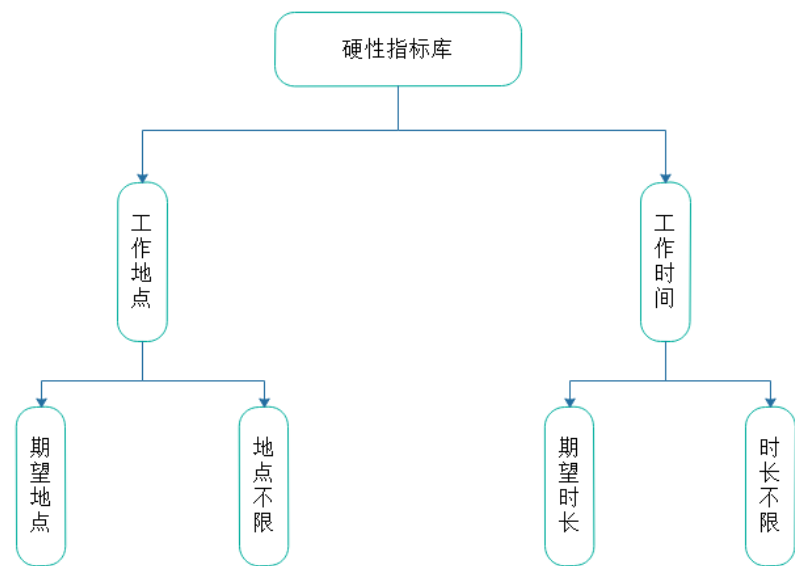


基于内容的推荐技术
编写者：李彦文

1. 设置硬性指标对招聘信息进行过滤。得到初步筛选结果。



2. 特征值的选取和量化

选取特征值：

本平台针对求职者选取以下五个特征字段：
工作经验，薪资要求，专业，岗位，学历。

量化：

工作经验量化表

工作经验	量化值
无	1
3 月以下实习经验	0.5
3~6 月实习经验	0.3
6 月以上实习经验	0.1

薪资量化表

薪资要求	量化值
10000 以上	1
5000~10000	0.5
3000~5000	0.3
3000 以下	0.1

专业量化表

专业	量化值
同专业	1
同大类不同专业	0.5
不同大类不同专业	0.1

岗位量化表

岗位	量化值
同岗位	1
相似岗位	0.5
不同岗位	0.1

学历量化表

学历	量化值
本科	1
研究生	0.5

3. 建立模型计算相似度

特征值权重表

特征值	权重
工作经验	0.3
薪资	0.2
专业	0.1
岗位	0.2
学历	0.2

假设每一份招聘信息的原始向量为 $D(k_1,k_2,k_3,...,k_n)$ ，如果都选为关键字，以关键字与信息结构相结合的方式，那么简历维度太高，太过复杂，不利于计算，影响效

果。所以将 2 中的五个特征值作为相似度计算的关键字。降维后的向量 $D(k_1, k_2, k_3, k_4, k_5)$ 。对一次筛选出来的招聘信息进行加权处理。

举例：

求职者

工作时间	工作地点	工作经验	薪资要求	学历	专业	岗位
956	福建省	无	无	本科	计算机	嵌入式开发

5 份招聘信息

招聘信息	工作地点	工作时间
D1	福建省	966
D2	北京市	956
D3	广东省	996
D4	福建省	956
D5	福建省	956

经过系统硬性指标筛选后，只剩下了 D4 和 D5 两份简历，再根据关键指标

$Q(\text{工作经验, 薪资, 专业, 岗位, 学历})$ ，进行二次筛选。

招聘信息	工作经验	薪资	专业	岗位	学历
Q4	6 月以上实习经验	11000	土木工程	建筑工程师	本科
Q5	3-6 月实习经验	8000	计算机	嵌入式开发	研究生

将以上两份简历的特征值根据其量化指标进行量化加权。

招聘信息特征值量化加权计算结果表

招聘信息	工作经验	薪资	专业	岗位	学历
Q0	0.3	0.2	0.1	0.2	0.2
Q4	0.03	0.2	0.01	0.02	0.2
Q5	0.09	0.1	0.1	0.2	0.1

相似度计算公式：

余弦方法

$$\cos(Q_i, Q_0) = \frac{\sum_{j=1}^n q_{ij} q_{0j}}{\sqrt{\sum_{j=1}^n q_{ij}^2 * \sum_{j=1}^n q_{0j}^2}}$$

根据余弦方法计算得到 Q4、Q5 的相似度，这个相似度是和 Q0 进行比较。结果越接近 1，相似度越高。

4. TOP-N

如果经过相似度计算之后得到的结果非常多，对结果按照相似度由大到小排序取前 N 个结果推荐给用户。