

团队成员：

鲍锋雄 程凯 陈振乾

陆纪慧 阳璐

基于加权多层K-Means 的企业分类系统e企查

企业推荐算法

e企查信息技术有限公司

大熊维尼队

指导教师：郑建炜

目录

一、概述 1

二、基本原理 1

三、实例示范 2

一、概述

当聚类算法实现时，其最具体的落地表现，就是用于向平台使用者推荐潜在的，与目标企业相近的企业，拓展金融信贷的业务可能，使得该无监督算法得到真正的体现。

具体的，为了达到这个目标，在推荐企业的过程中，会从多个依据，结合聚类得到的信息，得到与目标用户企业最相近的企业。当然这个算法可以推广到根据用户平时的浏览习惯来对用户进行目标企业的推荐。

二、基本原理

在《企业画像算法》中，我们通过等级化的方式，为企业计算出了它的综合等级，记作 `ent_module_type`；通过 K-Means 聚类算法，对企业的 7 个维度的属性进行聚类，对企业最终归入的簇，给其聚类标签，记作 `inner_type`；将目标推荐的企业个数设定为 `n`。

按照推荐优先级，根据行业门类(`industryphy`)>聚类标签(`inner_type`)>企业等级(`ent_module_type`)，得到如图 2-1 所示的推荐流程：

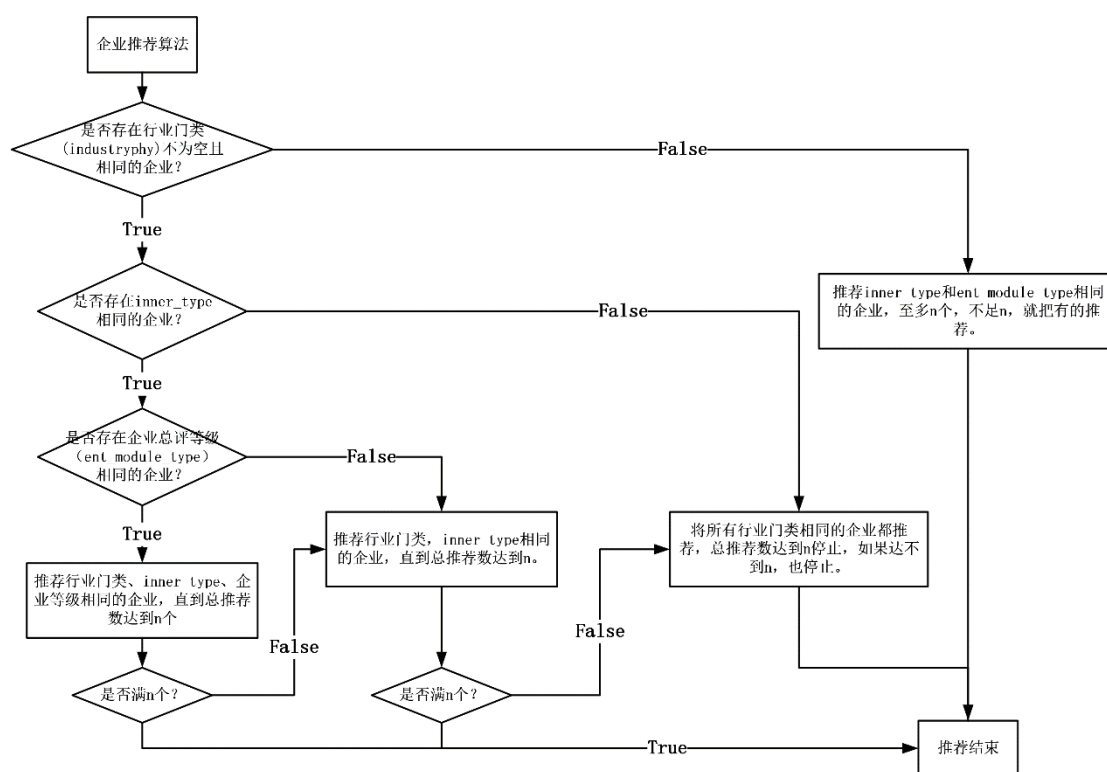


图 2-1：企业推荐流程

图 2-1 的流程算法的伪代码可以根据如下步骤解释：

步骤 1：判断数据库中是否存在与当前企业行业门类(`industryphy`)相同的企业。如果存在，则跳转到步骤 2；否则跳转到步骤 5。

步骤 2：推荐行业门类(`industryphy`)相同，聚类标签(`inner_type`)相同，且企业等级(`ent_module`

_type)相同的企业，查询结果加入到查询集合中。如果查询集合中满 n 个企业，跳转到步骤 6；否则，跳转到步骤 3。

步骤 3：推荐行业门类相同，聚类标签相同，企业等级不同的企业，查询结果加入到查询集合中。如果查询集合中满 n 个企业，跳转到步骤 6；否则，跳转到步骤 4。

步骤 4：推荐行业门类不同，聚类标签不同的企业。查询结果加入到查询集合中。当查询集合满 n 个，或者行业门类相同企业推荐完毕，跳转到步骤 6。

步骤 5：推荐聚类标签和企业等级相同的企业。查询结果加入到查询集合中。当查询集合满 n 个，或者聚类标签和企业等级相同的企业推荐完毕，跳转至步骤 6。

步骤 6：推荐结束。

三、实例示范

提供样例数据表（中间省略了较多属性），如表 3-1 所示：

表 3-1：推荐算法样例数据表

entname	industryphy	inner_type	ent_module_type
ent1	畜牧业	7	2
ent2	畜牧业	1	5
ent3	畜牧业	7	3
ent4	互联网	8	1
ent5	互联网	8	1
ent6	互联网	8	3
ent7	互联网	8	1
ent8	互联网	1	3
ent9	互联网	8	4
ent10	金融业	8	1
ent11	互联网	1	2

提供如下测试样例：

- 搜索 ent1，规定 n 的个数为 5，推荐相似企业。
 - ①由于存在与 ent1 行业门类相同的企业，进入步骤 2。
 - ②首先找 inner_type 和 ent_module_type 相同的企业，此时查询集合中企业个数为 0，进入步骤 3。
 - ③找到 inner_type 相同但是 ent_module_type 不同的企业，ent3 符合要求，加入到查询集合中，此时查询集合中企业个数为 1，进入步骤 4。
 - ④找到畜牧业中，inner_type 不同的企业，此时 ent2 加入到查询集合中，跳转至步骤 6。
 - ⑤推荐结束。所以推荐企业的集合是 {ent3, ent2}。
- 搜索 ent4，规定 n 的个数为 5，推荐相似企业。

- ①由于存在与 ent4 行业门类相同的企业，进入步骤 2。
 - ②找到 inner_type 和 ent_module_type 相同的企业。ent5 和 ent7 加入集合。此时企业个数为 2，查询集合为{ent5,ent7}，进入步骤 3。
 - ③找到 inner_type 相同，但是 ent_module_type 不同的企业。ent6 和 ent9 进入集合，此时查询集合为{ent5,ent7,ent6,ent9}，进入步骤 4。
 - ④找到 inner_type 不同的企业，此时有 ent8 和 ent11 两个企业，由于 n 的个数为 5，此时只能将 ent8 加入，最终的查询结合为{ ent5,ent7,ent6,ent9,ent8}，跳转至步骤 6。
 - ⑤推荐结束，最终推荐集合为{ ent5,ent7,ent6,ent9,ent8}。
- 搜索 ent10，规定 n 的个数为 5，推荐相似企业。
 - ①由于没有与它行业门类相同的企业，进入步骤 5。
 - ②找到 inner_type 与 ent_module_type 相同的企业，将 ent4，ent5，ent7 加入到查询集合中，跳转至步骤 6。
 - ③推荐结束，最终推荐集合为{ ent4，ent5，ent7}。