

购物篮子分析

2022 年 4 月 2 日

1 使用 Apriori 算法进行购物篮子分析

购物篮子分析是大型零售商用来发现商品之间联系的关键技术之一。它通过寻找交易中经常出现的商品组合来发挥作用。换句话说，它使零售商能够识别人们购买的商品之间的关系。

在本次作业中，使用 Groceries¹ 数据进行数据挖掘。同一消费者在同一天购买的商品集合视为一个事务 (transaction)，请实现 Apriori 算法并在此数据上进行数据挖掘，并根据挖掘的结果回答以下问题。

- 请列举 $k=1$ 的前 10 个频繁项集。(2 分)
- 请列举 $k=2$ 的前 5 个频繁项集，以及对应的关联规则，support 和 confidence。(4 分)
- 请列举 $k=3$ 的前 5 个频繁项集，以及对应的 support。(2 分)
- 请根据数据挖掘结果得到有价值的结论。(2 分)

注意事项：Apriori 需要自己手动实现，关联规则和结论需要根据自己实现的 Apriori 算法给出。

¹<https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset>