数据挖掘大作业二：关联规则挖掘

滕博文 5720182060

1. 数据源

Youtube视频集

2. 实验内容

1. 数据集处理，关联规则挖掘数据形式；

选择“category\_id，comments\_disabled，ratings\_disabled，video\_error\_or\_removed四个属性，每个属性对应一到多个值，每个值为一个项，每个事务以list存储进行分析。

1. 频繁项集；

#频繁1项集

[frozenset({'video\_error\_or\_removed=False'}),frozenset({'ratings\_disabled=False'}), frozenset({'comments\_disabled=False'})]

#频繁2项集

[frozenset({'comments\_disabled=False', 'ratings\_disabled=False'}), frozenset({'video\_error\_or\_removed=False','comments\_disabled=False'}), frozenset({'video\_error\_or\_removed=False','ratings\_disabled=False'})]

#频繁3项集

[frozenset({'video\_error\_or\_removed=False', 'comments\_disabled=False', 'ratings\_disabled=False'})], []

1. 关联规则，支持度(support)和置信度(conf)

[(frozenset({'ratings\_disabled=False'}), frozenset({'comments\_disabled=False'}), 0.5782472149798531, 0.9838267638405069), (frozenset({'comments\_disabled=False'}), frozenset({'ratings\_disabled=False'}), 0.5782472149798531, 0.9868507893252149), (frozenset({'comments\_disabled=False'}), frozenset({'video\_error\_or\_removed=False'}), 0.5855575030251619, 0.9993267050595587), (frozenset({'video\_error\_or\_removed=False'}), frozenset({'comments\_disabled=False'}), 0.5855575030251619, 0.9779594451707667), (frozenset({'ratings\_disabled=False'}), frozenset({'video\_error\_or\_removed=False'}), 0.5873585658861541, 0.9993287682499423), (frozenset({'video\_error\_or\_removed=False'}), frozenset({'ratings\_disabled=False'}), 0.5873585658861541, 0.9809674613385281), (frozenset({'ratings\_disabled=False'}), frozenset({'video\_error\_or\_removed=False', 'comments\_disabled=False'}), 0.5778526964483977, 0.9831555320904493), (frozenset({'comments\_disabled=False'}), frozenset({'video\_error\_or\_removed=False', 'ratings\_disabled=False'}), 0.5778526964483977, 0.9861774943847735), (frozenset({'video\_error\_or\_removed=False'}), frozenset({'comments\_disabled=False', 'ratings\_disabled=False'}), 0.5778526964483977, 0.9650913863959539)]

1. 使用Lift对规则进行评价

[[frozenset({'ratings\_disabled=False'}), frozenset({'comments\_disabled=False'}), 1.6790227316340869], [frozenset({'comments\_disabled=False'}), frozenset({'ratings\_disabled=False'}), 1.6790227316340869], [frozenset({'comments\_disabled=False'}), frozenset({'video\_error\_or\_removed=False'}), 1.6690094226021404], [frozenset({'video\_error\_or\_removed=False'}), frozenset({'comments\_disabled=False'}), 1.6690094226021404], [frozenset({'ratings\_disabled=False'}), frozenset({'video\_error\_or\_removed=False'}), 1.6690128684063739], [frozenset({'video\_error\_or\_removed=False'}), frozenset({'ratings\_disabled=False'}), 1.6690128684063739], [frozenset({'ratings\_disabled=False'}), frozenset({'video\_error\_or\_removed=False', 'comments\_disabled=False'}), 1.6790076585325597], [frozenset({'comments\_disabled=False'}), frozenset({'video\_error\_or\_removed=False', 'ratings\_disabled=False'}), 1.6790041920933205], [frozenset({'video\_error\_or\_removed=False'}), frozenset({'comments\_disabled=False', 'ratings\_disabled=False'}), 1.6689944393931564]]