**Thinking1:如何使用用户标签来指导业务(如何提升业务)?**

答: 通过对用户选择的商品,给用户打上相应的标签,然后制定策略推送给用户一定比例的、相同种类的商品,标签相当于用户的特征,通过不同的特征来描绘这个用户,然后猜测这个用户可能喜欢的上品。

**Thinking2:如果给你一堆用户数据,没有打标签,你该如何处理(如何打标签)?**

答: 1.涉及到冷启动的问题,在没有打标签的情况下,一般采用推荐热门商品或者协同过滤

的方式推荐商品,因此也可以使用协同过滤的方式,找到相似的用户去打上相似用户

的标签。

2.或者通过现有的用户数据,使用无监督的分类方式,自动去给用户去分类,然后再去打标签.

**Thinking3:准确率和精确率有何不同(评估指标)?**

答: 在预测过程中,可能把对的预测成对的,也可能把对的预测成错误的,相反,也可能把错

的预测成对,也可能把错的预测成错的,一共有四种情况。

而准确率是指预测正确的(把对的预测成对的,把错的预测成错的)的概率;

精确率是指真实正确的的数量占预测正确的数量的百分比。

**Thinking4:如果你使用大众点评,想要给某个餐厅打标签.这时系统可以自动**

**提示一些标签,你会如何设计(标签推荐)?**

答: 如果是已经在大众点评上存在很久的餐厅,可以直接使用以前用户经常给这个餐厅打的 标签,如果是新的餐厅,可以查找与这个餐厅相似的餐厅的标签。

**Thinking5:我们今天使用了10种方式来解MNIST,这些方法有何不同?那你还有其他方法来解决MNIST识别问题么(分类方法)?**

答: TOPT是使用自动机器学习的方式,而其他的方式都是自己选择分类方法.可以使用手写

神经网络的方式,使用不同的激活函数去分类.但是只是知道,并不会写.

**Action1:针对Delicious数据集,对SimpleTagBased算法进行改进(使用NormTagBased、TagBased-TFIDF算法)**

答: 1.NormTagBased算法



2.TagBased-TFIDF算法



**Action2: 对Titanic数据进行清洗,使用之前介绍过的10种模型中的至少2中(包括TPOT)**

答: 1.TPOT



2. 贝叶斯分类器

