基于属性加权贝叶斯分类器的 O2O 网站优惠券使用预测

摘要：本文通过使用带权重的贝叶斯方法结合 L-BFGS 算法来预测 O2O 网站用户在领取优惠券后 15 天以内的核销情况。权重的学习采用极大似然估计的方法，结合 L-BFGS 算法来获取似然函数取极大值时对应的权重值，从而突出了各特征对预测结果的不同影响。为了更充分地挖掘具备预测价值的因素，本文从商户、用户、商户用户匹配关系、优惠券四个方面构建了 12 个额外地有用特征。经过实际预测后，发现传统分类器和本文提出分类模型都能有较好的预测效果，且本文提出模型更胜一筹。

关键词：属性加权的贝叶斯 L-BFGS 优惠券 预测

1. 绪论

1.1 业务背景

O2O，全称 Online To Offline，线上线下电子商务，是把线上的消费者带到现实的商店中去：在线支付线下商品、服务，再到线下去享受服务。通过打折（例如团购）、提供信息、服务（例如预定）等方式，把线下商店的消息推送给互联网用户，从而将他们转换为自己的线下客户。这样线下服务就可以用线上来揽客，消费者可以用线上来筛选服务，还有成交可以在线结算。

以优惠券盘活老用户或吸引新客户进店消费是 O2O 的一种重要营销方式。然而随机投放的优惠券对多数用户造成无意义的干扰。对商家而言，滥发的优惠券可能降低品牌声誉，同时难以估算营销成本。 个性化投放是提高优惠券核销率的重要技术，它可以让具有一定偏好的消费者得到真正的实惠，同时赋予商家更强的营销能力。因而，准确地预测优惠券能否被用户使用，是一个具备商业意义的问题。

本项目着力于解决预测优惠券是否会被使用的问题。本项目已知用户在 2016 年 1 月 1 日至 2016 年 6 月 30 日之间真实线上线下消费行为数据，基于此预测用户在 2016 年 7 月领取优惠券后 15 天以内是否会核销优惠券。

1.2 算法背景

朴素贝叶斯(Naive Bayes)是一种简单而且高效的分类模型，它在属性条件独

立性假设的前提下利用贝叶斯定理进行分类。然而，朴素贝叶斯有以下主要缺点：1）朴素贝叶斯中每一个特征对于预测结果的贡献都是相同的，无法区分度更高的因素和区分度更低的因素。2）理论上，朴素贝叶斯模型与其他分类方法相比具有最小的误差率。但是实际上并非总是如此，这是因为朴素贝叶斯模型给定输出类别的情况下,假设属性之间相互独立，这个假设在实际应用中往往是不成立的，在属性个数比较多或者属性之间相关性较大时，分类效果不好。而在属性相关性较小时，朴素贝叶斯性能最为良好。对于这一点，有半朴素贝叶斯之类的算法通过考虑部分关联性适度改进。3）需要知道先验概率，且先验概率很多时候取决于假设，假设的模型可以有很多种，因此在某些时候会由于假设的先验模型的原因导致预测效果不佳。

1.3 本文贡献

本文主要贡献包括：(1)基于业务特征和预测目标，从商户、用户、商户用户匹配关系、优惠券四个方面挖掘了 12 个额外地对预测有价值的特征。(2)提出了一种属性加权的贝叶斯分类器，结合 L-BFGS 算法优化参数。(3)基于所设计的分类器，对 O2O 网站所领取优惠券的核销情况进行了预测。

2. 方案设计

2.1 特征构建

针对预测目标，本文从多个方面进行了特征的分析与提取。影响预测结果的因素主要包括商户、用户、商户用户匹配关系、优惠券特征这四个方面，因此构建的特征也是从以上四个方面来进行。

对于用户、商户、用户商户关系的每一个方面，其主要构建的特征围绕用户商户之间的距离、用户在商户内的优惠券领取情况、核销情况、用户使用优惠券消费占比情况、用户消费频繁情况等等角度来构建的。

对于优惠券的特征构建，主要从优惠折扣率、优惠额度、优惠形式几个角度进行。

2.2 模型选择

本项目采用基于属性加权的贝叶斯模型来完成预测，其中权重参数的学习通过结合 L-BFGS 算法的极大似然估计方法来完成。

朴素贝叶斯(Naive Bayes)本身是一种简单而且高效的分类模型，但是它的前提假设是各属性相互度量且各属性对于结果预测的贡献相同。对于本项目 O2O

优惠券的使用预测而言，影响预测结果的因素包括商家、用户、优惠券以及三者之间相互关系等多种因素，这些因素对预测结果的影响不同。为了反映不同属性对实例类属性归属的影响差异，本项目采用了基于属性加权的贝叶斯模型来完成预测。

对于模型参数的学习过程，考虑到时间复杂度取决于优化算法求解过程的收敛速度，采用了 L-BFGS 算法来作为优化算法，这种算法比较适合在大规模数据集上进行计算，它具有牛顿法收敛速度快的特点。

3. 预测模型

3.1 属性加权的贝叶斯

在一个测试实例 x 上预测其所属类别 c 的后验概率公式为：

$$\mathrm{P(c|x};w_x)=\frac{P(c)\prod_i P(a_i|c)^{w_i}}{\sum_{c'}P(c')\prod_i P(a_i|c')^{w_i}} \tag{1}$$

且满足条件$\sum_{c'}P(c')=1$，$\sum_j P(a_{ij}|c)=1$

其中，$\mathrm{w_x}=(\mathrm{w_0},\mathrm{w_1},\dots\dots,\mathrm{w_{s-1}})$是对于记录 x 对应不同属性的权重向量，s 为属性的个数；而w_i是第 i 个属性的权重，a_i是指记录 x 的第 i 个属性所对应的取值，a_{ij}表示第 i 个属性A_i可能取到的第 j 个值，c'是类别变量 y 所可能取到的值。

为了防止后验概率因乘积中的某个因子为 0 而产生 0 的结果，先验概率 $P(c)$ 和条件概率$P(a_i|c)$的估计均采用 m 估计。

3.2 参数学习

需要确定的参数是权重向量 w，为了确定权重向量 w 的取值，采用了极大似然估计法。本例中，对数似然函数为：

$$\begin{aligned}\mathrm{L(w)}&=\sum_x P(\mathrm{c|x};w_x)\\&=\sum_x\left(\ln\left(P(c)\prod_i P(a_i|c)^{w_i}\right)\right.\\&\left.-\ln\left(\sum_{c'}P(c')\prod_i P(a_i|c')^{w_i}\right)\right)\end{aligned} \tag{2}$$

对数似然函数的梯度为：

$$\frac{\partial \mathrm{L}(\mathrm{w})}{\partial \mathrm{w}_i} = \sum_x \left(\frac{w_i}{P(a_i|c)} - \frac{w_i}{\sum_{c'} P(a_i|c')} \right) \quad (2)$$

结合计算出的对数似然函数和对应的梯度，使用 L-BFGS 算法来求得使目标函数-L(w)取得最小值时所对应的权重向量 w 值。之所以采用 L-BFGS 算法，是因为 L-BFGS 算法具有具有收敛速度快、内存开销少等优点，适合于本项目超过 1300 万条记录的大型数据集。

3.3 整体实现思路

该算法的整体实现思路如下表 1 所示：

表 1 基于属性加权的贝叶斯模型实现思路

算法：基于属性加权的贝叶斯模型
输入：训练集 D、测试实例 x
输出：测试实例 x 取值为 1 的概率p_X
1. 计算所有的先验概率$P(c)$和条件概率$P(a_i\|c)$
2. 初始化权重向量 w
3. 利用式(2)计算对数似然函数 L(w)
4. 利用式(3)计算对数似然函数的梯度
5. 用 L-BFGS-M 优化程序计算函数- L(w)取最小值的解向量 w
6. 利用式(1)计算对于每一个测试实例 x 的后验概率p_X
7. 返回测试实例 x 的后验概率

4. 应用分析与实证

4.1 数据说明

本赛题提供用户在 2016 年 1 月 1 日至 2016 年 6 月 30 日之间真实线上线下消费行为，预测用户在 2016 年 7 月领取优惠券后 15 天以内的使用情况。

详细数据含义：

Table 1: 用户线下消费和优惠券领取行为

Field	Description
User_id	用户ID
Merchant_id	商户ID
Coupon_id	优惠券ID：null表示无优惠券消费，此时Discount_rate和Date_received字段无意义
Discount_rate	优惠率：x \in [0,1]代表折扣率；x:y表示满x减y。单位是元
Distance	user经常活动的地点离该merchant的最近门店距离是x*500米（如果是连锁店，则取最近的一家门店），x\in[0,10]；null表示无此信息，0表示低于500米，10表示大于5公里；
Date_received	领取优惠券日期
Date	消费日期：如果Date=null & Coupon_id != null，该记录表示领取优惠券但没有使用，即负样本；如果Date!=null & Coupon_id = null，则表示普通消费日期；如果Date!=null & Coupon_id != null，则表示用优惠券消费日期，即正样本；

Table 2: 用户线上点击/消费和优惠券领取行为

Field	Description
User_id	用户ID
Merchant_id	商户ID
Action	0 点击， 1购买，2领取优惠券
Coupon_id	优惠券ID：null表示无优惠券消费，此时Discount_rate和Date_received字段无意义。“fixed”表示该交易是限时低价活动。
Discount_rate	优惠率：x \in [0,1]代表折扣率；x:y表示满x减y；“fixed”表示低价限时优惠；
Date_received	领取优惠券日期
Date	消费日期：如果Date=null & Coupon_id != null，该记录表示领取优惠券但没有使用；如果Date!=null & Coupon_id = null，则表示普通消费日期；如果Date!=null & Coupon_id != null，则表示用优惠券消费日期；

为了测试预测结果，将线下消费数据集中的 80%设为训练集，剩下 20%设为测试集。

4.2 预测因子选取

本文中，从商户、用户、商户用户匹配关系、优惠券四个方面分别提取了对结果预测性较强的特征。

(1) 用户相关特征

用户特征提取了所消费过商户距离的中位数与最大数，体现了用户的活动半径，同时提取了用户使用优惠券的比例，以及用户使用优惠券消费占总消费的比

例。这一些特征即反映了用户使用优惠券的偏好，也反映了用户是否“不差钱”而直接进行消费，且消费过了商家的某些产品。

(2) 商户相关特征

商户相关特征首先提取了卖出商品优惠券占比、优惠券使用率这两项，直接关系到优惠券是否会使用。同时，提取了优惠券消费用户与商户的距离中位数与最大值，以体现商户优惠券辐射半径与辐射密度。

(3)用户-商户相关特征

在用户-商户相关特征方面，共提取了以下 5 个特征：用户在商户消费次数、用户领取商户优惠券次数、用户在商户使用优惠券消费次数、用户对商户的优惠券转化率、用户对商户使用优惠券消费占总消费比例。以上的各特征都能体现用户对商户的偏好，即挖掘了用户与商户之间的匹配关系。

(4) 优惠券相关特征

在优惠券特征方面，共提取了优惠券折扣率、优惠券的消费阈值、折扣类型 3 个特征。这些特征与优惠券的使用预测直接相关联。

4.3 实验结果与分析

本项目将线上线下数据集合并，完成数据清洗和特征工程后，选取 80%作为训练集，剩余 20%作为测试集。本项目分别将数据集应用于本文所提出的模型、朴素贝叶斯模型、随机森林模型、逻辑回归模型，以 ROC 作为评价指标，得到的实验结果如图 1 所示。

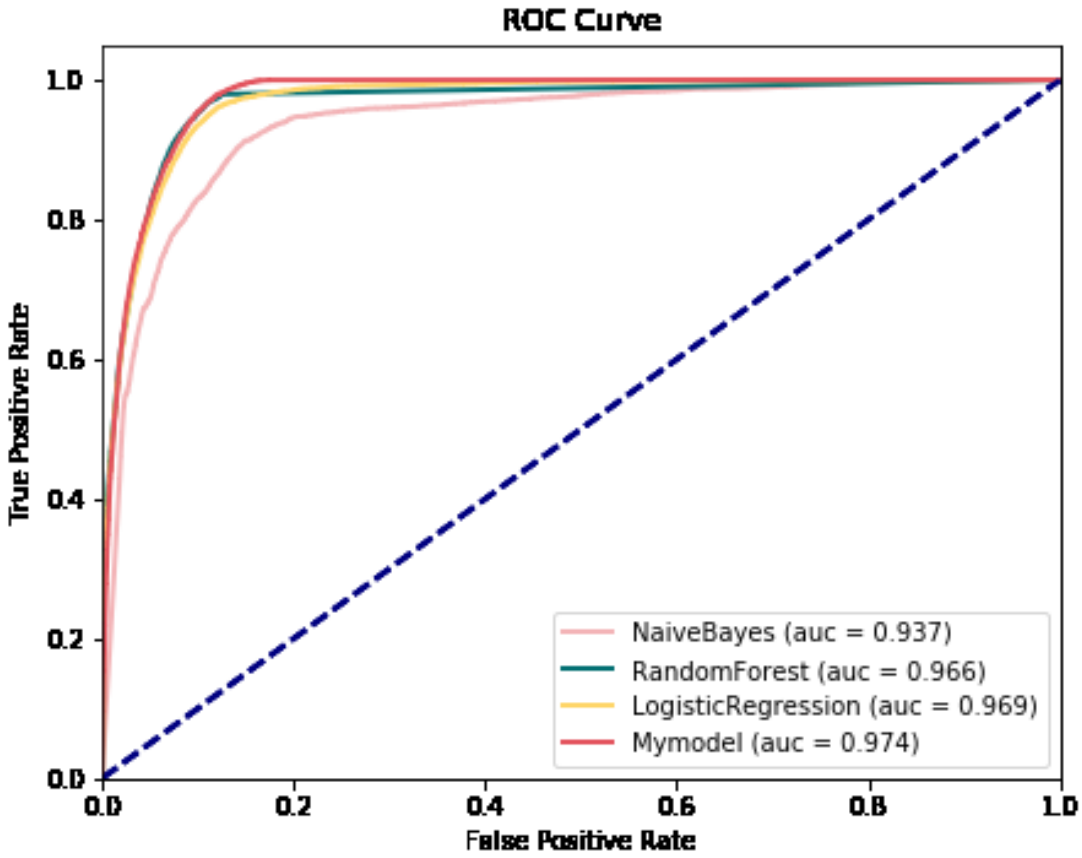

图 1 实验结果(注：本文所提出模型对应图例名为 Mymodel)

由此可见，无论是所设计的算法还是传统的方法，auc 的值都在 0.93 以上，这得益于恰当充分的特征构建。对比本文所提出模型和其它传统分类模型的表现，基于属性加权的贝叶斯模型将 auc 值提高至 0.974，超过了其它传统模型的表现。

对比其它传统算法，本文所提出的基于属性加权的贝叶斯模型在时间复杂度上表现较差，这主要是在求解似然函数极值点过程中需要反复迭代。在这一点上，需要后续的改进修改。

5. 结论

本文从商户、用户、商户用户匹配关系、优惠券四个角度挖掘了对预测优惠券核销概率有价值的共 12 个特征，并分别将完成特征工程后的数据集应用于的基于属性加权的贝叶斯模型、朴素贝叶斯模型、随机森林模型、逻辑回归模型。无论是传统分类器还是所设计的基于属性加权的贝叶斯分类器，预测结果的 auc 值均达到了 0.93 以上，这证明了特征工程的合理性。对比本文所提出模型和其它传统分类模型的表现，本文所提出模型将 auc 值还小幅提升至 0.974，这证明了本文提出模型在分类效果上优异的表现。

基于属性加权的贝叶斯分类器在时间的角度上表现较差。在这一点上，需要后续的改进。

参考文献

[1] Yukuo Chen,Jing Zhang,Gaofei Wang. Trust Relationship Prediction in Alibaba E-commerce Platform

[2] Jensen F. An Introduction to Bayesian Networks. New York: Springer, 1996.

[3] Broom BM, Do KA, Subramanian D. Model averaging strategies for structure learning in Bayesian networks with limited data. BMC Bioinformatics, 2012, 13(Suppl 13): S10.

[4] 乔秀全，李晓峰，廖建新. 基于贝叶斯网络的业务上下文认知模型构建方法. 电子与信息学报, 2008, 30(2): 464-467.

[5] Huang S, Li J, Ye JP, et al. A sparse structure learning algorithm for Gaussian Bayesian network identification from high-dimensional data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(6): 1328-1342. DOI:10.1109/TPAMI.2012.129

[6] De Campos LM, Castellano JG. Bayesian network learning algorithms using structural restrictions. International Journal of Approximate Reasoning, 2007, 45(2): 233-254. DOI:10.1016/j.ijar.2006.06.009

[7] 陈友玲，张岳园，凌磊，等. 基于贝叶斯网络的个体隐性知识测度方法研究. 计算机应用研究, 2019, 36(6). DOI:10.3969/j.issn.1001-3695.2017.12.0799