基于商品聚类的电商销量预测①

王建伟

(中国矿业大学 计算机科学与技术学院, 徐州 221116)

摘 要: 随着我国大力推进电商行业的发展, 越来越多的电商企业加入到线上的竞争之中. 随着销量的增大, 第三方电商企业所掌握的销售数据也越来越多, 这些分类上零散的销售数据给数据处理预测带来了一定的难度, 常常导致在预测过程中数据不完备或者预测结果存在非常大的偏差. 为了改善这一问题, 这里提出了一种基于销售数据的产品重分类预测模型, 利用产品销售共性提取产品聚类簇, 再使用时间序列模型得出预测结果并通过隐马尔科夫预测模型给出预测结果的概率分布. 通过实验分析, 利用以上模型的预测获得较好的预测结果, 对电商企业制定营销策略具有一定的参考价值.

关键词: 电商; 聚类; 时间序列; 隐马尔可夫; 预测

Online Sales Volume Prediction Based on Items Clustering

WANG Jian-Wei

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: By promoting of our government, more and more electronic business enterprise join the competition of online sales. With the sharp increase in sales, an ever increasing number of sales data is accumulated by the third party enterprise, and these sales data which is too scattered in original classification, it brings some difficulty in sales forecasting, in detail, it would lead to incomplete condition or severe deviation of predicted value. To improve this problem, a prediction model which is based on goods re-classification is constructed in this paper. This model used common sales features of products to extract the product cluster, then it used time series forecasting model to give the predicted value which is decorated by HMM in probability distribution aspect. Through experimental analysis, the final predicted values preferable fit the true values, and this achievement will provide the reference value to enterprise in establishing policies of distribution.

Key words: electric business; clustering; time series; HMM; sales forecast

近年来我国电子商务行业发展迅猛, 且一直保持着较快的增长势头, 2012 年, 交易额就达 81000 亿元, 2013 年, 仅天猫双十一购物狂欢节支付宝成交额变达到了 571.1 亿元, 电商行业全年销售额更是达到了惊人的 123000 亿元. 2012 年 3 月, 工信部出台了《电子商务“十二五”规划书》, 首次将电子商务战略升级至国家发展计划, 并指出到 2015 年, 电子交易额翻两番, 突破 18 万亿元 1. 由此可见, 电子商务的发展已经得到了国家战略层面上的关注.

在电商行业销售额不断增大的背后, 是各层电商间的相互角力竞争, 国内天猫, 淘宝, 京东, 亚马逊等在 C2C 平台领域各有优势, 而借助这些平台的大中小型电商的竞争也日趋白热化, 如何在平台中用更低的投入获取更高的营销回报是大中型第三方商家一直在追求的目标.

作为日益成长的非平台电商企业, 随着销量的增长, 品类的扩张, 积累了越来越多的销售数据, 一方面, 数据的增加给数据处理分析带来了更为可靠的保障, 另一方面, 数据的激增又给中小型电商处理数据带来了新的挑战. 作为销售多品类商品的买家, 随着

① 收稿时间:2016-02-16;收到修改稿时间:2016-03-31 [doi:10.15888/j.cnki.csa.005423]

商品品类的增多，数据会进一步分化，如何将这些零碎的数据重组在一起，再利用重组后的数据发现销量之间变化的规律，给出具有营销指导性意见的结果，对于中小型非平台电商利用自由数据提升销量有着重要的意义.

1 相关工作

对于电子商务营销策略的研究，文献[2]从消费者消费行为的角度进行了分析，通过浏览记录，搜索记录，评价记录等数据，利用统计的方式，对用户行为进行了系统的分析，并利用分析结果对电商营销给出对应的策略. 文献[3]指出了精准营销在电子商务中的重要作用. 文献[4]通过利用消费数据，利用 RFP, RFM 两张模型，对用户的购买情况进行分析，针对购买情况给出了营销策略，从数据层面上给出了一种制定针对客户的营销策略方案.

文献[5]详细的介绍了时间序列技术在电商市场预测中的作用，并对不同模型的实际应用做了分析，通过实验的方式论证了时间序列在实际应用中的可行性. 文献[6]注意到了有些电商销售数据的季节性变化特征，针对这一特征， 综合利用稳定季节性模式与支持向量回归模型对销量进行预测. 文献[7]利用了马尔科夫模型及时间序列模型预测了外汇汇率，这种组合预测的方法给本文在电商领域引入外部因素分析销量模型提供非常重要的借鉴与参考价值. 文献[8]利用隐马尔科夫模型，利用 4 个隐含状态，对股票走势进行建模，文献[9]也基于时间序列模型，结合人工智能，数据挖掘等领域的知识，深入分析了其在股市预测中的作用. 文献[10]通过研究商品销量与气温变化，提出了基于温度的销量预测方法.

2 模型假设

本文首先要解决多品类商品数据碎片化的问题，希望通过对数据的处理利用新的商品分类方法替换掉原有的商品分类. 然后再在新的分类下，利用预测模型对销售序列进行预测，但是目前常用的时间序列预测模型，其在预测的时候存在忽略动态变量的缺点，这里引入隐马尔科夫预测模型，利用定性的方法将时间序列模型的预测值进行定界，便于分析人员更高效准确的对预测值进行利用. 本文所假设一般处理模型如图 1 所示.

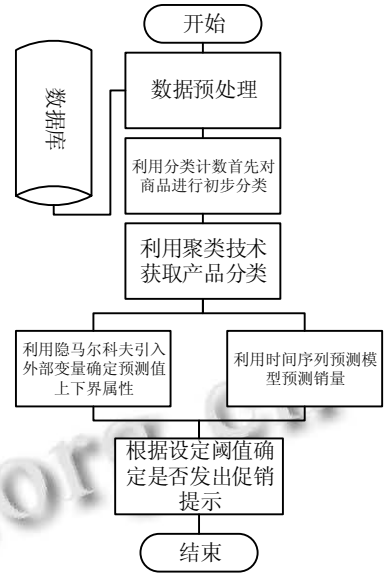

图 1 销量预测模型图

本文主要对利用聚类技术对商品重新分类部分与商品预测部分进行分析与可行性研究，对产生的预测值可能要利用到的处理规则与专家系统暂不做讨论.

3 基于聚类技术的商品重分类

3.1 数据预处理

本文主要研究的电商数据来自于某中型第三方电商企业，此类电商企业除了拥有自己的自建商城销售平台，大部分销售任务是通过各大电商平台进行的，因此这个级别的电商数据特征是分散于异构.

不同平台之间所使用的数据结果不尽相同，并且并非所有平台都提供数据接口供调用，淘宝店铺的商品数据提供文件导出功能，导出格式为 CSV 或者 excel，而销售数据可以通过 API 拉取获得. 自建平台的数据可以直接通过访问数据库获得，因此对于中型电商企业的数据而言，需要建立三种多规则的数据汇集程序. 汇集程序列表如表 1 所示.

表 1 数据汇集程序列表

数据获取方式	说明
API 数据获取	通过平台电商提供的 API 接口对数据进行读取转存
CSV/Excel 文件解析	编写文件解析程序解析文件内容对数据进行转存
数据库读取	直接读取数据转存

获取的汇总数据常常会存在字段丢失或者字段错误的情况，电商数据中除了销售价格，成本价格，折扣等字段外大部分都属于属性字段，例如产品名称，产品型号，收件人电话，地址等等．因此对于数据字段的缺失，不能采用均值，中位数等方法进行替换，但是由于例如商品信息及地址信息等存在大量冗余字段，因此采用建立冗余字段互补程序自动填充缺失字段．另外对于数据冲突的情况，由于中型电商企业的数据规模一般能达到百万条每年，因此对于低频次的冲突数据采用抛出人工处理的方式．数据预处理流程图如图 2 所示．

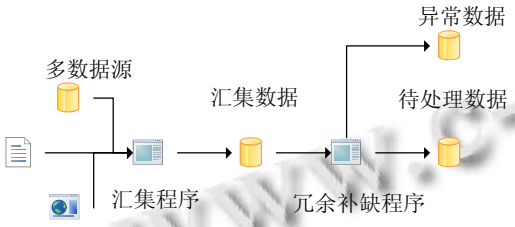

图 2 数据预处理流程图

3.2 商品数据的重分类

正常电商企业所生成的销售数据结构通常包含有商品自身的分类属性，但是其分类主要是以方便检索为目的，将相似的商品人为的或者按照某一商品属性化归为一类，这种固有的商品分类对于数据挖掘而言，存在着多种弊端．首先，当分类过于细化，会导致分类内的商品数量非常少，进而导致分类内商品的销量数据也比较少，影响数据处理中对于数据样本的规模要求，再来，对于相近商品而言，可能其具有本质的区别，同为饰品的丝巾与围巾，其在销售变化上是截然不同的．因而在对电商销售数据进行处理前，基于按地域划分的销售数据，对商品进行重新划分是非常重要的，这样才能反应出从特定角度具有相似特征的商品集合．

根据实际需求，可以从销量变化，折扣力度等角度对数据重新分类，而由于在分类之气按，实际上是无法确定商品能够分为几类，分类的标准是什么等，因而采用聚类技术，通过对组间的距离平方和除以整体距离平方和(between_ss/total_ss)收敛的情况进行判断来进行聚类分析．

3.3 利用决策树保留商品信息

上一小节中通过聚类的方法，对商品进行的重新划分，解决了商品原有分类对数据分析的一项，能够让具有一般共性销售特征的商品聚集在一起，便于数据处理和分析．但是，这种处理方式虽然能够屏蔽掉原有分类的干扰，同时也就损失了商品的一些相对重要的销售属性，或者当营销策略制定者希望能够区分某些特定商品时，当这些商品的销售序列特征又呈现相似特点时，如果直接使用聚类方法的话，就会导致丢失属性信息．

为了能够保有足够的商品信息，又能够发掘商品之间所共同具有的销售特征，在进行聚类前，可利用商品属性具有的信息熵大小构造满足要求的决策树，将商品划归到决策树中，再利用聚类算法对决策树中叶子结点中商品数据进行计算，获得特定分类下的商品聚类特征．

4 基于重分类商品的销量预测

4.1 时间序列

时间序列分析的主要目的是根据已有的历史数据对未来进行预测．电商的产品销售数据，是典型的时间序列数据，基于这样的时间序列，利用相应的时间序列模型，理论上可以通过对历史数据的拟合回归，对未来的销量进行预测．但是，不同产品的销量序列还需要区别对待[11]．

4.1.1 ARMA 模型

ARMA 模型即自回归移动平均模型(Auto-Regressive Moving Average Model,简称 ARMA)，该模型基本是由 Box-Jenkins[12]建立的，ARMA 又可分为三个子类型：AR 自回归模型，MA 移动平均模型和 ARMA 自回归移动平均模型[13]．ARMA(p,q)的形式为：

$$Xt=\alpha_1 X_{t-1}+\alpha_2 X_{t-2}+\ldots\ldots\alpha_p X_{t-p}-\varphi_1\varepsilon_{-1}-\varphi_2\varepsilon_{-2}\ldots\ldots\varphi_j\varepsilon_{-q}$$

这类模型一般用于处理平稳时间序列，在产品销量数据中，可以将这一模型应用于无趋势的，无季节周期的销售数据．

4.1.2 SARMA 模型

SARMA(Seasonal Autoregressive Moving Aveage)平稳季节时间序列模型，实际上季节模型本质上还是需要将序列的季节特性进行提取，在利用 ARMA 模型进行拟合．常用的处理方法有：将具有季节特性的数据利用周期特性进行相减或者利用滑动平均的方法消除季节特征．

4.1.3 ARIMA 模型

ARIMA 即 (Autoregressive Integrated Moving Aveage),非稳定时间序列通过差分等方法, 将时间序列转化为平稳序列, 再利用 ARMA 模型进行求解. 这一模型可以运用在夏粮具有一定趋势的销售序列中. 不同的商家, 由于发展各不相同, 因而其销量并不是完全随着市场的需求进行随机波动的, 而是具有一定的趋势特征, 如新兴商家, 通过合理经营与促销的手段, 让自身的销量在数年间增长数倍, 那么这个增长数倍就是贯穿整个销售数据的一个趋势.

4.2 隐马尔科夫

利用时间序列预测模型, 通过对不同品类商品的拟合, 能够从历史数据的角度给出一个可解释的预测值, 其实这样的事件序列预测值, 其中已经蕴含了诸如季节变化, 定期的促销活动, 因此时间序列预测模型的预测结果更像是黑盒测试, 其预测结果具有一定不可解释性. 因此, 基于时间序列预测模型的预测结果具有一定的局限性, 首先这样的预测值无法带入与历史差异因素, 对于以年为周期的销量预测, 诸如今年比去年温度更低, 促销力度更大这样的因素不能够在时间序列模型中更好的反应出来. 再来, 模型的预测值, 没有一个判断标准, 这个预测值应该是最大值还是最小值, 并没有一个合适的判断标准. 因而未来解决历史差异问题, 给时间序列预测值一个上下界的参考标准, 这里引入隐马尔科夫预测模型, 将一些可统计因素作为观测变量, 销量变化作为隐含变量. 用量化的方法, 对预测结果进行定性分析.

在使用隐马尔科模型前, 首先交代一下马尔科夫链所必须满足的假设:

(1) 0t+1 时刻系统状态的概率分布只与 t 时刻的状态有关, 与 t 时刻以前的状态无关, 即: $P(x_{t+1}|x_t \ldots\ldots x_1)=P(x_{t+1}|x_t)$;

(2) 从 t 时刻到 t+1 时刻的状态转移与 t 的值无关.

隐马尔科夫模型参数如下:

① $\mathrm{S}=\{S_1,S_2,\ldots\ldots S_N\}$:有 N 个值的状态集合.

② $\mathrm{V}=\{\mathrm{V}_1,\ \mathrm{V}_2\ldots\ldots\mathrm{V}_M\}$:有 M 个值的观测集合.

③ $\mathrm{A}=[a_{ij}]$:状态转移矩阵.

④ $\mathrm{B}=[b_j(k)]$:观测值的概率矩阵(混淆矩阵)

$$\mathrm{b}_j(k)=P(o_t=V_k|q_t=S_j)$$

$$\mathrm{j}\in[1,N],t\in[1,T]$$

⑤ $\pi=\{\pi_i\}$:初始概率分布.

$$\pi=\mathrm{P}(\mathrm{q}_1=\mathrm{S}_i),\mathrm{i}\in[1,\mathrm{N}]$$

这样, 一个马尔科夫模型可被标记为:

$$\lambda=(N,M,A,B,\pi)$$

其中, q_t 为 t 时刻的状态值, o_t 为 t 时刻的观测值[14].

这里以温度与销量变化作为两个观测序列为例. 每个月份的温度相对于去年同期增减情况作为观测序列, 那么观测序列就为{增长, 不变, 降低}, 增长变化的转移概率举证可以同统计方法获得. 如:

A: 下月相对温度上升; a: 本月相对温度上升

B: 下月相对温度不变; b: 本月相对温度不变

C: 下月相对温度降低; c: 本月相对温度降低

由全概率公式可得:

$$(P(A),P(B),P(C))=P(P(a),P(b),P(c))*A$$

$$A=\begin{bmatrix}P(A|a) & P(A|b) & P(A|\mathrm{c})\\ P(B|a) & P(B|b) & P(B|c)\\ P(C|a) & P(C|b) & P(C|c)\end{bmatrix}$$

其中转移矩阵 A 可以可通过统计气象历史数据获得. 这里的相对温度, 采用平均高温与平均的文的加权数值替换.

向量变化序列则为: {增加, 不变, 降低}, 温度变化关系与销量变化关系可以通过对销量变化统计获得, 即混淆矩阵也可以通过统计的方法获得.

假设向量序列为: {x,y,z}, x:增长, y:不变, z:降低, 则可以通过统计历史销售数据与气温变化关系, 其流程关系可见图 3, 得出混淆矩阵:

$$\mathrm{B}=\begin{bmatrix}P(x|A) & P(y|A) & P(z|A)\\ P(x|B) & P(x|B) & P(x|B)\\ P(x|C) & P(x|C) & P(x|C)\end{bmatrix}$$

$$(P(x),P(y),p(z))=(P(A),P(B),P(C))*B$$

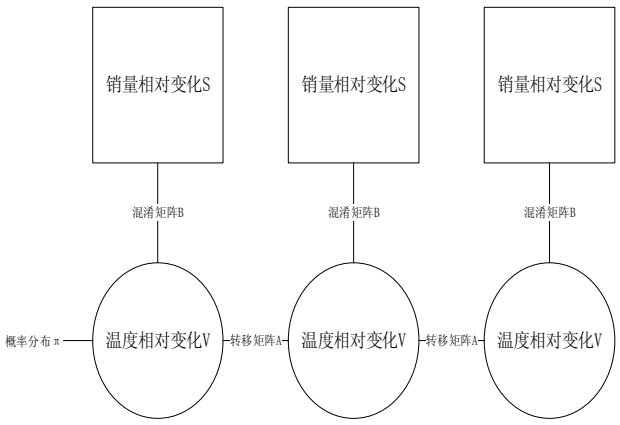

图 3 单因素隐马尔科夫预测模型图

通过转移矩阵与混淆矩阵，最终获取了下个月销量变化的概率分布，根据概率分布情况，通关过销售人员的经验规则，可以适当调整营销策略.

本文隐马尔科夫模型主要使用的是其一个外部因素观测值的情况下所做出的预测结果，对于多观测值的预测结果，还需要对各观测值之间的相关性做进一步研究. 如果两种观测值之间相互独立，则可以直接使用一个观测变量的隐马尔科夫预测模型进行直接叠加使用，分别给出两种因素在预测中所占的比例系数，两个预测值乘以比例系数后相加得到最终预测数值. 模型流程图如图 4 所示.

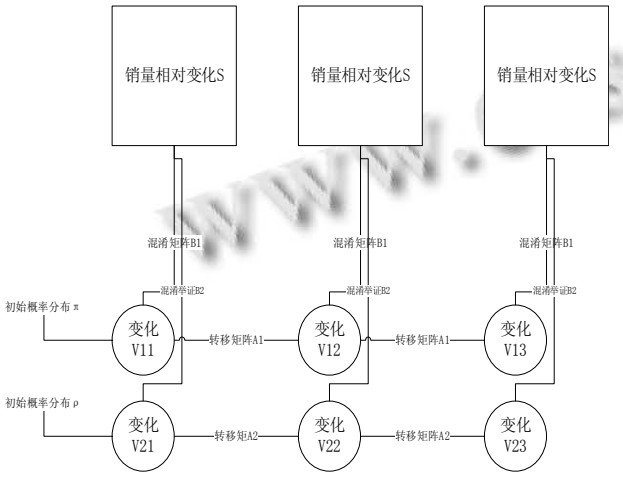

图 4　多因素隐马尔科夫预测模型图

对于非独立因素由于各因素之间存在相关性，相关性对于混淆矩阵的存在着一定的影响，如果不对相关性做出定量分析，混淆矩阵个比较难以得出. 故本文对具有相关性的多因素隐马尔科夫预测模型暂不做分析讨论.

4.3 预测值的应用

获取到两个预测模型的预测结果，还需要给出是否调整营销方案的结果，针对单品类商品可以利用变化百分比进行营销预警，这里可以利用营销人员的经验构造专家系统.

首先将时间序列预测模型的预测值作为历史预测值，其中不包含外边变化因素，这里可认为，在理论上如果所有历史条件不发生改变的话，这一预测值将会趋近于真实值. 但是每年处理与历史值相同的历史因素影响销量外，其他的一些可观测的与历史相异的因素也将影响销售，这个时候就通过观察隐马尔科夫预测模型的预测向量，对时间序列预测值进行边界定性分析.

5　实验

实验数据基于某运动服饰类电商 2013 来的真实销售数据，数据集规模超过 500W 条，字段包含，购买 ID，购买地址，商品货号，商品尺码，折扣价格，原始价格等. 数据包含大量商品，由于商品品类差别较大，且商品众多，因而在正对商品预测时，显然使用传统的分类方法有着极大的局限性，因而利用本文所提到的商品聚类算法，能够很好的得到可供时间序列分析的销售序列. 由于数据来源于第三方单品类商品卖家，故本文实验中跳过利用信息熵构造决策树的过程.

5.1 数据预处理与商品重分类

首先将销售数据按照地域，商品货号进行汇总，销售数据是按照销售顺序利用自增 ID 进行排列的，如果直接采用数据原有分类进行处理的话，将会极大增大模型个数和复杂度，如图所示，原有分类销量折线图，如图 5 所示.

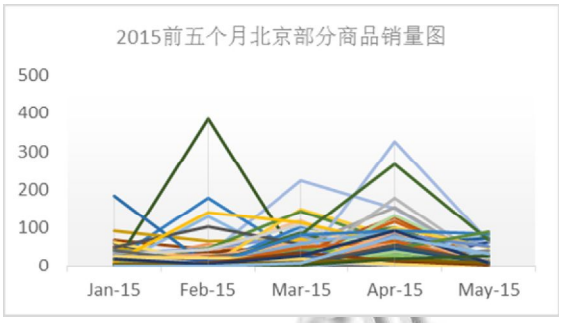

图 5　2015 年北京部分商品销量图

这里采用 K-means 聚类算法，对销售数据序列进行聚类，由于类团数量无法事先确定，理论上，越多的类团就会有更好的聚类效果，但是过多的类团将会影响数据的致密性，增加预测模型的复杂度，因而这里采用组间的距离平方和除以整体距离平方和(between_ss/total_ss)收敛的情况进行判断，当类团数量超过一定值时，其值会呈现收敛状态，如图 6 所示.

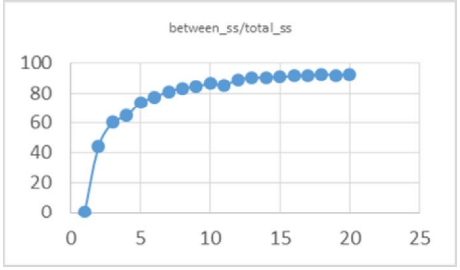

图 6　between_ss/total_ss

表 2 聚类结果表

类别	Jan-15	Feb-15	Mar-15	Apr-15	May-15	计数
1	0.1007	0.2606	0.5196	0.1092	0.0098	19
2	0.3491	0.2883	0.2411	0.1102	0.0113	40
3	0.0501	0.7583	0.1177	0.0690	0.0049	6
4	0.0533	0.0764	0.6146	0.2180	0.0377	16
5	0.0000	0.0000	0.0192	0.2446	0.7362	9
6	0.0034	0.0044	0.0750	0.8546	0.0626	34
7	0.0250	0.0447	0.8972	0.0331	0.0000	8
8	0.0037	0.0061	0.0862	0.6297	0.2743	28
9	0.1630	0.1968	0.3631	0.2542	0.0228	39
10	0.6826	0.1714	0.0981	0.0432	0.0047	13
11	0.0404	0.0780	0.4941	0.3532	0.0343	32
12	0.0241	0.0380	0.2861	0.5604	0.0915	45

类团中心折线图图像如图 7 所示.

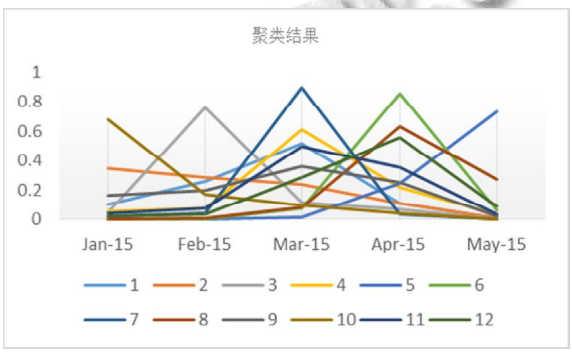

图 7 聚类结果图

5.2 销量预测

5.2.1 定量预测--时间序列预测模型

通过聚类技术获取的商品类别 1 其近年来销量序列如表 3 所示.

表 3 销量表

年份\月份	1	2	3	4	5	6
2013	1146	950	827	543	123	12
2014	982	931	754	400	76	24
2015	1096	925	763	310	33	
年份\月份	7	8	9	10	11	12
2014	7	9	27	110	531	782
2015	17	3	34	128	476	824

利用 spss[15] 工具的时间序列建模工具, 将数据输入 spss.创建时间序列, 并将 2015 年前五个月作为模型检验值进行预测. 最终预测结果如图 8 所示, 其中红色线条代表真实值, 蓝色线条代表预测值.

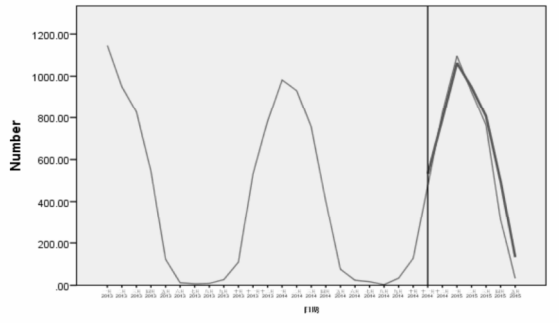

图 8 时间序列预测模型预测拟合曲线

预测结果表如表 4 所示.

表 4 时间序列模型预测结果

模型	2015-01	2015-02	2015-03	2015-04	2015-05
预测	1050.37	932.42	787.96	474.48	107.98
UCL	1116.37	1002.66	864.53	559.50	203.44
LCL	984.36	862.17	711.38	389.45	12.51

5.2.2 定性预测—隐马尔科夫预测模型

通过对北京月平均高温历史数据(数据见表 5)的统计得温度状态的转移矩阵为:

$$A=\begin{bmatrix}\frac{4}{13} & \frac{5}{13} & \frac{4}{13}\\ \frac{3}{7} & \frac{1}{7} & \frac{3}{7}\\ \frac{5}{12} & \frac{1}{12} & \frac{1}{12}\end{bmatrix},$$

混淆矩阵:

$$B=\begin{bmatrix}\frac{2}{7} & \frac{4}{7} & \frac{1}{7}\\ \frac{2}{5} & 0 & \frac{3}{5}\\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5}\end{bmatrix}.$$

利用 2014 年 12 月气温降低构造初始概率向量(1, 0, 0), 则利用转移矩阵预测 2015 年一二月气温变化向量为: (0.31, 0.38, 0.31), (0.39, 0.2, 0.41),从温度向量中可以得知, 在去年 12 月温度降低的条件下, 1 月温度比去年 1 月降低的概率为 0.31, 不变概率为 0.38, 升高概率为 0.31.分别用温度向量与混淆矩阵进行相称, 的销售变化向量为: (0.36,0.3,0.34),(0.36,0.39,0.25). 该序列意义为: 在去年 12 月温度高于前年的前提下, 1 月份该品类商品销量降低, 不变, 升高的概率分别为(0.36,0.3,0.34), 2 月份该类商品销量降低, 不变, 升高的概率为(0.36,0.39,0.25).

表 5 北京历史月均温统计表

日期	1	2	3	4	5	6
2012	1	4	11	21	29	30
2013	0	4	12	28	28	28

2014	5	3	16	23	28	31
2015	5	7	14	22	28	30
日期	**7**	**8**	**9**	**10**	**11**	**12**
2012	31	30	26	21	9	-1
2013	32	32	26	19	12	6
2014	33	31	25	19	12	4
2015	31	32	26			

6 结语

本文利用聚类技术, 改善了在处理电商销售数据时, 由于传统分类方法导致的数据割裂不完整的问题, 又通过两个角度利用两种预测模型对销售数据进行定量定性预测, 提高了预测值的可参考价值, 如果进一步与销售主管沟通构造专家系统, 能够一定程度上减少对销售人员经验的依赖, 降低误差. 从实验结果来看, 文中所建立的模型, 对选定商品的拟合程度非常高, 预测较为准确. 但是, 在商品聚类过程中, 依然会存在聚类结果不是非常满意的情况, 多个地区, 多个时间段的聚类结果之间存在的差异以及聚类数量都需要通过人工修正确认, 在隐马尔科夫预测模型中, 目前只引入了单变量, 对于实际影响销量的复杂因素模拟不足, 未来还有很大的改进空间.

参考文献

1 李博群.我国电子商务发展现状及前景展望研究.调研世界,2015(1):15–18.

2 马海霞.基于消费者信息行为的电子商务营销策略的研究.2006.

3 潘毅.精准营销在电子商务领域中的应用及策略研究[学位论文].北京:北京邮电大学,2013.

4 徐翔斌,王佳强,涂欢,等.基于改进 RFM 模型的电子商务客户细分.计算机应用,2012,32(5):1439–1442.

5 陈远,王菲菲.基于时间序列的电子商务市场预测系统研发.情报科学,2009,(12):1820–1823.

6 Ye F, Eskenazi J. Sales forecast using a hybrid learning method based on stable seasonal pattern and support vector regression. Emerging Technologies for Information Systems, Computing, and Management. Springer New York, 2013: 1251–1259.

7 Zahari A, Jaafar J. Combining hidden Markov model and case based reasoning for time series forecasting. Communications in Computer & Information Science, 2015, 513: 237–247.

8 余文利,廖建平,马文龙.一种新的基于隐马尔可夫模型的股票价格时间序列预测方法.计算机应用与软件,2010, 27(6):186–190.

9 李嵩松.基于隐马尔可夫模型和计算智能的股票价格时间序列预测[博士学位论文].哈尔滨:哈尔滨工业大学,2011.

10 辽宁省专业气象台 沈阳.夏季气温与商品销量市场预测及效益评价.气象与环境学报, 2002,2:22–23.

11 郭顺生,王磊,黄琨.基于时间序列模型预测汽车销量研究.机械工程师,2013(5):8–10.

12 潘红宇.时间序列分析及应用.2011.

13 Darcy S, Pegg S. Towards strategic intent: Perceptions of disability service provision amongst hotel accommodation managers. International Journal of Hospitality Management, 2011, 30(2011): 468−476.

14 侯雅文.基于隐马尔可夫模型的股票价格指数预测[硕士学位论文].广州:暨南大学,2007.

15 王周伟.SPSS 统计分析与综合应用.上海:上海交通大学出版社,2012.