

聚类算法在期货价格预测中的应用

李磊磊 吴友琦

(中央财经大学信息学院 100081)

【摘要】随着大数据时代的到来,越来越多的领域需要使用数据挖掘的算法来处理海量数据,以获得潜在有用的知识和信息。期货市场作为一个发展较成熟的市场,已经聚集了大量的数据需要决策者分析。本文使用聚类算法对期货市场的价格进行分析,获得一些有意义的信息,证明了聚类算法在期货市场应用的可行性。

【关键词】数据挖掘;聚类;期货市场

买卖期货合约,一部分人想投机获利,另一部分人则想回避价格风险。因此,找到一种期货市场价格分析方法预测期货价格走势,对于那些希望从市场价格波动中牟利的投资者,或希望未雨绸缪、保障自己不受价格急剧变化影响的生产经营者来说极具吸引力。

一、期货市场分析方法概述

期货分析方法可分为基本分析法和 technical 分析法。基本分析法是一种运用供应和需求信息预测未来市场价格变化的分析方法。基本分析法所依据的经济学原理是:商品的价格由需求和供给的变化决定,商品的市场价格由供求力量的均衡形成的。不同于基本分析法,技术分析法是一种利用历史价格、交易量、空盘量和其他交易数据预测未来价格趋势的价格分析方法。技术分析法是通过对期货品种的价格、交易量与持仓量资料,按照时间顺序绘制图形或图表,然后针对这些图形、图表或指标系统进行分析研究,以预测期货价格走势的方法。

期货市场的分析需要综合使用各种方法,取长补短,才能使预测更为准确。此外,人们也试图用回归分析等统计手段建立模型来预测期货市场,但由于期货市场没有一个科学的明确的分类,这些方法往往简单的对各种期货进行研究分析,盲目而效果不佳。我们需要借助更复杂的计算机技术,获得较好的效果。

随着大数据时代的到来,数据挖掘作为处理海量数据的利器,逐渐受到各领域的重视。

二、聚类算法

聚类作为数据挖掘研究中的一个重要的研究方向,逐渐受到人们的重视,已经成为数据挖掘研究领域中的一个十分活跃的研究课题。聚类就是将数据对象分组成为多个类或簇,划分的原则是在同一个簇中的对象之间具有较高的相似度。聚类算法可以分为:基于模型的方法、基于网格的方法、基于密度的方法、层次法、分裂法。各种聚类算法都有其各自的特点,有些方法对任意大小、形状的类识别能力非常强(如 clustering using border and nearest (CUBN)),有些原理简单、执行效率高(如 K-means),有些能够很好地过滤噪声数据(如 DBSCAN),但这些算法都有其各自的局限性,如 CUBN 的时间复杂度高($O(n^2)$),K-means 只能识别近似球形类的数据。

三、期货价格的聚类与分析

本文使用 k-means 算法对期货合约价格进行研究,通过对聚类结果分析,发现一些对期货价格预测有帮助的结论。

(一) k-means 算法思想

K-means 算法首先从 n 个数据对象中任意选择 k 个作为初始聚类的中心,该中心称为所在聚类的质心,再根据剩下的数据对

像到这 k 个质心的距离,将剩余数据对象分配给离其最近的(相似度最高)的质心所在的聚类中。然后以减小目标函数值为方向,通过迭代不断调整每个类的质心和数据对象在类中的分布,使得类内随迭代渐趋紧凑,类内对象相似性不断增大;而类与类之间逐渐独立,类间对象相似性越来越小。当迭代进行使得目标函数收敛,相邻两次算法的聚类中心相同时,聚类调整结束。

(二) 数据选取与数据预处理

本文选取中国期货市场具有代表性的农产品期货(豆一、豆二、玉米、鸡蛋、豆粕、棕榈油、豆油、粳稻、菜油、早籼、菜籽、白糖)为样本进行研究,数据来源于大连期货交易所和郑州期货交易所官网。选取期货合约的开盘价、最高价、最低价和收盘价四个指标进行聚类,期货样本时间选取 2014 年 2 月份每日数据。对原数据中的不完整数据清洗,对不一致数据进行归一化处理,使数据整齐、有效。

(三) 聚类结果分析

本文使用 matlab 软件实现 k-means 算法,将处理后的数据输入程序,对聚类结果观察分析得出以下几个结论:

1、大部分期货品种的价格稳定在一个簇内,这些期货品种与其他期货品种价格差距较大,而且价格波动较小,有的价格很稳定,几乎不怎么变动;

2、除了固定在一个簇内不发生变动的期货品种外,其他期货品种大体呈现三种趋势:

(1) 在相邻两个簇之间不规则的摆动,且摆动频繁,说明该期货品种价格波动较大且频繁,对这类期货投资需谨慎;

(2) 在相邻两个簇之间偶有摆动,或是长期处于高价簇,偶尔向相对低价的簇摆动,或是长期处于低价簇,偶尔向相对高价的簇摆动,这种摆动必然是收到当时市场上的政策或其他因素的影响,但该因素对期货价格的影响较小,对于投机者需要把握这种机会;

(3) 从一个簇摆动到另外一个簇,并长期保持不变。或是从低价簇摆动到高价簇,这种摆动意味着期货价格在这一段时间内会持续走高;或是从高价簇摆动到低价簇,这种摆动意味着期货价格在这一段时间内会持续走低。

3、结合期货市场的技术分析指标,对于比较稳定的期货品种价格可以用简单移动平均线(MA)来预测未来走势,而对于在簇间摆动的期货品种,可以用平滑异同移动平均线(MACD)来预测价格走势。

经过对期货价格折线图分析,以及期货价格和以上两种技术分析指标间相关系数的计算,证明以上基于聚类结果的价格分析都是正确的。同时也证明了数据挖掘的聚类算法在期货市场价格分析中应用的可行性。

【参考文献】

- [1] 田军.基于数据挖掘技术的期货投资分析与预测系统研究与实践[M].华南理工大学,2003
- [2] 杨丽华,魏宗有.谈技术指标在期货价格预测中的应用[J].商业时代,2011(2)
- [3] 毛国君,段立娟等.数据挖掘原理与算法(第二版)[M].清华大学出版社,2012