## 《计算机新技术讲座》报告写作指导书

叶宏庭 （学号：71118415）

**一、报告内容**

|  |
| --- |
| 1.报告信息：讲师王君，IBM资深大数据分析架构师、数据科学家。 作为技术带头人，目前除了带领IBM SPSS研发团队为IBM大数据分析产品如DSX，WA，Modeler，Statistics等开发机器学习算法，同时也为企业做大数据分析方案，多次主持并带领团队为企业量身打造大数据分析平台，涉及的领域有交通、制造业、银行等。本次讲座使用腾讯会议，线上开展。  2.报告内容：众所周知，在行业内部及行业之间的各种股票，它们存在着相互影响的关系。 如何在众多的有纷繁复杂关系的股票中识别出一些关键的具有代表性的股票，而这些股票往往是对行业内外的其他股票有着决定性的影响，挖掘并量化这些影响并将其用于股票的量价分析交易中有着很大的“钱“途。然而完成这项工作却有着很大的挑战性。我将详细介绍如何应用IBM Watson Studio产品中的时间因果和时空预测机器学习算法来解决这一难题，同时也会分享我们是如何使用IBM WS以团队协作方式轻松高效地进行股票量价分析的过程。 |

**二、收获和扩展学习**

|  |
| --- |
| 1.报告收获：如何分析行业之间的关联性？这就用到了二维时空分析法。主要流程如下：首先进行数据探索，探索完之后进行处理。因为这个模型要求做时间的预测，要求时间的序列必须得是完整的，连续的，不能缺失。所以就需要去做数据预测模型，要求得到真实的可以读取的数据。最后提取出真正影响比较大的，当然这只是初步的，有可能有偏差。然后通过技术做时间上的分析，分析出两只股票变化趋势。然后再进一步做空间上的分析，利用关联性做一些时间分析上的修正。然后将这些数据提供给有关部门，有关部门就可以去分析，其背后的原因是经营策略还是干预手段，同时为检验规划是否有效做侧面印证。  2.扩展学习：用数据创造价值不单单是拥有数据，而是要了解数据，了解商业模式，将对数据的理解和挖掘出的内在的逻辑和洞察应用到商业模式中去，才能真正地体现数据金矿的价值。数据探索，即用统计学指标对数据做尝试性分析，基于不同的统计指标，将数据划归为不同类型的问题，然后选取模型进行分析。有了初步印象，后续操作就更有目的性。数据预处理，即在同一尺度下补充缺失值、修正异常值、将连续数据离散化。这依赖于一个自动处理的算法和数据筛选，软件会筛选出一些关键的因素，最终利用这些因素去进行操作。建模中一种方法可根据输出结果的离散与连续性不同可以分为回归和分类问题。例如，有一个银行信贷员，他已经测算过700个客户的信贷业务，收集到了这700个客户的基本信息，对这些信息进行分类，主要目的是为了判断给哪些人贷款风险高。预测方法包括时间预测、空间预测。例如，运营商决定在哪些地方增加网线密度时，则需通过数据分析，得出特定范围内用网量变化较大的时段与地点，进行网线的添置，满足客户需求，又避免资源浪费。又如气温也是随时间和空间进行连续性变化，想要预测温度，则需使用聚类，将相似的归到一起，找到关联性。集成，即将多种模型的预测进行整合。因为一个事件往往会受到多种因素的影响，可能无法满足客户对高精度的要求，而集成正是为了弥补这一缺陷。如何分析行业之间的关联性？这就用到了二维时空分析法。主要流程如下：首先进行数据探索，探索完之后进行处理。因为这个模型要求做时间的预测，要求时间的序列必须得是完整的，连续的，不能缺失。所以就需要去做数据预测模型，要求得到真实的可以读取的数据。 |