Rossmann销售预测开题报告

# 项目背景

作为销售计划的中心任务之一，销售预测是指通过一定的手段和方法，对未来特定时间内全部产品或者特定产品的销售数量或销售金额的估计。在操作层面上，销售预测是在充分考虑了未来各种影响因素的基础上，结合本企业的过往销售业绩，运用一定的分析方法提出切实可行的销售目标。

影响销售的因素很多，包括有需求变化、经济变动等成分的外界因素，和有营销策略、生产状况、销售人员等成分的内部因素，但销售预测对于完善客户需求管理、指导运营、优化供应链、提高企业利润方面都具有重大促进作用，降低企业的业务计划的不确定性，因此，提出合理的销售预测一直也是人们孜孜不倦的追求，是企业辅助决策的重要工具。

多年来，人们已经形成了定性分析法和定量分析法两类分析方法。其中，定量分析法中的趋势预测分析法和因果预测分析法在实际应用中也能取得一定的预测效果。他们可以基于历史数据，运用一些直观的算法，如算术平均法、指数平滑法，来进行预测，相比定性分析法，预测效果也有了一定的提高[1]。

但随着经济全球化，商业网络化的进程，市场竞争日趋激烈，业务复杂化，数据规模海量化，传统的预测方法已经越来越力不从心。预测的精确性围城销售预测的核心痛点，人们迫切希望一些性能更高，更智能的预测方法。数据挖掘[2]技术由计算机自动从大量数据集中提取隐含的、事先不知道但有潜在应用价值的信息，可用于学习复杂销售的复杂特征对于销售的影响，从而得到较好的预测效果。本课题将应用机器学习算法来实施销售预测。

# 问题描述

Rossmann是欧洲的一家连锁药店，在欧洲7个国家拥有超过3000家药店。这是一个Kaggle[比赛项目](https://www.kaggle.com/c/rossmann-store-sales)，本课题将按照项目中的说明，需要为Rossmann在德国的1115家药店做出提前6个星期的每日销售预测。

对于Rossmann的销售预测问题将是一个具体领域的销售预测问题，作为药店连锁店，具有一定的行业特征，这些将体现为数据特征。如上节所述，基于机器学习算法的数据挖掘技术会是一种可能的得到一个合适预测模型的办法。

本课题得到的预测模型将用于输出未来6个星期里每天的销售量，预测结果可以和实际销售情况对比，从而衡量预测效果。在实际使用中，还可以随着时间的推移，不断学习和预测。

# 数据特征

作为一个Kaggle比赛项目，Kaggle提供了项目数据，包括训练数据train.csv，测试数据test.csv，已经商店信息数据store.csv。本项目将以离线的方式训练模型，并作出预测，以比赛项目的训练数据作为项目数据，分割出部分（0.5%）作为测试数据。

训练集数据的全部字段说明参见：<https://www.kaggle.com/c/rossmann-store-sales/data>，本项目将选取如下特征值作为销售预测：

* Store：商店编号
* Date：数据统计日期
* Open：当天门店是否开放，节假日商品可能不开放。
* Promo：当天门店是否进行促销活动
* StoreType：门店类型
* Assortment：分类标准
* CompetitionDistance：最近竞争门店的距离
* CompetitionOpenSinceMonth：竞争门店开店月份
* CompetitionOpenSinceYear：竞争门店开店年份

而标签，即要预测的数据为：

* Sales：当天销售额

时间序列的数据是本项目数据的一大特征。

# 解决办法

神经网络是一种非线性自适应系统模型，

# 基本模型

模型是什么

# 评估指标

指标是什么

# 设计大纲

设计方法为

# 参考文献：

1. <http://wiki.mbalib.com/wiki/%E9%94%80%E5%94%AE%E9%A2%84%E6%B5%8B>
2. 数据挖掘：<https://en.wikipedia.org/wiki/Data_mining>

文献：

<http://www.chinaaet.com/article/15625>

https://github.com/Cospel/kaggle-rossmann