

基于矩阵分解算法的智能租车系统

苏斓涛

(武汉东湖学院计算机科学学院 湖北武汉 430212)

摘要:在如今世界越来越向快捷化和智能化方向发展,从而高效率 and 便利性是人们越来越注重的方面。而对于广大的居民和游客们来说,租车是一种很常见的出行方式。针对这种情况,基于人工智能技术的租车系统应运而生。基于此目的以机器学习技术为切入点,探讨机器学习技术在租车系统中的运用。本篇文章运用矩阵分解算法来提升车辆信息检索效率。我们通过 ML.NET 技术来实现这种算法。结果,矩阵分解算法使信息检索效率提升了 29.79%。

关键词:矩阵分解,智能租车系统

一、引言

在如今世界越来越向快捷化和智能化方向发展,从而高效率 and 便利性是人们越来越注重的方面。而对于广大的居民和游客们来说,租车是一种很常见的出行方式。在当今以互联网为基础,以人工智能为助力,AI+正在全国大力发展,租车领域也应当顺应潮流加入 AI+的领域中,从而给租车行业带来新的活力。AI+背景下的租车系统对以下几点有较高要求,第一是管理大量的车辆,第二是归档客户租聘车辆和归还车辆的数据。租车系统应该适应新时代的技术变革,运用人工智能技术于租车系统服务中,建造出集信息化、智能化和服务化为一体的管理体系,从而转变传统租车的理念。为保证车辆信息检索的优质性,加强数据存储、采集和分析,租车系统应引入机器学习进行相关的建设和升级。

本文的目的是在矩阵分解算法的基础上运用 ML.NET 开发智能租车系统。本文的其余部分阐述如下。在下一节中,我们介绍了机器学习与 ML.NET 的一些基本概念。在第 3 节中,我们运用矩阵分解算法来开发租车检索系统。在第 4 节中,我们总结了本文并给出了一些意见。

二、前言

(一) 机器学习技术

机器学习技术用资料 and 答案,来得到 AI 模型,再把没有答案的资料给这个 AI 模型,从而得到答案。注意:有时候是没有答案的。机器学习主要分为监督式学习和非监督式学习俩大块。监督式学习是知道答案的,例如告诉 AI 哪些是猫,给猫的资料让 AI 去学习,AI 就能分

辨出哪些是猫。非监督式学习则是不知道答案,但通过非监督式学习算法,AI 能自动去把相似的资料进行归类。例如给 AI 一些图片,但图片没有标记也不知道从何标记,那么非监督式学习 AI 就能初步把这些图片分类,然后再进一步进行标记。我们在此租车系统中使用的矩阵分解算法是无监督学习算法。

(二) ML.NET

ML.NET 是微软的一套开源的,跨平台的 AI 开发框架。开发者可以轻松的使用 C#来做 AI 的开发,有很多内建好的机器学习算法让你快速的做出成果。你可以在 .NET 应用程序中把机器学习加入进去。通过 ML.NET 可以对数据进行自动预测。应用程序不需要进行显示编程,而是利用 AI 模式来进行预测。

三、矩阵分解算法在租车信息检索系统中的应用

本系统中预测用户给予特定车辆的评分 (1-5) 并在评分高于定义的阈值时推荐该车辆(评分越高,用户喜欢特定车辆的可能性就越大)。首先创建一个名字为“CarRecommender”的 C# 控制台应用程序,运用 .NET 6 作为使用框架。在项目中创建一个数据文件夹来存储数据集文件。

(一) 加载数据

准备两个数据集,分别为 Train 和 Test 数据集。训练模型用 Train 数据。Test 数据用于使用经过训练好的模型来进行预测并评估模型性能。在这里我们采用 80/20 来对 Train 和 Test 数据进行拆分。

数据集中包含三张表,分别为租车用户表,车辆表和车评表。*.csv 文件中用到的字段主要有 userID, carID,

carRating。在机器学习技术中，用于进行预测的字段称为 Features，带有返回预测的列称为 Label。所以我们用 carRating 字段作为 Label。其他两列，userID，carID 都是作为 Features 从而用来预测 Label。车评表对应的 CarRating 类，含有 userID，carID，Label，这 3 个属性。再创建另一个类 CarRatingPrediction，用来表示预测结果，该类含有 label 和 score 两个属性。

用 MLContext 类的 LoadData 方法加载 *.csv 文件里的数据，并返回 Train 和 Test 数据。

(二) 制造并训练模型

创建 BuildAndTrainModel 方法，在该方法中选择矩阵分解算法来作为训练算法。当你掌握用户过去如何评价车辆的数据时，通常建议使用矩阵分解方法。在本智能租车系统中，Matrix Factorization 算法使用了一种称为“协作筛选”的方法，该方法假设如果用户 1 和用户 2 对车辆的评分相似，那么用户 2 更有可能欣赏用户 1 已使用并给出很高评分的车辆。然后把前面提供的训练数据集用 fit 方法来训练模型。最终生成训练好的模型。

(三) 评估模型

在训练模型后，我们使用测试数据来评估模型的执行情况。创建 EvaluateModel 方法，用 model 的 transform 方法对测试数据集里的多个的输入行进行 carRating 预测。获得预测集后，EvaluateModel 方法会对模型进行评估，该模型会将预测值与测试数据集中的实际 Labels 进行比较，并返回有关模型执行情况的指标。在此输出中，有 20 次迭代。在每次迭代中，误差测量值均会减小并逐渐趋于最小值 0。root of mean squared error 用于度量模型预测的值与测试数据集观察到的值之间的差异。从技术上讲，它是误差的平方的平均值的平方根。指标越低，模型就越好。R Squared 指明数据与模型的适应程度。范围从 0 到 1。值 0 表示数据是随机的，否则就无法适应模型。值 1 表示模型与数据完全匹配。通常会希望 R Squared 分数尽可能接近 1。

(四) 使用模型

创建 UseModelForSinglePrediction 方法，在该方法里使用 PredictionEngine 来预测车辆评分。PredictionEngine 是一个简便 API，可使用它对单个数据实例执行预测。

然后，使用 Score 或预测评分来确定是否要将特定车辆推荐给特定用户。Score 越高，用户喜欢特定车辆的可能性就越大。在这种情况下，假设系统推荐预测评分大于 3 的车辆给用户。基于内容的推荐、基于关联规则的推荐和矩阵分解算法的性能比较如表 1 所示。表中使用平均绝对误差(mean absolute error, 简称 MAE)和均方根误差(root mean squared error, 简称 RMSE)作为实验结果的评价指标。

表 1 Performance comparison

| Metric | Content-based Recommendation | Association Rule-based Recommendation | Matrix decomposition algorithm | Improve(%) |
|--------|------------------------------|---------------------------------------|--------------------------------|------------|
| MAE | 1.103 | 0.983 | 0.849 | 30.09 |
| RMSE | 1.392 | 1.215 | 1.075 | 29.49 |

(五) 保存模型

创建 SaveModel 方法保存模型，从而在最终用户应用程序中使用该模型进行预测。

四、结语

综上所述，智能租车系统中，机器学习技术的应用赋予系统相关功能的自动化、智能化，经过数据信息的多方面分析和预测，从而提高车辆信息检索的精准度，效率提升了 29.79%。租车的服务质量得到了全面的提升。不过在信息检索方面，对于从来没有做过车辆评论的用户，在今后可以采用其他算法或策略来优化这部分用户的信息检索体验。因此在后期发展中，租车应该进一步加强机器学习技术在各个方面的研发与应用，加快智慧租车系统的升级和改造。

参考文献:

[1]吕云翔,马连韬,刘卓然,张凡,张程博. 机器学习基础.[M]北京:清华大学出版社,2018

[3]牟少敏. 机器学习与大数据技术.[M]北京:人民邮电出版社,2018

[3]Dino Esposito, Francesco Esposito. 跨平台机器学习: ML.NET 架构及应用编程.[M]北京:清华大学出版社,2022