

机器学习技术在数据挖掘中的商业应用

文/张昊泽

摘要

现如今,数据挖掘中机器学习技术的应用非常广泛,遍及金融、医疗等领域,这就要求企业应该提升自身辨别方法和技术优劣的能力,面对不同的环境,选择应用与任务相匹配的数据挖掘技术。本篇文章首先对机器学习技术的界说作了简述,其后对机器学习技术在数据挖掘任务的商业应用作了两点详细的研究,以期促进企业的可持续发展。

【关键词】机器学习技术 数据挖掘 商业应用

随着科学技术的发展,存在于数据挖掘领域的机器学习技术凭借其强大的数据处理能力和在商业领域上巨大的应用潜力越发得到数据挖掘领域学术界以及商界的广泛关注。因此,对机器学习技术在数据挖掘中的商业应用进行探究有着十分重要的现实意义。

1 机器学习技术的界说

通过自动化的计算方法来获取知识的学习就是机器学习。人工智能研究中,机器学习占据着十分重要的位置。可被称作真正的智能系统,一定具有非常强悍的学习能力,由此可得出,以往的大多数系统都不能称作智能系统。比如,不能进行错误的自我校正;不能经过经验改良系统性能;不能自发的取得和找寻系统所需的大量知识。这些系统的推理都是演绎推理,缺少归纳推理,这使得仅能在最大程度上证明已知的定理和事实,不可能有新的发现。这个局限随着人工智能发展的逐渐深化变得更加明显。在此种情况下,机器学习技术已然是人工智能的研究重心之一。其应用广布人工智能领域的智能机器人、自然语言理解等分支。

2 机器学习技术在数据挖掘任务的商业应用

2.1 机器学习技术的分类

机器学习可依据不同的推理方法,划分为归纳、分析式以及非符号这三种学习方式。在例子存在的空间中,归纳学习的系统利用一般操作和特殊操作的方法对给出的正、反这两种有关于某个概念的集合例子进行在空间中的搜查和检索,还要进行经过学习倾向标准的抉择过程,最终得到这个关于某概念的一般描述。这是从特殊到一般的推理,可称做从事实出发来进行理论的最终建立过程。学习进程一般由大量的专业领域知识的运用来驱动,是分析式

学习的一个特点。此时实力已在特定的知识背景环境中得到大量的考察,不再是无关特征的、单纯的几何,对学习结果进行了倾向影响的消除。其还可划分为类比、基于案例以及解释这三种学习方式。神经网络与算法基因都属于非符号学习的范畴,将学习行为在数值层次中体现是其主要的特点。博斯和马哈帕特拉将机器学习技术在数据挖掘中的应用,作了下述五种归类。

2.1.1 归纳推理

依据各不相同的特征,应用决策树进行数据分类。应选择一个可作为因变量的变量在数据源中,再应用组间差别扩大与组内差异差别缩小的统计方法进行变量的分组,找寻对因变量影响最大的预测变量,此时可将该变量作为决策树中的一个节点来设置。决策树的优点是直观,缺点是其分支随着数据复杂度的提升会逐渐增多,这使管理变得困难。强大的大数据集处理能力,任务的适于预测和分类性,解释简易的结果,技术实施的简易性于一体。

2.1.2 神经网络

由与人脑神经元类似的节点来构成,是一个多层的网络结构,通过隐藏节点将输入和输出这两个节点连接起来而构成。将凡是通过和到达该节点进行信号输入的进行加权所得出的加权和该节点的信号输入总量。该网络的学习方式是将样本的历史数据进行大量和反复性的训练,在此期间,可通过数学方法——学习规则使节点进行数据的汇总与转化,并进行链接节点权值的调节。其有互相间相连的输入、中间以及输出这三种层次构成,网络工作中的绝大部分由多节点构成的中间层来完成。对分析数据的执行结果进行输出是输出层的主要工作。可进行复杂性问题的精确性预测是该网络的优点。但是其存在着大数据集的处理效率过低的缺点,使用该方法的用户需具备大量的专业知识。

2.1.3 事例推理

关于问题的描述和解决办法是每个事例都具备的内容。问题提出之后,学习系统会进行匹配性事例和解决办法的寻找。较强的污染与缺失数据处理能力是其优点,常适用于存有大量数据的领域。

2.1.4 遗传算法

是一种优化组合方法,基于生物进化的适者生存思想,有繁殖、杂交以及变异这三个基本的操作过程。个体适应度的计算都源于繁殖操作过程。进行信息某一部分的交换过程称为变异操作过程。通过某一部分信息随机改变得到的新个体属于变异操作过程。最有个体的特色,往往得益于重复性的变异操作。易于集成较强的污染与缺失数据处理能力是其优点,但使用该方法的用户需具备大量的专业知识。

2.1.5 归纳性逻辑程序

用一级逻辑属性来进行概念的描述与定

义。先对正、负两面的例子进行定义,其后对新例子的等级进行划分。强大的大数据集处理能力和理解容易的模型是其优点,但使用该方法的用户需具备大量的专业知识。这五种方法由上到下依次是商业应用最广泛的归纳推理方法。

2.2 机器学习技术在数据挖掘任务的应用

博斯和马哈帕特拉将机器学习技术在数据挖掘中的商业应用归结为下述四种任务类型:

2.2.1 分类

比如,从商业数据库中,应用数据挖掘进行有效信息的挖掘,依据统一偏好或是年贡献估计额等标准来进行全部客户的分类。

2.2.2 预测

比如,当顾客有贷款的需要时,银行系统应在第一时间对其的信用状况进行审查,应用机器学习技术,就可在日常中对存在于数据库中的源数据进行不断的学习和修正,得出的信息也就是最具参考性的。

2.2.3 关联

潜存于实体间或属性间的联系规律进行关联性的分析。第四,侦察。进行异常现象、模式以及离群数据的寻找,并对决策给出支持作用的解释,是其主要目的。

3 结束语

总的来说,很多领域都应用到了数据挖掘中机器学习技术,这就要求企业应该知晓不同方法和技术的优劣,面对不同的环境,选择应用与任务相匹配的数据挖掘技术。本篇文章首先对机器学习技术的界说作了简述,其后对机器学习技术在数据挖掘任务的商业应用中机器学习技术的分类和机器学习技术在数据挖掘任务的应用这两点作了详细的研究,以期促进企业的可持续发展。

参考文献

- [1] 朱天元. 机器学习算法在数据挖掘中的应用[J]. 数字技术与应用, 2017(03): 166-166.
- [2] 张绍成, 孙时光, 曲洋等. 大数据环境下机器学习在数据挖掘中的应用研究[J]. 辽宁大学学报(自然科学版), 2017, 44(01): 15-17.

作者简介

张昊泽(1996-), 男, 黑龙江省大庆市人。大学本科学历。大连海事大学航运经济与管理学院。主要研究方向为信息管理与信息系统。

作者单位

大连海事大学 辽宁省大连市 116026