

基于深度神经网络的大学生就业情况分析 with 推荐

李 昊

(郑州轻工业大学, 河南 郑州 450000)

摘要: 随着高等教育的普及化, 高校内大学生数量与日俱增, 与此同时大学生的就业情况也备受瞩目。由于学生数量众多, 且个人的就业偏向和特点不同, 这对大学辅导员为学生记性就业指导带来了一定的困难。本文依据现有的机器学习方法, 结合现有数据, 对大学生就业情况进行了分析并给出推荐结果。文章首先收集了 3000 份学生数据, 数据包括了个人爱好, 学科成绩, 就业偏向的样本特征, 采用机器学习算法对样本数据集进行仿真分析。利用机器学习的方法, 可以大大提高推荐效率, 且实验验证表明, 本文设计的模型推荐准确率极高。

关键词: 就业; 神经元; 隐层; 深度神经网络

中图分类号: TP183

文献标识码: A

文章编号: 1673-1131(2020)08-0196-04

0 引言

随着高校招生规模不断扩大, 截止 2019 年, 我国高等教育毛入学率超过 50%, 高等教育从精英教育阶段进入普及化教育阶段^[1]。毕业生就业需求逐年增加^[2], 企业对毕业生的专业技能, 综合素质要求越来越高。以往传统的就业指导模式

已经不能适应新的就业供需关系。如何为毕业生提供个性化、精准化的就业服务, 成为高校工作的重点。

数据显示, 2019 年全国大学毕业生达到 834 万^[3], 再创新高。当前世界贸易保护愈演愈烈, 经济增长处在新旧动能替换中, 这些因素叠加使我国经济存在较大下行压力。与此同

收稿日期: 2020-07-01

作者简介: 李昊(1990-), 男, 河南郑州人, 硕士研究生, 助教, 研究方向: 模式识别。

性和可靠性^[4]。依托统一的校务管理平台共享各部门间的数据, 密切各部门间、师生、教师之间的信息沟通, 也能够让职业院校的领导获得更多有价值的信息, 并科学的分析, 为决策管理提供科学的参考。总之要积极开发管理职业院校的信息资源, 并建立完善的、规范化的信息化标准与评估体系, 为信息化建设的顺利推进打下良好的基础。

2.4 准确定位信息化管理机构

为顺利推进信息化建设, 职业院校要对信息化管理机构进行建立与定位, 明确其职权范围, 使其具有独立性, 实现集中性的管理、同时建立专业化技术小组, 为信息化建设中基础设施的建设、维护、管理等提供保障。建立领导小组、专家小组和实施小组。由领导小组对整个建设进行组织协调与决策; 专家小组对信息化建设的评估、论证与指导^[5]; 实施小组对信息化建设的实际操作负责。此外, 职业院校各系、学院等要建立数字化信息联络员, 为管理协调提供便利。职业院校领导小组统一规划、指导, 实现数据、设备、应用等全部集中, 顺利推进学校信息化建设。

2.5 强化信息化人才培养

职业院校要结合信息化建设发展的需要, 建立完善的人才培养机制, 结合行政管理、岗位需要等确定用人和人才引进机制, 吸引更多高水平、素质能力强的专业化信息人才进入到职业院校中, 为信息化建设服务。首先, 要注意专业技术人才的培养, 使其掌握现代化的信息理论知识与技术, 从而建立技术硬、责任意识强、满足信息化建设发展需要的信息技术维护人员, 科学的设计信息系统, 使其科学有效运行, 保证信息化建设中的问题得到妥善的处理。其次, 要强化职业院校教师信息化技术水平。注重教师思想理念的转变, 使其认识到信息化技术在现代教育领域中的重要作用, 树立信息化的意识理念, 并善于运用信息技术开展教育教学工作。同时加强教师基本技能的培训, 掌握信息技术的知识与操作技能, 使其具备获取、处理、应用教学信息的能力, 将信息技术与学科知识相结合, 建立新型的教育理念。

2.6 发挥先进信息技术手段

信息化时代背景下, 云计算、物联网技术等快速发展, 其在职业院校信息化建设中发挥着至关重要的作用。云计算技术能够集中信息资源、网络、服务器、储存等, 利用云信息技术将其定义为虚拟服务, 通过“租赁”的方法提供给用户^[6], 其在职业院校中应用有助于信息化革命的开展, 促进学校新一代数据中心的建设。物联网技术推进了平安校园的建设, 建设物联网智能图书馆、物联网课程等, 实现学生的有效管理。这些信息技术的发展与延伸, 使职业院校逐步实现智慧化校园建设。

3 结语

总而言之, 新时期职业院校评价标准中, 信息化水平已经成为重要的指标, 职业院校在信息化建设中要遵循自身特点, 结合职业院校的定位制定科学的信息化建设方案, 完善机构、制度、人才等建设, 强化资金投入, 为信息化建设奠定坚实的基础, 促进数字化资源、信息技术等高效应用, 为职业院校现代化发展实现提供支持。

参考文献:

- [1] 李洁. 互联网背景下高职院校财务信息化建设存在的问题与改进研究[J]. 中国管理信息化, 2020, 23(08): 73-74.
- [2] 李传良. 论新时期职业院校信息化建设中的问题与解决方案[J]. 信息通信, 2020(04): 147-148.
- [3] 马红麟, 徐楠楠, 周文楷. 新时代高职院校信息化建设研究——以北京电子科技职业学院为例[J]. 中国管理信息化, 2020, 23(05): 238-240.
- [4] 关利. 以信息化建设推进职业教育现代化[J]. 职业, 2020(03): 26-27.
- [5] 段瑞静. 高职院校信息化建设有效服务教学科研方法与途径研究[J]. 天津职业院校联合学报, 2020, 22(01): 95-99.
- [6] 李晓旋, 徐康泰, 李江飞, 梅彦利, 王秀梅, 冯学洋. 高职院校信息化建设存在的问题及建议[J]. 承德石油高等专科学校学报, 2019, 21(06): 54-56+79.
- [7] 邵雨萍. 高职院校教育信息化建设的问题及解决策略[J]. 信息与电脑(理论版), 2019, 31(23): 240-241+244.

时,随着互联网大数据、云计算、人工智能等新技术涌现,用人岗位也发生了深刻变化。中小微企业、新业态、新模式对高校就业服务提出了新要求。高校毕业生一方面面临较大就业压力,另一方面要适应经济新常态带来的变化。目前,高校就业服务多以简单的信息推送为主,通过对所有学生无差别的推送用人信息,期望达到较高的就业率。这种传统的“大水漫灌式”的服务模式虽然能覆盖全体毕业生,但忽略了学生的个性化需求,不够精准的就业信息反而阻碍了毕业生高质量就业。

辅导员作为高校就业服务主体,人员有限。面对逐年增加的高校毕业生,如果仅依靠辅导员个人来对学生进行就业分析和推荐,对辅导员的学生工作将产生巨大影响,且难以对数量众多的学生均进行有效和客观的分析,因此本文借助机器学习算法设计模型,来减轻一线辅导员的工作量,并大大提高就业推荐的准确性和效率。

1 算法设计

1.1 神经网络

神经网络^[12]是上个世纪提出的一种智能学习算法,用来拟合数据,发现数据的内在规律。神经网络是机器学习的一种,与其他机器学习算法不同,神经网络不需要人为过多的干预,尽量依靠算法本身去寻找概率统计特征。这类算法有着其独特的特点,当数据量过小时,不能很好的拟合数据,适用于大数据量的场景。

1.2 深度神经网络

深度神经网络^[13-15]的特征提取能力比神经网络更强,更适用与大数据场景。深度神经网络模型的建立过程可以分为两步。第一步将原数据量化后输入输入层,经过深度神经网络的计算,到达输出层。第二步,通过输出结果与正确结果的对比,从而修正模型中的参数。

对于神经网络来说,不同的参数设置,会给模型的准确性带来不同的效果。对于模型参数的选择,目前尚无有效的结论供我们选择,因此,为了进一步保证模型的可靠性,本文通过对比不同的参数设置模型,来提高模型的可靠性。

2 实验设置

2.1 模型架构

本文模型的训练流程图如图 1 所示。首先对原始数据进行预处理,将语义特征转化为数字特征,使其适用于模型的训练。在模型训练之前,还需要对数据进行正则化操作。标准化操作可以使得模型更快的达到拟合。然后数据被分为训练数据和测试数据。首先模型对训练数据进行训练,当误差小于阈值时,即判定模型达到收敛。此时停止训练。进行测试。

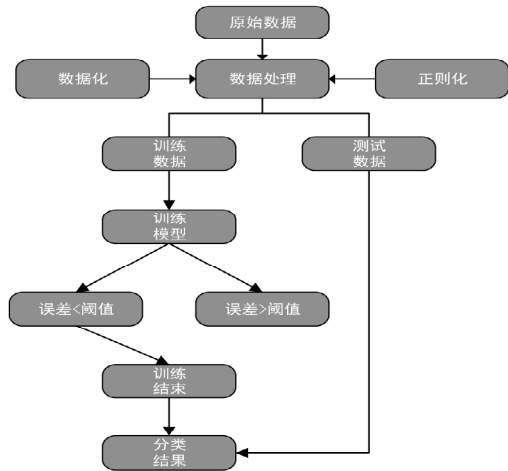


图 1 模型训练流程图

2.2 数据集

目前对于大学生就业情况分析尚无公开的数据集,本文的数据集为本校的学生的真实数据集,共 5000 份。样本特征包括了性别,个人爱好,课程成绩以及就业偏向等共 10 项特征,对应的标签为学生的实际职业。

本文将统计特征分为几大类。性别类分为男女两类;爱好类分为球类,田径类,乐器类,摄影,电竞,美术,歌唱,旅游;成绩类包含了政治成绩,专业类成绩,英语成绩和体育成绩;就业职业偏向包括了程序员,教师,公务员,事业编,个体户,国企,私企,外企;就业地偏向分为超一线城市,郑州市,郑州所属县级市,原籍和其他中心城市;另外,特征还包括了期望薪资。样本数量如表 1 所示:

表 1 样本数据集概括

数据类型	训练样本	测试样本
程序员	711	162
教师	124	42
公务员	235	75
事业编制人员	1321	375
个体户	608	207
国有企业	504	99
私人企业	413	29
外资企业	84	11
共计	4000	1000

2.3 数据预处理

了训练集和测试集的样本数量。关于训练集和测试集的划分,采用程序随机选择的方式。样本中的私企和国企中均不包含程序员,即程序员只单独划分为一类。关于样本中的语义特征,本文采用 one-hot 编码,将语义映射到高维空间,转化为数字特征。为了加速训练,使得训练结果更准确,我们采用了标准化对样本数据进行处理。

2.4 实验结果

本文采用谷歌的 tensorflow 作为深度神经网络框架,编程语言为 python。硬件环境为 Intel Core i5-8400, 16G RAM, 512 SSD, GPU 为 GTX1066。为了便于数据梳理,本文将模型运行在 ubuntu 16.04 系统上。

本文的原始数据集有十维特征,经过数据预处理,可扩展为 28 维特征,因此训练模型的输入神经元个数为 28。本文首先对隐藏层数不同的模型进行实验,并给出实验结果,如表 2 所示。

表 2 模型超参数选择对比

模型序号	隐藏层设置	AC	Tims(s)
1	10-9-8-8-8	94	10.08
2	10-9-8-8	96	7.06
3	10-9-8	93	4.12
4	10-9	85	2.88
5	10	72	0.75

模型序号置 ims(s)-9-8-8-80.08-9-8-8.06-9-8.12-9.88.75.2 给出了隐藏层数不同的五个模型在本数据集下的实验结果。通过表 2 对比不同模型的比较可以发现,当隐藏层数量为 4 的时候模型的结构达到最优。随后我们对同一深度下,不同隐层结构的模型进行训练和测试。

表 3 给出了 6 个模型的训练结果。从模型 3 和模型 5 的对比中我们可以看出,当模型的隐藏层数量相同,神经元的个数相同时,模型的性能表现也会有所差异。我们分析认为,这是由于神经网络中不同层所学习到的特征不同。根据上述分析,在实验的设置上虽然隐藏层同为 4 的模型,但是当神经元个数不同时,模型的训练准确率也不同。基于我们的实验及分析结果,我们选择结构为[10-7-7-10]的模型作为我们的最终模型。

表 3 同一深度不同隐层结构对比

模型序号	隐藏层设置	ACC	Time(s)
1	10-10-10-10	94	10.23
2	10-9-8-8	96	7.06
3	10-7-7-10	97	6.99
4	10-3-3-8	93	4.28
5	7-10-10-7	94	7.02
6	3-7-7-3	89	4.33

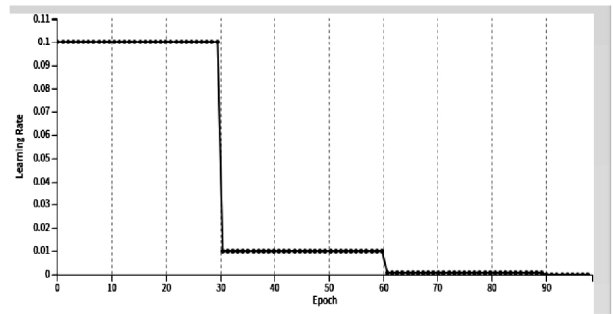


图 4 学习率变化曲线

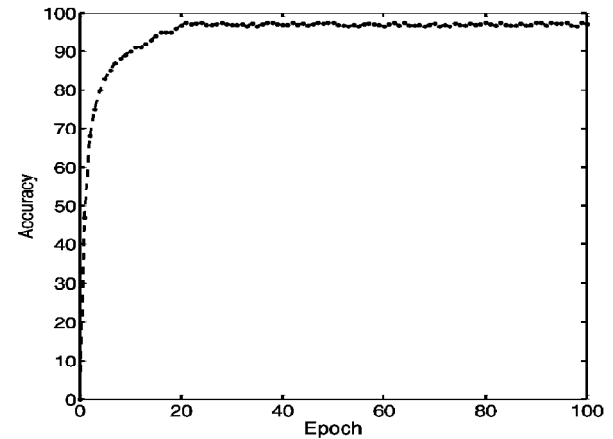


图 5 准确率变化曲线

在迭代次数的选择上,本文选择模型的迭代次数为 100 个 epoch。模型的学习曲线如图 3.5 所示。在图 4.5 中,前 20 个 epoch 迭代时增长最快。在前 10 次迭代上,准确率可以达到 80% 以上。在第 30 次迭代时,基本达到最优。随着迭代次数的增加,模型达到收敛。因此我们在实际的操作中,可以将迭代次数设置在 30-40 之间。

在图 6 中我们给出了模型训练时的损失函数变化曲线。模型的损失与模型的准确率曲线相反,模型的损失在前 20 次迭代时减小最快,在第 30 次迭代时,基本达到最优,随着迭代次数的增加,模型的损失达到平衡。对应于准确率的变化,也符合模型在第 30 次迭代达到收敛。

为了说明本文设计模型的有效性和正确性,本文将数据集用其他的传统机器学习方法也进行了训练分类,选用了常用的

SVM,KNN,和随机森林方法,并与本文的模型进行了对比。

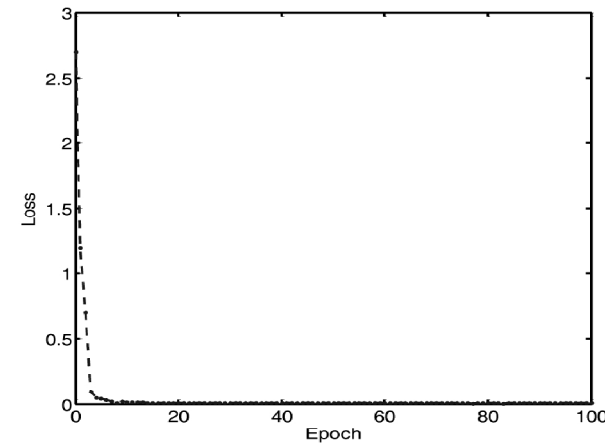


图 6 损失函数变化曲线

本文所用到的对比算法如下:

SVM^[16]: 又称为支持向量机, Support Vector Machines。SVM 是一种常用的分类模型,该模型的学习策略是寻找一个超平面,使得特征空间上的间隔最大化。

RF^[17]: 又称为随机森林, Random Forest。这是一种基于树思想的机器学习方法。随机森林利用多棵树对样本及逆行训练并预测。

KNN^[18]: 又称为 K 最近邻, K-NearestNeighbor。KNN 的分类决策是只依据最邻近的一个或者几个样本的类别来决定样本的所属分类。

表 4 几种算法识别率比较

Method	ACC%
SVM	92
RF	89
KNN	76
OUR	97

其中,本文的方法采用了上述分析的模型结构。表 4 给出了本文的方法和其他机器学习方法的对比实验结果。从表中的数据我们可以看到本文的方法在对数据集的测试准确上明显超过了其他方法。相较于 KNN,我们提出的方法准确率提升了超过 20%。在这些方法中, SVM 的效果较优,也和我们的方法有一定的差距。显然,在针对于本文提出的场景和数据集下,本文提出的深度神经网络模型可以达到更高的准确率。

3 结语

传统的机器学习方法无法有效应对大数据量时的高维特征分类任务,因此本文基于神经网络,设计了一套适用于特定样本集的模型。针对与神经网络参数无法有效确定的缺点,本文进行了有效的,不同的对比实验,从不同角度为模型的参数选择提供了依据。目前针对于大学生就业指导尚无较好的机器算法,仅依靠人力来进行工作给辅导员带来了较大压力,面对日益严峻的就业形势,本文设计了一种基于 tensorflow 的深度神经网络,根据学生性别、兴趣爱好、学习成绩、职业偏向、就业地偏向、期望薪资等特征,为学生提供精准就业推荐,避免现阶段高校采取“大水漫灌”式的就业指导方法,有效缓解了毕业生“慢就业”、“缓就业”的情况,为毕业生与用人单位精准对接提供了依据。

数据库虚拟表技术在用户数据安全保护中的应用

张 晶

(晋中师范高等专科学校,山西 晋中 030600)

摘要:数据库是一种能够进行共享的资源,但是因为数据共享将会导致数据库的安全受到严重的威胁。而数据库的安全性是当前社会所重点关注的问题,数据库的安全性将直接关系到数据的安全。在数据库系统的理论学习以及实际应用方面,为了保护数据表的安全,对于用户数据管理具体需求进行分析,基于数据库字典管理平台,在用户数据安全保护中引用数据库虚拟表技术。

关键词:数据库虚拟表;用户数据安全;保护

中图分类号:TP311.13

文献标识码:A

文章编号:1673-1131(2020)08-0199-02

数据库虚拟表就是数据库视图,由一个或是多个基本表导出形成,但是数据依旧会存在于基本表中,因此和基本表不同,两者之间存在着差异。数据库虚拟表技术在数据库中的作用主要体现为数据库显示数据的窗口,通过数据库虚拟表技术,用户可以查看自己所需要的数据,并且还可以看到数据所出现的变化情况。所以,对于数据库虚拟表技术应用在用户数据安全保护中具有十分积极的意义。

1 用户数据安全保护

国际标准化组织(ISO)对于计算机系统安全的定义为:为数据处理系统建立和采用的技术和管理的安全保护,保护不会受到破坏、泄漏和更改。因为,计算机网络安全可以理解为:通过各种技术与管理措施,保障网络系统的稳定运行,从而确保数据的完整性、保密性以及可用性。因此,构建网络安全保护的目的是为了使用户数据在进行传输、交换过程中,不会出现各种问题。

数据安全有两个方面的含义,并且含义呈现出对立面。首先,是数据本身安全,指的是采用现代密码算法主动保护数据;其次为数据防护的安全,采用现代信息储存手段主动防护数据。数据安全是一种主动的包含措施,数据本身的安全必须基于可靠的加密算法与安全体系,主要是有对称算法与公开密钥密码体系两种。

而数据处理的安全则指的是如何防止数据在使用过程中,因为各种问题而导致数据库损坏,从而出现用户数据安全无

法保障的情况,导致部分敏感数据或保密数据被不具备资格的人员或操作员阅读,从而出现数据泄密,造成严重影响。

而数据存储的安全是指数据库在系统运行之外的可读性。如果数据库被盗,那么即便没有原来的系统程序,也同样可以另外编写程序,从而查看数据库。基于这一方面来说,未加密的数据库不够安全,容易出现数据泄密,因此必须要针对数据防泄密进行研究。而这一问题涉及到了计算机网络通信的保密、软件保护以及安全等问题。

对于静态存储的信息数据来说,其巨大的信息数据量背景下,传统保护方法,是将数据保存在本地本件夹中,然后对于完整性能进行验证。而基于计算机中的动态数据,云计算背景下的多租户特点,可以使用户铁耗服务流程对于数据进行访问。而共享访问的过程载体也成为了用户权利的重点所在。共享漏洞威胁必须要通过权限隔离机制来保护用户权限,而基于云计算下,能够有效降低计算基础设施方面的投入成本,从而使储存数据与计算之间实现合理利用,从而促进用户数据安全保护的发展。

2 数据库虚拟表技术及其应用

根据用户的个性化条件下所提出的不同目标和管理需求,系统管理人员在目标数据库内对于具体对象所对应的个性化数据结构进行定义以及标识,这一步骤就是数据库虚拟表技术。基于数据管理下,管理不同要求所生成的逻辑结构。根据数据库虚拟表的相应数据充足和功能定义的来源等方面的

收稿日期:2020-06-24

作者简介:张晶(1981-),女,山西祁县人,硕士研究生,讲师,研究方向:数据库。

参考文献:

- [1] 苏建贤.大学生职业生涯规划存在的问题及对策[J].亚太教育,000(21):267-268.
- [2] 钟秋明. Review and Prospect of College Graduates' View on Employment Since the Enrollment Expansion%扩招以来高校毕业生就业观研究综述与前瞻 [J]. 大学教育科学, 2015, 000(004):32-36,43.
- [3] 唐博. 转变就业观念,创业走进基层[J]. 中国大学生就业, 423(09):18-20.
- [4] 崔岩. 中国制造 2025 背景下的现代职业教育发展战略[J]. 中国职业技术教育, 2016(30):92-93.
- [5] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436.

- [6] Hinton G, Vinyals O, Dean J. Distilling the Knowledge in a Neural Network[J]. Computer ence, 2015, 14(7):38-39.
- [7] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
- [8] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multi-box detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
- [9] Susan Athey. The Impact of Machine Learning on Economics[J]. Nber Chapters, 2018.
- [10] Cabitza F, Rasoini R, Gensini G F, et al. Unintended Consequences of Machine Learning in Medicine[J]. 2017, 318(6):517.
- [11] 周非, 李阳, 范馨月. 图像分类卷积神经网络的反馈损失计算方法改进[J]. 小型微型计算机系统, 2019, 40(7).