

基于大数据技术的农业决策信息挖掘系统设计

朱小栋, 付峻宇

(上海理工大学管理学院, 上海 200093)
✉zhuxd@usst.edu.cn; fujunyu0319@163.com



摘要: 在乡村振兴战略背景下, 针对农业生产者、消费者及政府政策制定者之间, 存在信息不对称和信息传递滞后性的问题, 以及农业数据的分散化问题, 以农业数据特征和大数据技术特点为研究对象, 提出了一套农业决策信息挖掘系统。首先, 该系统通过数据库设计对多源异构数据进行整合, 打通了不同数据之间的壁垒, 并基于前后端分离的系统开发方案, 为农业生产者和消费者提供决策支持; 其次, 该系统利用长短期记忆网络(Long Short-Term Memory, LSTM)等大数据处理技术, 对序列农业数据进行深度学习趋势预测, 挖掘农业数据的潜在信息, 并设计了在Web平台中融合数据处理技术的具体应用方案。分析结果表明, 该系统能够帮助农业生产者和消费者充分发掘农业数据中的潜在价值。

关键词: 决策信息挖掘; 大数据处理技术; 系统开发

中图分类号: TP311 **文献标识码:** A

Design of Agricultural Decision Information Mining System based on Big Data Technology

ZHU Xiaodong, FU Junyu

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)
✉zhuxd@usst.edu.cn; fujunyu0319@163.com

Abstract: In the context of rural revitalization strategy, this paper proposes to design an agricultural decision-making information mining system for the problems of information asymmetry and information transmission lag among agricultural producers, consumers and government policy makers as well as the decentralization of agricultural data. This research takes agricultural data characteristics and the features of big data technology as the research objects. Firstly, the system integrates multi-source heterogeneous data through database design, breaks the barriers between different data, and provides decision support for agricultural producers and consumers based on the system development scheme of front and back-end separation. Secondly, big data processing technologies such as Long Short-Term Memory (LSTM) is used to predict the deep learning trend of sequential agricultural data, mine the potential information of agricultural data, and design a specific application scheme integrating data processing technology in the Web platform. The analysis results show that the proposed system can help agricultural producers and consumers fully explore the potential value of agricultural data.

Keywords: decision information mining; big data processing technology; system development

1 引言(Introduction)

2021年2月21日发布的中央一号文件《中共中央 国务院关于全面推进乡村振兴加快农业农村现代化的意见》指出, 民族要复兴, 乡村必振兴, 要把全面推进乡村振兴作为实现

中华民族伟大复兴的一项重大任务, 举全党全社会之力加快农业农村现代化^[1]。农业生产是乡村建设的重要方面, 近年来国务院出台的《数字乡村发展战略纲要》, 第一次明确提出了“数字农业”的概念, 旨在通过推动农业的数字化转型来实

现乡村建设的现代化。“数字农业”是数字化发展模式在农业领域的应用实践，是互联网技术与农业生产的有效结合^[2]。

但目前我国实际的农业生产方式在信息化和现代化等方面仍存在许多不足，例如，信息传递的滞后性带来的生产资源分配不合理、信息整合能力不足导致的生产碎片化，等等。以上种种不足导致农业生产者、消费者及政府政策制定者难以做出科学、有效的决策。因此，要推动乡村振兴，实现农业生产现代化，就必须思考如何提高农业生产和销售信息化水平，为农业生产的各个环节提供有力的决策支持。

本文旨在设计一套农业决策信息挖掘系统，通过大数据技术从农业数据中挖掘决策信息。决策信息挖掘系统的构建以优化农业生产为目的和业务逻辑，融合新兴的互联网技术，力求打通农业系统之间的信息壁垒^[3]。为乡村农业生产提供决策信息挖掘系统作为指导工具符合“数字农业”的理念，也是国家推动乡村振兴战略，实现农业现代化的有效方式。

2 农业决策信息挖掘系统构建(Construction of agricultural decision information mining system)

根据农业生产、销售的现实需要，农业决策信息挖掘系统的主要功能包括：(1)对原始数据进行整理，进行可视化处理和信息查询操作，打通信息壁垒；(2)通过深度学习技术，对现有数据进行未来的趋势预测，为农业生产提供一定程度的指导，如图1所示。

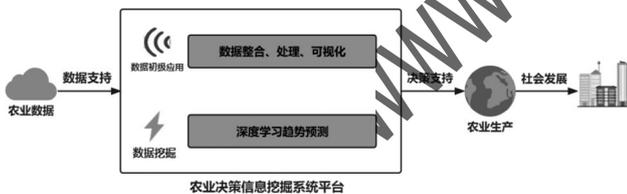


图1 农业决策信息挖掘系统平台功能

Fig.1 Platform function of agricultural decision information mining system

系统的实现过程主要包括：(1)系统架构的搭建；(2)系统数据库的设计；(3)深度学习算法的应用；(4)数据挖掘模型在Web平台中的融合。

2.1 前后端分离的系统架构

首先，系统架构是实现农业决策信息挖掘系统的基础，决定了系统业务的运行方式。传统的系统开发模式包括MVC模式等，但为了灵活应对农业决策信息挖掘系统中复杂多样的数据流，本文采用前后端分离的MVVM模式来构建平台。

MVVM模式在逻辑上将系统分为Model、View、

ViewModel三部分。Model是数据模型，在农业决策信息挖掘系统中代表不同的农业数据类型，用于定义修改数据和操作数据的业务逻辑。View代表农业决策信息挖掘系统界面组件，它将农业数据模型转化成页面的可视化组件，在实际的农业生产场景中提供给用户使用。ViewModel是一个用来同步View和Model的对象，在MVVM架构中，View对象和Model对象之间没有直接的联系，而是通过ViewModel进行交互。除此之外，Model和ViewModel之间的交互是双向的，因此View对象数据的变化会同步到Model对象中，而Model对象数据的变化也会立即反映到View对象上^[4]。在农业决策信息挖掘系统中，农业数据来源多样，在不同的模块中数据的处理不尽相同，所以MVVM模式可以有效地应对数据在前端的更新问题。MVVM架构如图2所示。

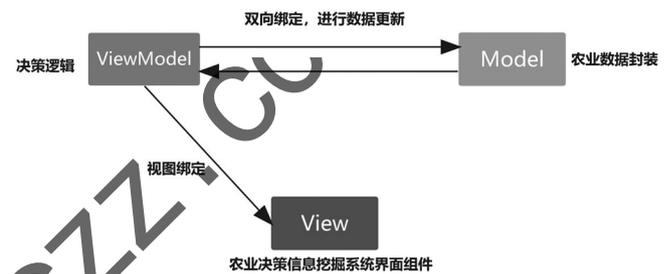


图2 农业决策信息挖掘系统MVVM模式架构图

Fig.2 MVVM model architecture of agricultural decision information mining system

农业决策信息挖掘系统的构建将系统分为前端和后端。前端负责页面内容，后端负责数据供给和各种数据库操作。在MVVM模式下，前后端可以做到完全分离，并且前后端可独立开发和部署^[5]。前后端分离的优势包括：(1)降低开发难度，提高开发效率，更好地实现系统各部分的需求，并且系统架构清晰，有利于后期对系统进行维护和进一步开发。(2)降低系统耦合度，使数据传输和渲染的过程更加清晰，大大提高了系统的健壮性。

另外，平台搭建过程采用的详细技术架构包括：(1)前端使用Vue作为总体框架，通过Axios与后端进行数据交流，利用Bootstrap和Element-UI作为辅助开发用户界面。(2)后端采用SpringBoot框架，使用MyBatis进行数据库读写，在Controller控制层对数据进行封装，生成API接口。

基于以上技术，我们可以完成农业决策信息挖掘系统主要的平台建设工作，并充分考虑其所具备的数据特征和业务逻辑特点。

2.2 农业决策信息挖掘系统数据库设计

各种农业数据是推进“数字农业”的基础，只有在数据的支持下平台才能生成对农业生产有利的服务产品，所以数

据的存储和调用过程至关重要。数据存储技术发展至今，已经由最初的文件存储系统发展到单机的关系型数据库系统，再到近几年兴起的分布式存储系统^[6]。分布式存储系统不再基于单一服务器的存储模式，并且具有更高的扩展性，因此可以更好地满足当下数据量大、业务逻辑复杂的应用场景。分布式系统是与传统RDBMS显著不同的数据库管理系统，其数据的存储可不需要固定的模式，而且通常避免连接操作^[7]。

农业决策信息挖掘系统中农产品种类繁多，数据量大，不同数据的数据结构差异较大，并且系统并发访问时对数据库压力较大，所以适合采用分布式存储的方式，在多台服务器中建立数据库系统，服务端可以根据数据需求访问不同的数据库，大大增强了系统的抗压能力，也可以使数据的存储和数据之间的关系更加清晰。分布式数据库集群如图3所示。同时，服务端在访问数据库时，根据业务逻辑需要对数据库表进行连接，以及进行各种级联操作，所以在单一服务器上仍然适合采用关系型数据库，例如MySQL等。

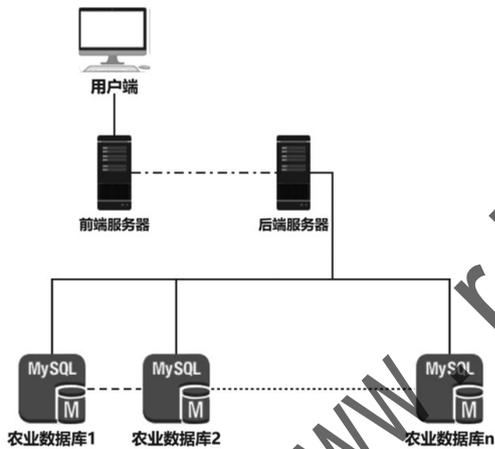


图3 数据库服务集群

Fig.3 Database service cluster

3 大数据算法应用(The application of big data algorithms)

3.1 基于LSTM算法的时间序列预测

上述系统架构及数据库设计是农业决策信息挖掘系统的平台基础及浅层数据应用，为了深度挖掘数据中的重要信息，需要引入有效的数据挖掘算法。LSTM时间序列预测算法在处理价格趋势等方面有着良好的性能，所以适合应用到农业决策信息挖掘系统中。

在农业生产过程中，农产品价格是指导农业生产的重要数据，也是农业生产者关注的重要信息，农业决策信息挖掘系统的数据整理和查询提供了过往价格数据信息，但仍需要对未来一段时间的农产品价格走向有一定的预测功能。

受一些市场及环境因素的影响，农产品价格会随时间波

动变化，而价格波动是影响农业生产者收益的重要因素。所以对农产品市场价格进行短期预测，有利于农业生产者规避市场风险，也是维护市场健康发展、制定农业产业政策的重要手段^[8]。采用基于LSTM的预测算法可以通过数据库中现有价格数据来训练模型，并预测未来一段时间的农产品价格信息。

3.1.1 LSTM时间序列预测算法

LSTM是经典的循环神经网络(RNN)。循环神经网络的前向计算利用了当前步的输入及上一时间步的隐藏状态，当前步的输入表征当前信息，上一时间步的隐藏状态表征过往信息^[9]。

LSTM引入了三种门和记忆细胞。三种门即输入门(Input Gate)、遗忘门(Forget Gate)和输出门(Output Gate)，三种门的输入均为当前时间步输入与上一时间步的隐藏状态，输出由激活函数为Sigmoid函数的全连接层计算得到。LSTM的输入门、遗忘门和输出门可以控制信息的流动。记忆细胞与隐藏状态形状相同，用于记录额外的信息。在LSTM中，长短期记忆的隐藏层输出包括隐藏状态和记忆细胞。其中，只有隐藏状态会传递到输出层。LSTM可用于解决循环神经网络中的梯度衰减问题，并更好地捕捉时间序列中时间步距离较大的依赖关系^[10]。

3.1.2 使用农产品价格数据训练预测模型

针对农业决策信息挖掘系统中的农产品价格数据构造模型并进行训练。构造模型和训练的过程使用PyTorch深度学习框架，其中包含各种神经网络层，可以快速构建LSTM的模型结构。

首先是数据的加载方式。农产品价格信息是一段长时间序列数据，设训练集的总长度为 L ，每次输入的批次大小为 $batch\ size$ ，每一个样本包含的时间步数为 num_steps 。用一个小批量最终时间步的隐藏状态来初始化下一个小批量的隐藏状态，从而使下一个小批量的输出也取决于当前小批量的输入，并如此循环。这对实现循环神经网络造成了两方面影响：一方面，在训练模型时，只需在每一个迭代周期开始时初始化隐藏状态；另一方面，当多个相邻小批量通过传递隐藏状态串联起来时，模型参数的梯度计算将依赖所有串联起来的小批量序列。同一迭代周期中，随着迭代次数的增加，梯度的计算开销会越来越大。为了使模型参数的梯度计算只依赖一次迭代读取的小批量序列，可以在每次读取小批量前将隐藏状态从计算图中分离出来。

其次是定义模型的网络结构。LSTM模型的结构包括输入门、遗忘门和输出门的计算，以及记忆细胞和隐藏状态的

计算，最后通过隐藏状态得到当前时间步的输出。所有的过程可以手动实现，添加一些自定义的设计，也可以直接调用PyTorch提供的LSTM模块，其中包含完整的LSTM计算过程，只需要输入训练数据即可。

最后是定义模型训练函数并进行训练。训练函数中需要指定模型、优化器及各种参数。模型即为定义好的LSTM模型。优化器选择PyTorch中封装好的优化函数，例如SGD随机梯度下降函数、Adam算法等。各种参数包括LSTM中隐藏节点的数量、批次大小、时间步长、学习率和梯度裁剪阈值等。定义好函数的参数后便可以开始训练过程的定义。将加载进来的农产品价格数据输入LSTM模型中，得到一个预测值，并计算这个预测值与真实的农产品价格之间的差距，即计算这一过程的损失值。损失函数使用均方误差损失函数MSE。之后通过损失值对模型中所有节点的参数进行求偏导，并使用优化器进行优化。这便是一次完整的训练过程，对所有输入数据循环进行此过程，最终得到预测模型。

不同农产品价格波动程度存在差异，所以要针对不同数据设置合适的参数大小，并进行多次训练得到尽可能准确的结果。

3.2 基于计算机视觉的农作物病虫害诊断

农作物的虫害、生理病变是影响农作物产量和质量的重要因素，因此对农作物病虫害的诊断与防治是农业生产中的关键活动。但传统基于经验的农作物病虫害诊断具有效率低、准确性差的特点，而且农业生产者学习专业的农业诊断知识困难较大。随着遥感技术的发展，对农作物长势和发展状态的观测手段越来越多，数据样本越来越丰富，这使得计算机视觉在农作物病虫害防治方面有了更好的应用^[11]。其主要的应用方式是对输入图像或视频做目标检测，提取所需关键目标。输入患病农作物所表现的外观图像，利用目标检测技术判断发病类型是大数据技术在农作物病虫害诊断上的有效应用。通过采集到的大量农作物发病表现的图像样本训练模型，来检测农作物图像中存在病理表现的目标，进而将农作物的发病原因提供给农业生产者，进行决策支持。

目前最常用的框架是YOLO模型。YOLO模型自2016年被提出发展至今已有多版本，其中最新的YOLOv5模型的性能最佳，更适合现实应用场景^[12]。

使用YOLOv5训练病虫害诊断模型的步骤分为：(1)数据标签圈定；(2)模型训练；(3)性能测试。整个过程需要依赖足够的农作物病虫害图像样本才能训练出有效的模型。

3.2.1 数据标签圈定

对于原始的农作物病虫害图像样本数据集，需要人工圈

定出图像中的发病表现区域，并设置该病理表现的病虫害类别，才能在训练过程中根据标签进行模型的训练。圈定过程使用的是LabelImg工具，它提供了一个可视化的图像数据标注环境，能够自动对数据进行读取和保存，并将标注结果转换成YOLOv5可以识别的数据格式。标注结果可以划分为训练集、测试集和验证集分别保存，用于后期训练和验证，其标注过程如图4所示。

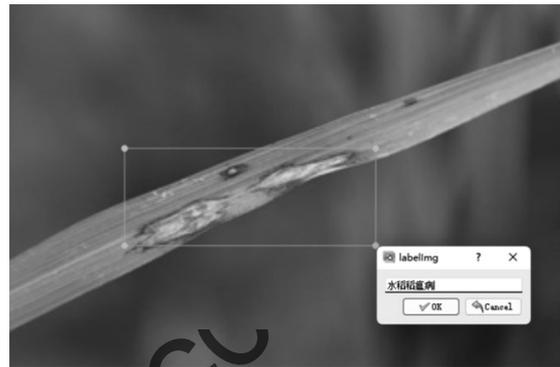


图4 使用LabelImg进行数据标签标注

Fig.4. Data labeling with LabelImg

3.2.2 模型训练

开始进行模型训练之前，要对训练过程进行配置，创建一个标记语言文件来指定训练样本保存位置、分类类别等信息。随后开始训练，通过命令运行YOLOv5模型根目录下的train文件，同时需要设定参数等信息。

3.2.3 性能测试

模型训练后可以得到一个针对某种病虫害的检测模型，同时YOLOv5会输出训练过程中的分类损失值、召回率等信息，用于评价模型的性能，根据这些信息可以调整训练数据集的划分和模型训练超参数的设定，逐步获取性能最佳的模型。

训练好的模型以模型参数文件保存，可以使用YOLOv5模型根目录下的detect文件来指定模型对数据进行检测。不仅如此，YOLOv5训练得到的模型还能用于不同系统或终端的部署，以融合到我们的农业决策信息挖掘系统。YOLOv5模型对样本的检测如图5所示。

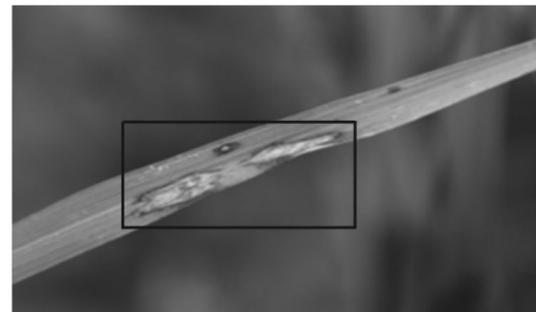


图5 水稻稻瘟病图像样本的检测

Fig.5 Detection of rice blast image samples

3.3 数据挖掘在Web平台中的融合

农业决策信息挖掘系统平台的Web服务端通过基于Java语言的SpringBoot框架构建,但数据挖掘算法模型基于Python语言,平台使用过程中需要通过Web端将输入数据传递到模型来进行预测,所以两者之间的通信存在一定的跨语言冲突,因此需要通过特定手段实现服务端对算法模型的调用。

在Java企业级项目开发中,RunTime.getRuntime().exec()函数提供了调用执行服务器命令脚本的功能,其包含Python脚本在内的多种命令执行功能,因此我们可以使用该方法来创建进程,执行指定的Python程序。该函数的输入参数是一个字符串数组,包含调用程序的命令及指定程序的路径和程序参数,需要注意的是,这里的路径必须为模型算法在服务器中的绝对路径。

算法模型被Java服务调用后生成相应的预测价格信息,价格结果通过Java的BufferedReader类回传到服务端中,再由服务端发送到前端,并渲染到用户所看到的Web界面。

4 农业决策信息挖掘系统的应用(Application of agricultural decision information mining system)

4.1 信息整合与利用

在传统生产和市场环境下,各种农业数据信息分布比较零散,难以联系在一起发挥作用,并且分布不均,农业数据信息掌握在少数部门中,部门之间缺乏有效的沟通渠道,这导致农业生产者、消费者及政府政策制定者之间存在信息不对称及数字鸿沟,妨碍了生产力的提高和供需关系的平衡。农业决策信息挖掘系统将各方面数据整合到一起并进行初步的整理和筛选,以直观、高效的方式提取出有效信息传递给用户。

在这个过程中,农业生产者可以通过农业决策信息挖掘系统了解不同地区及不同产品的价格走势,进而选择在合适的时间进行售卖,以及针对自身的生产能力进行生产在不同作物之间的分配;对于消费者,在大数据农业决策信息挖掘系统中可以了解更多供应渠道及潜在需求,根据自身需要选择更合适的供应商及高品质产品;对于政府政策制定者,可以了解市场中供需关系的状态,更合理地根据实际情况制定有效政策,并提高政策制定的时效性。

4.2 挖掘数据的潜在价值

农业数据体量庞大,其中蕴含着可观的潜在价值,通过现有的人工智能等技术与农业决策信息挖掘系统相结合,可以延伸出多种应用场景^[13]。例如,农业预测预警技术是以农

业物联网收集到的海量数据为依据,利用大数据技术建立预测模型,对农业对象未来发展的趋势进行科学预测,对农业环境中可能出现的不利于农作物正常生长的极端情景进行提前预警,最大限度地消除或减轻有可能遭受的农业灾害。另外,农业人工智能技术可以从优化的角度促进农业生产过程高效化,对农业生产、销售等过程实施动态管理,提升资源利用率和劳动效率,减少生产成本和劳动强度^[14]。

5 结论(Conclusion)

首先,针对目前我国农业生产中存在的信息传递滞后性和信息整合能力不足等问题,以农业数据特征和大数据技术特点为研究对象,提出了一套农业决策信息挖掘系统设计方案。该系统将各渠道收集的农产品数据存入系统数据库,利用LSTM等深度学习模型预测序列数据及图像数据。其次,根据农业数据特征和业务逻辑,设计了农业决策信息挖掘系统的平台架构和数据库架构,优化了运行效率和用户体验,并设计了在Web平台中融合数据处理技术的具体应用方案。最后,通过分析农业决策信息挖掘系统的实际应用场景,证明该系统能够有效挖掘农业数据中的潜在价值,促进农业生产和消费升级,并延伸出其他人工智能算法在农业中的应用。

参考文献(References)

- [1] 单天振.乡村振兴战略体制机制研究[J].农业经济,2021(12):32-33.
- [2] 孙哲君.论乡村振兴视域下“数字农业”的创新实践研究[J].中国市场,2021(31):60-61.
- [3] 朱小栋,王晨悦,陈美芳,等.新时代新农科人才情报库系统的设计与思考[J].软件工程,2021,24(03):24-27.
- [4] 霍福华,韩慧.基于SpringBoot微服务架构下前后端分离的MVVM模型[J].电子技术与软件工程,2022(01):73-76.
- [5] SONG J H, ZHANG M, XIE H. Design and implementation of a Vue.js-based college teaching system[J]. International Journal of Emerging Technologies in Learning, 2019, 14(13):59-69.
- [6] 王珊,撒师焯.数据库系统概论[M].北京:高等教育出版社,2014:442
- [7] 申红雪,郭延哺.分布式数据库研究新趋势[J].软件,2021,42(10):31-34.
- [8] 王晓蕾,张艳,柳平增,等.基于堆叠式LSTM与熵值法的苹果价格组合预测模型研究[J].中国农机化学报,2021,42(10):157-164.
- [9] 林昕,朱小栋.基于Attention机制的LSTM股价预测模型[J].重庆工商大学学报(自然科学版),2022,39(02):75-82.

- [10] ZHANG C P, LI S, YE M, et al. Learning various length dependence by dual recurrent neural networks[J]. Neurocomputing, 2021(466):1-15.
- [11] 何芳.遥感技术在农作物病虫害预警监测中的应用[J].南方农机,2022,53(04):50-52.
- [12] 郭磊,王邱龙,薛伟,等.基于改进YOLOv5的小目标检测算法[J].电子科技大学学报,2022,51(02):251-258.
- [13] 王艳雨,刘萍.基于云计算与物联网技术的数据挖掘分析[J].

科技创新与应用,2021,11(35):94-97.

- [14] 欧非凡,张超群.农业信息处理技术研究与应用进展[J].中国农学通报,2021,37(20):113-118.

作者简介:

朱小栋(1981-),男,博士,副教授.研究领域:数据挖掘与深度学习,云计算与大数据管理,电子商务.

付峻宇(1999-),男,硕士生.研究领域:神经网络,深度学习,推荐系统,软件工程.

(上接第62页)

表1 数据库表与实体类关系表

Tab.1 Relational table of database table and entity class

| 表 | 实体类 | 说明 |
|-------------|-------------|------|
| Number | Number | 药品编号 |
| Picture | Picture | 图片 |
| Purchase | Purchase | 进价 |
| Sell | Sell | 售价 |
| Medicine | Medicine | 药品名称 |
| Type | Type | 药品类型 |
| EXP | EXP | 保质期 |
| Description | Description | 简单描述 |
| Name | Name | 患者姓名 |
| Time | Time | 挂号时间 |
| Department | Department | 挂号科室 |

6 结论(Conclusion)

本医药系统网站基本实现了从医生开药到患者取药的整个流程,包括录入及查询患者信息、后台数据库管理、药品信息更新、药箱状态信息更新等一系列功能,通过开发医生端、药房端、自提柜端三端一体化的服务平台,实现三端个性化同步数据的需求,极大满足了用户对医院取药系统的需求。目前还有一些地方做得不完美,前端页面布局、色彩分布等方面还需要进一步改善^[12]。

希望在未来可以对本医药系统网站加以优化和改进,以提升用户的使用体验。通过“互联网+”和医药系统相结合,以科技带动生产力,进一步促进中国医疗行业的创新和发展,为打造高品质、高效率、高智能的智慧医院注入力量。

参考文献(References)

- [1] 马国强,王哲.基于JavaWeb的智慧医疗问诊系统设计与实现[J].信息技术与信息化,2017(9):105-106.
- [2] 贾娜.基于Java的医院挂号系统的设计[J].福建电脑,2016,32(11):132-133.
- [3] 王鑫,许智鸿,胡华昌,等.智能外卖自提柜系统的设计与实现[J].电脑知识与技术,2020,16(19):10-13.

- [4] JACKSON C M, CHOW S, LEITCH R A. Toward an understanding of the behavioral intention to use an information system[J]. Decision Sciences, 2010, 28(2):357-389.

- [5] 陈帅,关玉蓉.基于JavaWeb的奖助学金系统设计与实现[J].科技广场,2017(03):190-192.

- [6] 陈信君.基于BS架构的成人高校教务管理软件的设计与实现[D].成都:电子科技大学,2012.

- [7] 周子淳.基于Java的校内社交网站设计与实现[J].电脑知识与技术,2020,16(36):89-91,94.

- [8] 王美华.基于CSS 3.0实现的网页动态效果研究[D].武汉:中南民族大学,2014.

- [9] 陈芳.基于JavaScript与ASP.NET的网站开发技术[J].工程技术研究,2017(11):82-83.

- [10] 喻盛甲,许新华,向启凤,等.“兰台e令”网站数据库设计与功能实现[J].信息技术与信息化,2019(06):105-108.

- [11] 辛力.在线视频分享网站数据库设计[J].科技风,2021(12):100-101.

- [12] 宋雅飞.JavaScript特效在企业网站设计中的应用[J].无线互联科技,2022,19(02):91-92.

作者简介:

宋婧荣(2001-),女,本科生.研究领域:智能系统.

杜佩尧(2000-),女,本科生.研究领域:控制理论与控制工程.

李君恺(2001-),女,本科生.研究领域:控制理论与控制工程.

梁鑫(2001-),男,本科生.研究领域:控制理论与控制工程.

陈玲(2001-),女,本科生.研究领域:控制理论与控制工程.

马帅(2001-),男,本科生.研究领域:控制理论与控制工程.

刘婷(1981-),女,博士,副教授.研究领域:智能信号与信息处理.