

文章编号: 2095-2163(2019)03-0103-04

中图分类号: TP311.5

文献标志码: A

印染需求智能分析系统的设计与实现

唐文辉, 刘国华, 王国栋

(东华大学 计算机科学与技术学院, 上海 201620)

摘要: 随着“工业 4.0”概念的提出和实施以及 C2M 互联网商业模式的兴起,传统纺织行业正面临着向智能化、服务化转型的难题。本文将传统纺织印染行业与网络信息技术相结合,以报价服务为切入点,应用深度学习以及有限状态自动机理论,基于微服务架构,设计了一套印染需求智能分析系统。为纺织企业的印染报价服务提供了一套高效、智能的解决方案。

关键词: 工业 4.0; C2M 模式; 结构化数据提取; 深度学习; 有限状态自动机

Design and implementation of intelligent analysis system for dyeing requirements

TANG Wenhui, LIU Guohua, WANG Guodong

(School of Computer Science and Technology, Donghua University, Shanghai 201620, China)

[Abstract] With the introduction and implementation of the “Industry 4.0” and the rise of the Customer-to-Manufacturer Internet business model, the traditional textile industry is facing the challenge of transforming into being intelligent and service-oriented. This paper combines the traditional textile dyeing industry with network information technology, taking the quotation service as the entry point, applying deep learning and the finite-state machine theory, and designs the Intelligent Analysis System For Dyeing Requirements based on microservice architecture. It provides an efficient and intelligent solution for the dyeing quotation service of textile companies.

[Key words] Industry 4.0; Customer-to-Manufacturer; extraction of structured data; deep learning; finite-state machine

0 引言

随着“工业 4.0”概念的提出与实施,全球制造业开始了新一轮的改革。在这样的时代背景下,越来越多的企业陆续采用 C2M (Customer-to-Manufacturer) 商业模式。印染行业作为中国传统制造业,已然进入了新的发展阶段,其面临的困扰正日益突显,诸如企业运营成本的不断提高,以及客户的个性化、小批量、多品种的新需求,企业的生存壮大也不可避免地会受到冲击与挑战。当前,互联网技术、智能技术的进步,为企业的转型升级带来了有利契机。

客户需求的结构化、规范化,是印染企业营销过程中实现快速准确报价的前提。传统的报价形式普遍存在处理效率低下、信息数据不规范、询价过程记录缺失等问题。所以,提升和保证报价功能的效率及规范性已然成为印染行业在结构转型中亟需解决的重要问题。由此,本文设计开发了印染需求智能分析系统。本系统将深度学习、文本预处理以及有限状态自动机理论等计算机理论与技术与业务实践

相结合,并基于微服务架构,为用户搭建了需求分析服务平台,实现了对用户需求的快速精确获取。

1 相关技术

1.1 Seq2Seq

Seq2Seq 的研发主要是为了解决传统 RNN 结构的输入和输出必须要等长,无法实现翻译功能及聊天机器人等服务。Seq2Seq 是一个 Encoder-Decoder 结构的网络,其输入和输出都是序列,Encoder 中将一个可变长度的信号序列变为固定长度的向量表达,Decoder 将这个固定长度的向量变成可变长度的目标的信号序列。

Seq2Seq 模型的基本思想是通过使用一个循环神经网络读取输入句子,将整个句子的信息压缩到一个固定维度的编码中;再使用另一个循环神经网络读取这个编码,将其解压为目标语言的一个句子。

1.2 有限状态自动机

有限状态自动机是为研究有限内存的计算过程和某些语言类而抽象出的一种计算模型。模型中拥有有限数量的状态,每个状态可以迁移到零个或多个

基金项目: 科技部国家重点研发计划(2017YFB0309800)。

作者简介: 唐文辉(1995-),男,硕士研究生,主要研究方向:计算理论、数据库理论;刘国华(1966-),男,博士,教授,博士生导师,主要研究方向:人工智能、大数据、关系数据库。

收稿日期: 2019-03-02

个状态,输入串决定执行哪个状态的迁移。用形式化的语言表示,有限状态自动机可以被表示为一个5-元组 $(Q, \Sigma, \delta, q_0, F)$,设计构成为:状态的有限集合 Q 、输入符号的有限集合 Σ 、转移函数 $\delta: Q \times \Sigma \rightarrow P(Q)$ 、初始状态 $q_0, q_0 \in Q$ 、接受状态的集合 $F, F \subseteq Q$ 。

1.3 微服务架构

微服务架构是一种架构概念,其主要作用是将功能分解到离散的各个服务当中,从而降低系统的耦合性,并提供更加灵活的服务支持。究其本质而言,就是用一些功能比较明确、业务比较精炼的服务,去解决更大、更实际的问题。

2 系统设计

根据对印染行业用户报价的需求分析,将印染需求智能分析系统分为三大模块,研究得到的功能模块设计如图1所示。在此,关于各模块主要功能可分述如下。

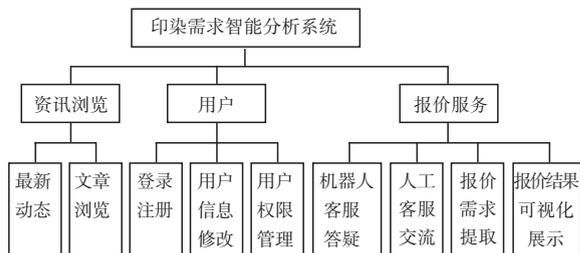


图1 系统功能模块

Fig. 1 Function modules of the system

(1) 资讯浏览模块。该模块分为最新动态和文章浏览,用户可以在最新动态子模块下浏览企业动态、新推出的产品、咨询企业文化等,在文章浏览子模块中,通过技术、咨询、发展等标签将印染领域相关文章做出分类,以随机推荐的形式展示给用户进行阅读。

(2) 用户模块。该模块主要管理用户的基本操作,包括用户登录注册、修改个人信息及密码等,同时对用户进行身份权限管理,对不同身份权限的用户开放不同的功能模块。

(3) 报价模块。通过新技术与传统报价流程相结合,为用户创建智能化的报价体验。可细分为如下4个功能:通过深度学习 Seq2Seq 模型训练的智能机器人客服为用户提供 24 h 印染领域问答服务;基于 WebSocket 搭建的人工客服交流平台,用户可以与人工客服在此平台实时交流,进行产品答疑、需求描述、订单确认等操作;基于有限状态自动机的报价需求提取服务,配合文本预处理等技术,为用户提

供了智能、高效的自动化需求提取流程;最后将提取出的用户结构化需求列表以及客服报价结果整理成表单,以可视化形式反馈给用户。

3 数据库设计

本系统采用 MongoDB 数据库。MongoDB 是一个基于分布式文件存储的开源数据库系统,在高负载的情况下,添加更多的节点,可以保证服务器性能。MongoDB 旨在为 Web 应用提供可扩展的高性能数据存储解决方案。关于重要数据表功能,研究中可做概述如下。

(1) 用户信息表。主要用来记录用户个人信息,包括用户名、密码、手机号、电子邮箱等信息,其中用户名为主键。

(2) 最新动态表。主要用来保存企业相关的最新动态,涉及企业新闻、最新产品、企业文化等内容。包括动态的标题、内容、附图 URI 路径、发布时间等字段。

(3) 文章信息表。主要用来保存印染行业的一些相关文章,供用户浏览。包括文章的标题、内容、发布时间、作者以及文章标签类型等字段。

(4) 报价信息表。主要用来记录每个用户的历史报价信息及报价结果。包括用户编号、报价产品的型号、外观及内在质量要求、执行标准、包装要求等产品需求信息、报价结果、处理客服编号等字段。

4 系统架构与功能实现

4.1 系统架构

系统基于微服务架构,通过前后端分离模式进行开发。与传统的 MVC 开发模式不同,前后端分离模式大大降低了前端和后端的耦合性,把后端根据业务划分成若干个服务组件,前、后端之间通过 API 网络通信进行数据交互,提升了系统的可维护性、开发效率、以及综合性能等。

4.2 智能机器人客服

系统报价模块中,智能机器人客服为用户提供 24 h 印染领域专业问答服务。主要以 2 种方式配合实现,对此可得阐释解析如下。

(1) 通过深度学习 Seq2Seq 模型训练的基于生成的聊天机器人。首先进行数据预处理,目前网络上流行的中文问答语料库在质量上参差不齐,且对话也存在噪音等,需要引入文本预处理,去除一些聊天噪音以及无意义对话等,对处理后的问答数据进行中文分词,并以统一的格式整理成所需的语料库。使用 Keras 初始化 Seq2Seq 模型并设定模型参数。

再将预期准备的语料库送入模型中进行训练。最终实验结果由语料库质量以及训练时间而定。通过该聊天机器人可以实现基础对话的问答交流。

(2)鉴于训练后的机器人并不能保证回答内容的专业性,因而需要预先设定印染专业领域问答库,根据阈值匹配问句,继而反馈对应的答句。研究中,需整理出印染专业领域问答,将问句和答句一一对应保存起来,当智能客服获取用户输入时,就会将用户输入与预设的问句进行相似度匹配,当相似度到达所设的阈值时,即向用户反馈问答库中与该问句对应的答句。从而实现印染领域的专业问答。

4.3 人工客服交流平台

系统除了配备智能机器人客服进行专业问答之外,也为用户提供了人工客服交流平台。该平台基于 WebSocket 进行通信。WebSocket 是一种在单个 TCP 连接上进行全双工通信的协议,可使得客户端和服务端之间的数据交换变得更加简单,允许服务端主动向客户端推送数据。在 WebSocket API 中,浏览器和服务端只需要完成一次握手,两者之间就直接可以创建持久性的连接,并进行双向数据传输。

本系统通过服务器同时与用户端和客服端进行 WebSocket 通信,当用户请求人工客服服务时,会通过调度算法为该用户实时分配当前最空闲的客服,并在服务器中保存该用户和客服的连接,由服务器来处理用户和客服的交流。

4.4 报价需求提取流程

本文通过有限状态自动机作为报价需求提取流程的理论基础,通过模拟真实报价需求提取流程,定义出对应的有限状态自动机,该自动机及其形式化定义如图 2、图 3 所示。

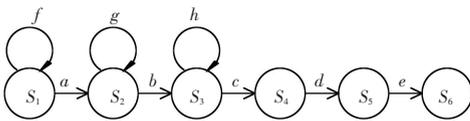


图 2 模拟需求提取流程的有限状态自动机

Fig. 2 The finite-state machine that simulates the requirements extraction process

形式化定义	
$Q=\{S_1, S_2, S_3, S_4, S_5, S_6\}$	状态转移函数
$\Sigma=\{a, b, c, d, e, f, g, h\}$	$\delta(S_1, a)=S_2$ $\delta(S_1, f)=S_1$
$S=S_1$	$\delta(S_2, b)=S_3$ $\delta(S_2, g)=S_2$
$F=\{S_6\}$	$\delta(S_3, c)=S_4$ $\delta(S_3, h)=S_3$
	$\delta(S_4, d)=S_5$
	$\delta(S_5, e)=S_6$

图 3 形式化定义

Fig. 3 Formal definition

在本方法中,用户的合理需求是一个正则语言,需要解决的理论问题为:用户的输入串是否属于这个正则语言。该自动机一开始运行即进入初始状态,等待用户输入串,每次读入输入串就会根据状态转移函数得出下一个状态,直到进入接受状态,则代表准确的结构化需求数据采集完毕,退出采集流程。

4.5 系统实现

印染需求智能分析系统通过前后端分离模式进行开发。其中,前端由 React Native 来操控视图展示,由 Mobx 来管理前端状态;后端基础模块通过基于 Node.js 的 Express 框架进行开发并封装成接口供前端调用;深度学习部分通过 Keras 进行聊天机器人模型的训练。前端视图捕获到用户的输入后会进行对应的逻辑处理,如果需要与后端发生交互,则通过 Ajax 调用 Express 以及 Keras 封装好的 Restful API 进行数据交互,获取到后端返回的数据后,再通过可视化的形式展示给用户。

5 结束语

在新的时代背景下,随着数据量的不断扩大以及移动化联网技术不断走向成熟,新的电子互联网商业模式必将兴起,作为中国的传统制造业,纺织行业也必须随之进行技术改造来推动自身向高端化、智能化、绿色化、服务化的转型。本文通过需求分析、系统设计、数据库设计以及对有限状态自动机理论的研究等,将技术与理论相融合,实现了印染需求智能分析系统。通过本系统,用户可以简洁方便地满足产品报价需求,同时还可以浏览印染行业资讯文章;有了智能机器人客服以及自动化采集报价需求流程,可以大大提升人工客服的工作效率,为用户高效准确地做出报价;用户的历史报价信息也可以被规范统一地加以整理,便于后续查询或分析。为印染行业报价服务提供了一套智能、高效的解决方案。

参考文献

[1] SIPSER M. Introduction to the theory of computation[M]. 3rd ed. USA: Cengage Learning, 2012.

[2] FAMILIAR B. Microservice architecture[M]//Microservices, IoT, and Azure. Berkeley, CA: Apress, 2015: 21-31.

[3] SRIRAM A, JUN H, SATHEESH S, et al. Cold fusion: Training Seq2Seq models together with language models[J]. arXiv preprint arXiv: 1708.06426, 2017.

[4] LIU Tianyu, WANG Kexiang, SHA Lei, et al. table-to-text generation by structure-aware Seq2seq learning[J]. arXiv preprint arXiv: 1711.09724, 2017.