

基于微博的用户相似度计算研究

郑志蕴 贾春园 王振飞 李 钝
(郑州大学信息工程学院 郑州 450001)

摘 要 针对传统的相似度计算方法和评价标准在准确、高效地度量微博用户相似关系时不理想的缺陷,提出一种新的微博用户相似度的计算方法。该方法针对不同的属性数据结构采用不同的计算方式,并根据属性统计信息对各个属性赋值,利用层次分析法确定各属性权值,最后构建综合相似度计算模型。实验结果表明,相对于传统的相似度计算方法,所提计算方法衡量用户相似的准确度提高了22.6%,召回率提高了12.7%,F1度量值提高了29.5%。

关键词 微博,用户相似度,属性权值,层次分析法

中图分类号 TP393 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.02.043

Computing Research of User Similarity Based on Micro-blog

ZHENG Zhi-yun JIA Chun-yuan WANG Zhen-fei LI Dun

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract For traditional similarity calculation methods and evaluation criteria have defects that accurately and efficiently measuring micro-blog users similar relationship is undesirable, a new method to calculate micro-blog user similarity was proposed. For different attribute data structure, different calculation methods are used, and based on the assignment of property statistics for each property, the method uses AHP to determine the weight of each attribute, finally building an integrated similarity calculation model. Experimental results show that the improved method of calculation to measure user similarity precision increases by 22.6%, recall rate increases by 12.7%, and F1 metric improves 29.5%.

Keywords Micro-blog, User similarity, Attribute weight, Analytic hierarchy process

1 引言

微博是一个以用户关系为基础的信息发布、共享、传播、获取的社交网络平台,是对信息传播方式的一种质的革新,对信息化的推广起到了重要的作用^[1-4]。区别于Web1.0的传统网络媒体,在以微博为代表的Web2.0新型媒体中,用户不仅仅是信息的获取者,网站发布的信息也不再是信息的主流来源。在微博中,用户既是基础又是核心,与微博内容共同构成大体框架。与传统的媒体相似,用户可以浏览并获取信息,同时用户也是一个独立的信息节点,主动发布信息,可以通过关注与自身感兴趣的用户频道建立关系网络,接收信息,并通过对微博的点赞、转发与评论推动信息的传播与共享。微博的关注与被关注功能使得用户之间建立起了一个庞大且结构复杂的社会关系,而点赞、转发与评论功能则使得热点信息呈爆发式扩散传播^[5]。

据中国互联网络信息中心(CNNIC)统计,目前新浪微博的注册用户已超5亿,腾讯微博注册用户已达7亿,微博月活跃用户数达到1.76亿,全年净增4700万人。庞大的用户基

数使得在搜寻感兴趣的用户及关注并建立互动关系时,数据过载的问题不容忽视。因此,通过用户自身的信息、用户的关系网络和互动信息来发现潜在的感兴趣的用户变得格外重要。很多在线社区用户推荐系统立足于用户相似度计算,可以满足用户获取感兴趣信息的需求,加强用户之间的互动性^[6]。相似用户的计算与发现在分类、聚类、社区发现、信息的传播、热门微博的预测和舆论的引导等方面都有着重要的意义。

目前,针对微博用户的相似度计算研究仍处在起步阶段,相关的研究和量化标准不够全面。本文将用户自身信息与用户互动信息相结合,提出根据属性的不同取值类型和权值计算用户相似度的模型。

本文第2节介绍相似度计算方法以及相关的研究;第3节分析用户各属性,确定不同属性的相似度计算方法;第4节构建基于层次分析法的属性综合相似度模型;第5节获取微博数据,并对实验结果进行分析;最后进行总结与展望。

到稿日期:2015-12-21 返修日期:2016-05-19 本文受郑州大学新媒体公共传播学科招标课题(XMTGGCBJSZ05),河南省科技攻关项目(144300510007),郑州市科技攻关计划项目(141PPTGG368)资助。

郑志蕴(1962—),女,博士,教授,CCF会员,主要研究方向为分布式计算、智能信息处理;贾春园(1991—),女,硕士生,主要研究方向为社交网络、智能信息处理;王振飞(1973—),男,博士,副教授,主要研究方向为大数据、社交网络;李 钝(1975—),女,博士,讲师,主要研究方向为信息检索、数据挖掘。

2 相似度计算相关研究

传统相似度计算方法是将数据进行转化、归一后统一计算度量^[7],但是描述微博用户的信息不仅包括用户自身的信息,还包括通过关注、被关注、转发、评论、点赞等关联起来的互动信息,且各个数据类型不同,简单地转化数据类型及计算并不能准确、合理地评价和度量用户,在转化过程中还会造成大量的数据缺失。

基于微博的相似度计算研究可以分为 3 大类:1)以 Krishnamurthy 等人^[8]为代表的通过关注与被关注的关系将 Twitter 用户分为 3 类,并以用户关系为基础构建网络拓扑,以计算用户相似度。该算法的核心思想是用户关注关系,没有考虑用户自身的背景信息以及除关注之外的其他用户关系。2)以用户之间的共同邻居数量作为相似度计算标准,即用户 A 和用户 B 之间的共同邻居好友越多,则 A 和 B 之间的相似度越高。CN(Common Neighbors)模型^[9]、Cosine 相似度模型以及 Jaccard 相似度模型、Hub Promoted(HP)相似度模型、HD(Hub Depressed)相似度模型^[10]等属于这类方法,它们的相同点均是以用户间的共同好友数量占自身好友数量的比例作为相似度衡量标准,不同之处在于各种方法对于比例的计算方法不同。但是这些方法并没有考虑到用户自身信息对相似度计算的影响。3)徐志明等人对微博相似度的研究,即将微博看作一个加权无向图,通过分析用户的背景信息、微博文本、社交信息等各项属性对相似度计算的影响力大小进行实验,通过 P@N 和排序准确率评估各项属性在相似度计算上的性能。该算法没有考虑用户的性别、年龄和点赞的信息,也没有给出各属性综合计算的推荐性能^[11]。

本文全面考虑用户的各个基础信息和互动信息,为各属性分配权值,综合计算用户相似度。

3 分类用户属性相似度模型

以新浪微博为例,分析与用户相关的各种属性与信息,根据属性的数据结构确定相应的具体相似度计算方法,根据获取的数据属性的数量、所占比例、属性分布情况等确定各个属性的权值,构建最终的微博用户相似度计算模型。

3.1 微博用户属性分析

微博,主要依靠用户注册时所填写的背景信息以及通过关注、被关注、评论、转发、点赞功能建立的关系网络描述一个用户。用户的背景信息包括性别、年龄、地理位置、工作、教育层次和标签信息^[12]。互动信息包括用户间的关注、粉丝、评论、转发、点赞关系。微博用户特征属性描述如表 1 所列。

表 1 微博用户特征属性

微博用户	背景信息	性别
		年龄
		地理位置
		工作信息
		教育信息
		标签信息
互动信息	关注	关注
		粉丝
		点赞
		转发

3.2 基于背景的相似度计算

假定 U_i 表示微博用户 i ,该用户的性别、年龄、地理位置、关注、粉丝、点赞、转发、评论属性分别用 s, a, p, c, f, l, t, d 表示; U_{ij} 表示用户 i 的属性 j ; w 表示权值,加下标表示相应属性的权值,如 w_s 表示性别属性权值。

(1) 性别

性别往往是衡量一个人的重要标准,性别的不同会对关注的领域和兴趣焦点造成一定的影响,如男性用户一般会对体育、汽车、时政等方面的内容更感兴趣,而女性用户则可能会更关注美容、化妆、饮食、瘦身等方面的信息。性别对于区分用户有一定的影响力。定义用户性别取值如式(1)所示:

$$U_s = \begin{cases} 0, & s = \text{“男性”} \\ 1, & s = \text{“女性”} \end{cases} \tag{1}$$

当两用户的性别相同时,在性别这一维度的相似度为 1,反之相似度为 0,那么用户性别的相似度如式(2)所示:

$$Sim_s(U_x, U_y) = \begin{cases} 0, & U_x \neq U_y \\ w_s, & U_x = U_y \end{cases} \tag{2}$$

(2) 年龄

年龄是一个用户分类的重要因素,不同年龄层的用户往往拥有不同的经历、阅历和关注点。一般而言,年龄差越小,年龄差占年龄的比例越低,用户的兴趣点越接近,相似度越高。但是仅用年龄差并不能准确地描述年龄相似度,年龄差与用户的年龄值之比也是重要的计算因素。如 4 岁的年龄差对于 5 岁和 9 岁的人而言是巨大的,但是对于 60 岁和 64 岁的人,他们之间的年龄相似度则非常高。式(3)给出了用户年龄属性的相似度计算模型。

$$Sim_a(U_x, U_y) = w_a \cdot (1 - \frac{|U_{xa} - U_{ya}|}{\max(U_{xa}, U_{ya})}) \tag{3}$$

其中, w_a 表示年龄属性的权值, $|U_{xa} - U_{ya}|$ 表示用户 x 和 y 之间的年龄差, $\max(U_{xa}, U_{ya})$ 表示两用户之中年龄较大的年龄值。

(3) 地理位置

人们往往更关注自身所在地区发生的新闻,传统习俗、地域文化等方面的区别都会对用户间的相似度造成影响,因此地理位置是在线兴趣推荐中十分重要的属性,也是计算用户相似度必须考虑的因素。本文采用国家-省份-地市三段数据格式来表示用户的地理位置属性,用转换次数来表示两用户位置间的相似度,转换次数是指将某一用户的位置信息变为与另一用户相同时所需变换的次数。转换次数越多,则相似度越低,反之则越高。例如,有 3 个用户, A 用户来自美国-纽约州-纽约, B 用户来自中国-广东-广州, C 用户来自中国-广东-深圳, AB, BC 和 AC 间的转换次数分别为 3, 1 和 3。显然,仅从地域层次考虑,不同的政治、经济、文化背景造成 B 用户与 C 用户之间的相似度大于 A 用户与 B 用户或 C 用户之间的相似度。用户地理位置的权值用 w_p 表示, $T(U_{xp}, U_{yp})$ 表示两用户地理位置属性的转换次数,其相似度计算如式(4)所示:

$$Sim_p(U_{xp}, U_{yp}) = \frac{w_p}{T(U_{xp}, U_{yp})} \tag{4}$$

用户工作行业的不同和教育层次的不同也会造成用户的

兴趣不同。但是由于工作行业的种类数量繁多,而工作性质的分类暂时没有统一的标准和明确的包含关系,并且在新浪微博中填写这两项信息的用户数量极少,因此本文算法不考虑这两项属性对用户分类的影响。

3.3 基于互动信息的相似度计算

在微博中,用户之间的常规互动行为包括关注其他用户以及评论、转发、点赞微博内容等,而关注同一用户、拥有相同的粉丝、对同样的微博内容进行操作等都表现了用户特征和兴趣点的相似。以关注为例,两用户共同关注的人越多,占据关注总数的比例越高,则两用户的关注相似度就越高。其他互动属性的相似度计算方法与关注属性一致。关注相似度的计算如式(5)所示。

$$Sim_c(U_{xc}, U_{yc}) = w_c \cdot \frac{U_{xc} \cap U_{yc}}{U_{xc} \cup U_{yc}} \quad (5)$$

其中, U_i 为考察的用户, S_i 为被考察用户的关注对象, F_i 为被考察用户的粉丝, 用户之间的关注-被关注互动关系如图 1 所示。

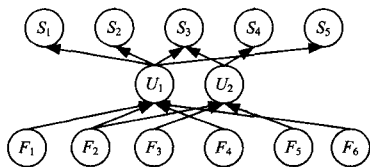


图 1 用户关注-被关注互动关系

在图 1 中, 用户 U_1 关注的人为 $\{S_1, S_2, S_3, S_5\}$, 用户 U_2 关注的人为 $\{S_3, S_4\}$, 两用户全部关注的用户为 $\{S_1, S_2, S_3, S_4, S_5\}$, 共包含 5 个对象; 共同关注的用户为 $\{S_3\}$, 其包含 1 个对象, 那么用户的关注相似度为 $0.2w_c$ 。用户 U_1 的粉丝为 $\{F_1, F_2, F_4, F_6\}$, 用户 U_2 的粉丝为 $\{F_2, F_3, F_4, F_5\}$, 两用户的全部粉丝为 $\{F_1, F_2, F_3, F_4, F_5, F_6\}$, 共包含 6 个对象, 共同粉丝为 $\{F_2, F_4\}$, 其包含 2 个对象, 那么用户的粉丝相似度为 $0.33w_f$ 。

4 基于层次分析法的属性综合相似度模型

微博用户的整体相似度由背景信息的相似度和互动信息的相似度两方面加和构成, 即 $Sim = Sim_b + Sim_i$ 。其中 Sim_b 表示背景信息相似度, 是各项背景属性相似度的加和, $Sim_b = Sim_s + Sim_a + Sim_p$; 而 Sim_i 则表示互动信息相似度, 是各项互动信息相似度的叠加。对结果进行归一化, 使其取值在 $[0, 1]$ 之间, 那么 $\overline{Sim} = \frac{Sim}{\sum w}$ 。综合相似度计算公式如式(6)所示:

$$\overline{Sim} = \frac{Sim_s + Sim_a + Sim_p + Sim_c + Sim_f + Sim_i + Sim_t + Sim_d}{w_s + w_a + w_p + w_c + w_f + w_t + w_i + w_d} \quad (6)$$

描述用户特征的属性很多, 但各个属性对计算相似度的影响力并不相同, 单纯地将所有属性视为同等重要是不合理的, 计算结果的准确性也无法得到保证。根据获取到的数据的统计信息, 存在基于属性取值和层次分析法两种属性权重分配的方法。本文借鉴这两种权重分配方法, 将基于属性取值的方法的结果作为层次分析法的判别依据, 结合两者构成综合权重分配方法, 权重表见 5.3 小节。本文分配的各属性权重适合大多数一般性的相似性计算问题, 特殊问题的计算

应考虑具体的应用场景, 通过筛选合适的实验数据和生成合理的判别矩阵动态调整权重。例如在给一位只关注“美妆”的用户进行好友推荐的问题中, 由于关注这一领域的基本全为女性, 性别属性的权重应降低。

基于属性取值的计算方法以属性的不同取值占属性出现次数的比例作为衡量标准, 即属性的不同取值越多, 该属性在相似度计算时的影响力越大, 权重也就越大^[13]。如微博用户的性别属性和地理位置属性, 性别有两个取值, 而地理位置的取值数量要远多于性别, 那么地理位置对于用户的分类的重要性要大于性别, 相应的权重也更大。属性 i 的权重计算如式(7)所示。

$$w_i = \frac{Num(D_i)}{Num(T_i)} \quad (7)$$

其中, $Num(D_i)$ 表示属性 i 在实验数据集中的不同取值数量, $Num(T_i)$ 表示属性 i 在数据集中出现的总次数。

层次分析法^[14]的核心思想是通过基于属性取值方法的计算结果及观察大量数据的统计结果, 对各属性进行两两比较, 生成判别矩阵, 确定各属性对影响决策所占的比重, 即属性权重。

定义判别矩阵为 A , 它共有 n 个属性, 第 i 行第 j 列的元素为 a_{ij} , 表示元素 i 对元素 j 的重要程度, 取值为 $1 \sim 9$, 1 表示两属性同等重要, 数字越大表示元素 i 相对元素 j 更重要。判别矩阵如式(8)所示:

$$A = (a_{ij})_{n \times n} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad (8)$$

其中, $a_{ji} = \frac{1}{a_{ij}}$, 表示元素 j 对元素 i 的重要程度。

5 实验及结果分析

5.1 实验环境

采用新浪微博、八爪鱼爬虫工具、Eclipse 作为数据的获取、统计以及属性权重、相似度计算工具。实验硬件配置为 Intel(R)Core(TM)i3-2100 @3.10GHz CPU, 4GB 内存。

5.2 实验数据

采用的实验数据主要来自新浪微博。数据获取过程包括采集和筛选两个阶段。在采集阶段, 利用数据采集工具获取用户的基本信息及用户间的互动信息; 在筛选阶段, 去除信息缺失严重的无效数据组, 共得到 2000 个用户的信息, 包括各项属性信息 593909 条。

5.3 属性权重分配

将实验数据代入式(7)可得基于属性取值的权重, 如表 2 所列。

表 2 基于属性取值方法的权重表

属性	性别	年龄	所在地	关注	粉丝	点赞	转发	评论
权重	0.001	0.015	0.353	0.942	0.822	0.978	0.961	0.970

微博用户背景信息判别矩阵为:

	性别	年龄	所在地
性别	1	1/3	1/6
年龄	3	1	1/2
所在地	6	2	1

微博用户互动信息判别矩阵为:

	关注	粉丝	点赞	转发	评论
关注	1	6	6	3	1
粉丝	1/6	1	1	1/2	1/6
点赞	1/6	1	1	1/2	1/6
转发	1/3	2	2	1	1/3
评论	1	6	6	3	1

微博用户背景信息及互动信息间的判别矩阵为:

	背景信息	互动信息
背景信息	1	1/6
互动信息	6	1

采用层次分析法计算各属性权值,归一化后的结果如表 3 所列。

表 3 基于层次分析法的属性权值

属性	性别	年龄	所在地	关注	粉丝	点赞	转发	评论
权值	0.014	0.043	0.086	0.321	0.054	0.054	0.107	0.321

5.4 评估指标

本文采用准确率 (Precision Rate)、召回率 (Recall Rate)^[15]、F1 度量值 (F1-measure) 作为实验结果的评估指标。

以用户所填标签以及用户互动内容作为用户分类的标准答案。使用本文算法计算各个用户之间的相似度,选取不同的阈值,若相似度超过阈值,则将相似用户归为一类。

准确率是指分类结果中正确分类用户个数 N_c 与分类总数 N_t 之比,比值越大,准确率越高。准确率的计算公式为:

$$Precision = \frac{N_c}{N_t} \tag{9}$$

召回率是指标准结果中正确分类用户 N_c 与标准结果总数 N_s 之比,比值越大,分类结果越优。召回率的计算方法如式(10)所示:

$$Recall = \frac{N_c}{N_s} \tag{10}$$

F1 度量值是综合准确率和召回率的评估指标,其值越大,说明分类结果越准确。F1 度量值的计算方法如式(11)所示:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

5.5 实验结果及分析

将非综合的基于用户基础信息的相似度算法、基于用户互动信息的相似度算法、经典的基于共同邻居好友的相似度算法以及本文提出的基于综合属性的相似度算法进行实验对比。当相似度阈值取不同值时,4 种算法的准确率、召回率、F1 度量值分别如图 2—图 4 所示。

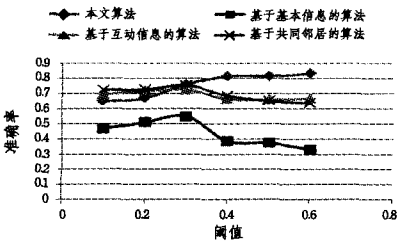


图 2 4 种算法准确率对比

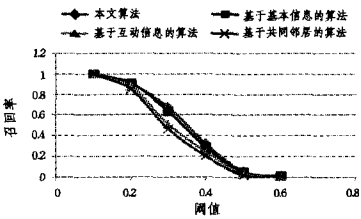


图 3 4 种算法召回率对比

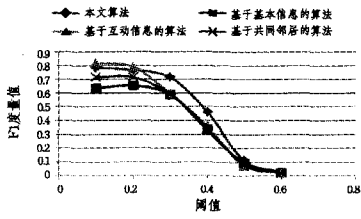


图 4 4 种算法 F1 度量值对比

当阈值大于或等于 0.3 时,本文算法的准确率高于另外 3 种算法,并随着阈值的增加准确率提高,较另外 3 种算法提高了 22.6%;在召回率方面,4 种算法起始值相近,随着阈值的增加,基于综合信息的算法的召回率高于另于 3 种算法,有 12.7%的提升;当阈值小于或等于 0.2 时,基于互动信息的算法的 F1 度量值最高,之后本文算法的结果优于另外 3 种算法,提高了 29.5%。因此,本文提出的基于综合信息的微博用户相似度计算方法的性能优于其他算法。

结束语 针对传统的相似度计算方法无法准确衡量微博用户的瓶颈问题,本文提出结合描述用户的基础信息和互动信息的综合属性相似度计算方法。首先,对微博用户相似度算法的相关研究进行了分类与比较。然后分析了微博用户属性,提出了不同属性的相似度计算方法。最后统计了获取的微博数据,采用层次分析法分配各属性权值,并依据准确率、召回率、F1 度量值将本体算法与仅基于基础属性的算法和仅基于互动属性的算法进行了评估和对比。实验表明,本文所提算法在性能和效果上具有很大优势。该算法仍然具有改进的空间,如用户发布的微博内容、评论内容等也是表明用户特征的属性,对感兴趣用户的推荐有着一定的影响。

参考文献

[1] SHARIFI B. Automatic Microblog Classification and Summarization [D]. Colorado Springs: University of Colorado, 2010.

[2] LIAO Y, MOSHTAGHI M, HAN B. Mining Micro-Blogs: Opportunities and Challenges [C] // Proceedings of Social Networks: Computational Aspects and Mining. Springer, 2011: 1-28.

[3] KAPLAN A M, HAENLEIN M. The Early Bird Catches the News: Nine Things You Should Know about Microblogging [J]. Business Horizons, 2010, 20(10): 1-9.

[4] XIE Y G, XU Y. Special Research: Microblog in China [J]. Modern Communication (Communication University of China), 2011(4): 75-80. (in Chinese)

谢耘耕,徐颖. 微博的历史、现状与发展趋势[J]. 现代传播(中国传媒大学学报), 2011(4): 75-80.

[5] YIN D, HONG L. Structural Link Analysis and Prediction in Microblogs [C] // Proceedings of 20th ACM Conference on Information and Knowledge Management (CIKM2011). Glasgow,

- Scotland, UK, 2011(10):24-28.
- [6] LI Q Q, GU J F. Activity driven modelling of online social network [J]. Journal of Systems Engineering, 2015(1):9-15. (in Chinese)
李倩倩, 顾基发. 用户行为驱动的在线社交网络建模[J]. 系统工程学报, 2015(1):9-15.
- [7] ZHANG B, ZHANG Y, GAO K N, et al. Combining Relation and Content Analysis for Social Tagging Recommendation [J]. Journal of Software, 2012, 23(3):476-488. (in Chinese)
张斌, 张引, 高克宁, 等. 融合关系与内容分析的社会标签推荐[J]. 软件学报, 2012, 23(3):476-488.
- [8] KRISHNAMURTHY B, GILL P, ARLITT M. A New Chirps About Twitter[C]//Proceedings of the 1st Workshop on Online Social Networks. 2008:19-24.
- [9] LU P, ZHANG S S, GAO Q Y. Research on BBV Mode with Limited Node Strength Based on Common Neighbors [J]. Computer Science, 2014, 41(4):49-52. (in Chinese)
逯鹏, 张姗姗, 高庆一. 基于共同邻居的点权有限 BBV 模型研究[J]. 计算机科学, 2014, 41(4):49-52.
- [10] QIN H Y. Research on Correlative Techniques of Hot-topic Discovery About Internet Public Opinion [D]. Habin: Harbin Engineering University, 2010. (in Chinese)
秦宏宇. 网络舆情热点发现相关技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2010.
- [11] XU Z M, LI D, LIU T, et al. Measuring Similarity between Microblog Users and Its Application [J]. Chinese Journal of Computers, 2014, 37(1):207-218. (in Chinese)
徐志明, 李栋, 刘挺, 等. 微博用户的相似性度量及其应用[J]. 计算机学报, 2014, 37(1):207-218.
- [12] WU S F, XU J M, WU X B. Similarity Measurement of Microblogging Users Merging User Tags and Relationships [J]. Journal of Intelligence, 2014(12):170-173, 126. (in Chinese)
吴树芳, 徐建民, 武晓波. 融合用户标签和关系的微博用户相似性度量[J]. 情报杂志, 2014(12):170-173, 126.
- [13] SONG D, HEFLIN J. Domain-independent entity reference in RDF graphs[C]//Proceedings of the 19th ACM International Conference on Information and Knowledge Management. ACM, 2010:1821-1824.
- [14] GUO J Y, ZHANG Z B, SUN Q Y. Study and Applications of Analytic Hierarchy Process [J]. China Safety Science Journal, 2008, 18(5):148-153. (in Chinese)
郭金玉, 张忠彬, 孙庆云. 层次分析法的研究与应用 [J]. 中国安全科学学报, 2008, 18(5):148-153.
- [15] SONG J F, LI G H. Analysis and Improvement on Evaluation of Information Retrieval Algorithm [J]. Journal of Chinese Computer Systems, 2003, 24(10):1800-1804. (in Chinese)
宋峻峰, 李国辉. 信息检索算法评价指标的分析与改进 [J]. 小型微型计算机系统, 2003, 24(10):1800-1804.

(上接第 243 页)

在进行本次测试时, 试验台上压气机的信号经放大后用磁带机记录, 进行同步采样, 采样频率为 5kHz。在服务器端进行信号的转换和滤波处理, 为了减轻传输的负担和客户端识别算法的运算负荷, 进行了重采样。从图 9 可以看出, 第一个开关信号作为失速故障信号的预警信号, 第二个信号是失速故障发生信号, 有 1 秒多钟的时间裕度, 这为进一步的控制提供了一定的预警时间, 有助于保障飞行安全。

结束语 本文利用 LabVIEW 的 GPU 解决了神经网络在线实现的计算问题。首先, 介绍了 RBF 神经网络及其持续激励条件, 这是保证神经网络能够准确逼近系统动态的前提。其次, 基于 LabVIEW 的 GPU 特性进行混合编程, 提出了针对神经网络建模和检测的具体方法, 为在线实现提供了解决途径。通过沿着系统状态轨迹布置神经网络实现了系统的局部准确建模, 利用持续激励水平和神经网络建模准确性的关系, 优化系统参数。最后, 利用 Duffing 振说明了沿着系统轨迹布局神经网络的依据, 并在一台低速航空发动机中实现了基于 LabVIEW 的失速检测系统。在线失速试验说明提出的方法可以保证实时的大规模神经网络计算。

参考文献

- [1] ANDERSON J A. An introduction to neural networks[M]. MIT press, 1995.
- [2] HAGAN M T, DEMUTH H B, BEALA M H. Neural network design[M]. Boston: Pws Pub., 1996.
- [3] ZENG D, GAO L, LIN L, et al. Application of LabVIEW in on-line monitoring and automatic control of fermentation process [J]. Control & Computer, 2006(22):48-50.
- [4] HORNG J H. Hybrid MATLAB and LabVIEW with neural network to implement a SCADA system of AC servo motor[J]. Advances in Engineering Software, 2008, 39(3):149-155.
- [5] LIU J M, XU Z Z, SUN D H. Virtual instrument and neural network in application of grain moisture detection[C]//Proceedings of the 2014 International Conference on Control Engineering and Information Systems. 2014:155-158.
- [6] LU F S, SONG J Q, YIN F K, et al. Survey of CPU/GPU synergistic parallel computing[J]. Computer Science, 2011, 38(3):5-9. (in Chinese)
卢风顺, 宋君强, 银福康, 等. CPU-GPU 协同并行计算研究综述[J]. 计算机科学, 2011, 38(3):5-9.
- [7] GAO Z, CECATI C, DING S X. A survey of fault diagnosis and fault-tolerant techniques-Part I: fault diagnosis With model-based and signal-based approaches[J]. IEEE Transactions on Industrial Electronics, 2015, 62(6):3757-3767.
- [8] WU Y, JIANG B, LU N, et al. ToMFIR-based incipient fault detection and estimation for high-speed rail vehicle suspension system[J]. Journal of the Franklin Institute, 2015, 352(4):1672-1692.
- [9] HWANG W, HUH K. Fault Detection and Estimation for Electromechanical Brake Systems Using Parity Space Approach[J]. Journal of Dynamic Systems, Measurement, and Control, 2015, 137(1):014504.
- [10] KURDILA A J, NARCOWICH F J, WARD J D. Persistency of excitation in identification using radial basis function approximants[J]. SIAM Journal on Control and Optimization, 1995, 33(2):625-642.
- [11] WANG C, HILL D J. Deterministic learning theory for identification, recognition, and control[M]. CRC Press, 2009.
- [12] POLYCARPOU M M, TRUNOV A B. Learning approach to nonlinear fault diagnosis: detectability analysis[J]. IEEE Transactions on Automatic Control, 2000, 45(4):806-812.