

《自动化学报》稿件加工样本

尚书林^{1,2} 左年明² 陈培颖³ 欧彦^{4,5} 张哲^{4,5}

摘要 近些年来,随着社交网络服务蓬勃发展,互联网世界快速进化,社交网络在全球拥有着大量用户,并且使用人数在日益增加,已然成为影响巨大的信息平台。掌握其中有利的信息例如社交网络中用户的行为特征、所具有以及发布的信息所具有的传播规律,甚至内容的定位等,不仅能帮助企业根据客户所需制定出更好更完善及用户友好型的产品,提供更有力的服务,还可以进行更加有效的网络媒体营销。另外,还可为政府及相关部门在舆论控制方面提供有利理论依据和执行的条件,从而优化社会风气。所以,在本科生毕业设计时,我选择了基于社交网络的数据采集以及分析。本文中提出并实现了一种,基于新浪微博程序接口和网络数据流相结合的方式对数据进行采集,利用cookie对新浪微博网页端进行了模拟登陆,实现了数据抓取采集。并且通过监测信息交互时传输的数据包。进行动态抓取页面信息。利用python语言网络爬虫和json等数据库函数,解析HTML XML,获得用户的数据信息。然后通过基于图的聚类的社交网络算法以及PageRank算法,进行用户节点数据的建模。收集用户的偏好以及计算其节点的重要性,对微博中用户与用户之间的社交网络关系进行分析,通过多种不同算法对节点的重要性进行了分析,对比多种算法之间的精准度,对其间的影响力评价模型进行了实现。本文中所讨论的社交网络用户数据信息的获取,主要问题是在于解决了微博近些年来日渐难以爬取大量数据的问题并通过数据挖掘对用户行为进行分析研究,得到了具体的用户影响力公式,并根据实际进行了验证与误差分析。

关键词 社交网络, 数据采集, 数据处理与分析, 数据挖掘, 图聚类

Preparation of Papers for Acta Automatica Sinica

SHANG Shu-Lin^{1,2} ZUO Nian-Ming² CHEN Pei-Ying³ OU Yan^{4,5} ZHANG Zhe^{4,5}

Abstract In recent years, with the rapid development of social networking services and the rapid evolution of the Internet world, social networks have a large number of users in the world, and the number of users is increasing. It has become a huge information platform. Mastering the beneficial information such as the behavioral characteristics of the users in the social network, the propagation rules of the information they have and the information they publish, and even the positioning of the content can not only help the company develop better and more complete and user-friendly The products provide more powerful services and more effective online media marketing. In addition, it can also provide favorable theoretical basis and conditions for the implementation of public opinion control for the government and related departments, so as to optimize social atmosphere. So, when I graduated from undergraduates, I chose to collect and analyze data based on social networks. This article proposes and implements a collection of documents based on the combination of Sina microblogging program interface and network data flow, and uses cookies to simulate the landing of Sina Weibo's web-site to achieve data capture and acquisition. And by monitoring the data packets transmitted during information exchange. Dynamically crawl page information. Using python language web crawlers and json database functions, parsing HTML XML to obtain user data information. Then the user node data is modeled by graph-based clustering social network algorithm and PageRank algorithm. Collect user's preferences and calculate the importance of their nodes. Analyze the social network relationships between users and users in Weibo. Analyze the importance of nodes through various algorithms and compare the accuracy of various algorithms. Degree, the influence evaluation model between them was implemented. The main problem of obtaining social network user data information discussed in this paper is that it solves the problem that microblogging is increasingly difficult to crawl large amounts of data in recent years and analyzes user behavior through data mining and obtains specific user influence. The force formula is based on the actual verification and error analysis.

Key words Social Network Data, Acquisition Data, Processing and Analysis, Data Mining, Graph Clustering

1 准备稿件

在正文开始之前,需要定义\cntitle来定义中文标题,用\cnauthor来定义中文作者,用\cnaddress来定义中文地址,用\email来定义\email地址,用\cnabstract来定义中文摘要,用\cnkeyword来定义中文关键词; 同样,\entitle来定义中文标题,用\enauthor来定

义英文作者,用\enaddress来定义英文地址,用\enabstract来定义英文摘要,用\enkeyword来定义英文关键词。最后,用\maketitle来输出这些定义。

稿件首页应包括下列内容:中英文标题、作者姓名、详细工作单位或通信地址和邮政编码(置于脚注位置)(新投稿件请删除全部作者信息)、摘要、关键词(3~5个)。如该稿件不是原始性稿件(如在某会议

上发表过),请作者务必在第一页用脚注注明. 获基金资助的课题请在首页用中英文脚注说明,基金项目英文名称请写全称而非缩写.

1.1 关于 `section` 标题的写法

对于`section/subsection/subsubsection`标题中的中文,我们使用黑体;当需要夹杂英文的时候,英文使用粗体. 要生成当前`subsection`标题,可以这样写

```
\subsection{关于~{\bf section}~标题的写法}
```

1.2 中文支持及中英文混排

在您的 \LaTeX 系统已经安装了 \LaTeX 中文字体的情况下,此模板支持6种中文字体,分别为宋体(song),仿宋体(fs),楷体(kai),黑体(hei),隶书(li) dsad dsad dsa dsa

和幼圆(you),括号内为字体引用名称,具体可以参考aas.cls 文件第64~69行. 当您编译.tex文件遇到类似以下的字体缺失警告时:

LaTeX Font Warning: Some font shapes were not available, defaults substituted.

请依次检查: 1) 您的L^AT_EX系统是否已经安装了您使用的字体; 2) 字体的引用名称是否正确,比如仿宋体的名称为fangsong,而不是fs, 此时需要相应的修改aas.cls 文件第64~69行.

1.3 pdf 的生成方法

编译生成顺序如下: 1) 运行CCT & Latex命令两次生成dvi和ps文件; 2) 使用dvi2pdf编译dvi文件生成pdf文件. 注意,一般不要使用的其他方法生成pdf文件.

2 录用稿的格式及要求

稿件要求论点明确,论证充分,语句通顺,文字简练,字迹工整. 录用后的稿件请根据录用通知中审稿意见认真修改原稿. 定稿时论文和综述版面不超过7页;短文不超过4页;长论文版面可稍放宽. 凡版面超出过多、文字不流畅、编排混乱的稿件不给予发表.

2.1 符号、计量单位和缩略词

全文同一名词术语、人名、地名须前后一致. 每个字母,符号所表示的物理意义须前后一致,并采用习惯表示法. 计量单位请采用国际单位制(SI),用规范符号表示,如压力单位应用P(帕斯卡),而不用“毫米汞注”. 文中中英文缩写词须在首次出现时注明全称. 文中统一使用英文逗号、句号,英文表述部分逗号、句号后面加一个空格,中文表述部分逗号、句号后面不必加空格.

2.2 数学公式

公式请用阿拉伯数字全文统一编号,外文字母大小写及文种易混淆的,如C、K、O、P、S、V、W、X、Y、Z、 α 和 α 、 r 和 γ 等请在打印稿上用铅笔标注“英大、英小、希大、希小”,上下角标位置明显,书写准确. 一般变量排斜体,专有名词及数学符号如微分、积分、偏微分符号、数学期望、转置等排正体. 向量(矢量)请排小写粗斜体,在打印稿上另用铅笔下划曲线表示,矩阵请排大写斜体. 大写希腊字母如 Γ Δ Θ Λ Ξ Π Σ Υ Φ Ψ Ω 排正体,为向量(矢量)时排粗正体. et al., etc., e.g., s.t.排正体.

式(1)、(2)是两个简单单栏数学公式的例子.

$$\begin{pmatrix} A_{cl}^T P + P A_{cl} & P B_1 \\ B_1^T P & -I \end{pmatrix} < 0 \quad (1)$$

$$g_i(t) = i \int_t^\infty e^{(A-SP_1)^T(r-t)} [P_1 A_1 x^{(i-1)}(r-\tau) + A_1^T P_1 x^{(i-1)}(r+\tau) + A_1^T g_{i-1}(r+\tau)] dr \quad i = 1, 2, \dots \quad (2)$$

式(3)和(4)是通栏的公式的例子. 可以用环境`\end{multicols} \begin{multicols}{2}`产生.

2.3 图表

文中图表只附最必要的,并用中英文双语表示. 图、表分别用阿拉伯数字全文连续编号. 曲线标明坐标. 图表中向量和正文一样在打印稿中用铅笔标注. 附图直接列在正文出现处,不要附在文末. 排版时图片格式应统一为.eps. 如果是仿真图.fig 格式,可直接在Matlab中选择打印成PDF文件,这样可以保留中文文字,然后,由Photoshop 转换为.eps 格式. 图片模式为灰度图,分辨率为600 像素/英寸, 1)尺寸:一般单栏图宽8 cm, 通栏图宽16 cm. 2)字体:图中标注文字英文用Times New Roman字体,中文用宋体,消除锯齿的方法为平滑,字号选8 点. 3)线条:图中实线粗细1 px以上,虚线、点画线3 px以上. 4)存储格式: eps 存储选项为: 预览无, 编码ASCII, 其他不选. 这些图片包括原图(.fig, .vsd, .doc 等) 随文章定稿标识清楚后压缩打包(.zip、.rar), E-mail 至学报编辑部. 以下是图的生成方式

```
\begin{center}
{\centering
\vbox{\centerline{\psfig{figure=1.eps,
width=6cm}}\vskip2mm
{\small 图1\quad
熔化率与比例、积分系数之间的关系\\
Fig.\,1\quad Comparing melting rate with
proportion and integral parameters of
the Consarc controller }}}
\end{center}
```

《自动化学报》的表格采用的是三线表格式. 表1是一个通栏的表格示例,用`\end{multicols} \begin{multicols}{2}`环境.

2.4 参考文献

参考文献只列公开出版的文献. 内部资料、未公开发行的论文集(学位论文除外)等不得作为文献

引用. 必须引用时,可放于出现页的脚注(书写格式与参考文献相同). 参考文献要按在文中出现的先后依次编号1,2,⋯. 所引文献在正文中务请标注. 参考文献中不使用任何缩略语. 作者人数超过六人以上可用“et al.”,否则请注明全部作者姓名. 注意文献条目中各项内容应该齐备,否则将返回作者重新进行查证标引,未被引用的文献不能列于参考文献列表中. 格式如本模板文末所示.

图1 熔化率与比例、积分系数之间的关系
Fig.1 Comparing melting rate with proportion and integral parameters of the Consarc controller

\TeX 的参考文献的引用比较方便,比如我们想引用Knuth的The \TeX book,可以输入“The \TeX book\$^{[1]}\$”得到想要的效果“The \TeX book^[1]”.再比如我们输入“多模态优化问题\,!\\$^{[2]}\$”得到想要的效果“多模态优化问

题^[2]”.

2.5 作者简介注意事项

文章末页需附作者中英文简介,简介不超过100字,包括作者姓名、职称、学习工作经历,以及研究方向.

作者照片请选用简单背景的证件照片,比例要求是5:4,大小做成3.0 cm × 2.4 cm.

在无照片情况下,采用biographynophoto环境来填写作者简历. 与正常情况的区别是,此时没有了照片文件这个参数.

3 其他注意事项

1) 因本刊文章为国内外公开发表,所有录用后稿件必须在本单位经保密审查. 修改加工后的定稿除根据录用通知要求发送一份至编辑部指定电子邮箱外,另请一式两份,注明原稿件编号后附本单位同意公开发表的介绍信寄至编辑部. 如无特殊原因,录用通知发出半年后才收到的修改稿,本刊不再发表.

2) 请作者仔细地按照录用通知和该模版中的各项要求认真加工修改稿,若修改稿不符合要求,编辑部将退回作者重新修改直至合乎要求为止,由此造成的稿件不能及时发表的责任由作者自负. 作者如无特殊声明,本刊默认第一作者为通信作者. 如果第一作者非文章通信作者或通信作者联系方式变动,务请及时通知编辑部.

$$T_{N-i}(m_{N-i}) = \left[\underbrace{O, \dots, O}_{Z_1^{N-i}(m_{N-i})} \underbrace{I \cdot h^{N-i}(n^{N-i}), I \cdot h^{N-i}(n^{N-i} - 1), \dots, I \cdot h^{N-i}(m^{N-i})}_{L^{(N-i)}} \underbrace{O, \dots, O}_{Z_2^{N-i}(m_{N-i})} \right] \quad (3)$$

$$\begin{bmatrix} \hat{\mathbf{x}}(k+1) \\ \mathbf{e}(k+1) \end{bmatrix} = \begin{bmatrix} A_0 + B_0F + L_o(C(k) - C_0) & L_oC(k) \\ (A(k) - A_0) + (B(k) - B_0)F - L_o(C(k) - C_0) & A(k) - L_oC(k) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}(k) \\ \mathbf{e}(k) \end{bmatrix} \quad (4)$$

表1 这是一个通栏的表格示例
Table 1 An exampletable in one column

对手(OPP)	比分(CSU : OPP)	团队决策所占比例	个体决策所占比例	团队决策与个体决策比值
Cyberoos2001	3:0	71.5	28.5	2.51
FCPortugal2001	1:0	68.4	31.6	2.16
Gemini	26:0	59.7	40.3	1.48
Harmony	3:0	69.9	30.1	2.32
Lazarus	11:0	57.3	42.7	1.34
MRB	2:0	63.2	32.8	1.93
SBCe	4:1	65.8	34.2	1.92
UvA_Trilearn_2001	1:0	54.9	44.1	1.24
UTUtd	10:0	70.7	29.3	2.41
WrightEagle2001	3:1	66.2	33.8	1.96
Average		64.8	35.2	1.84

(a)控制参数1
(a) Controller parameter 1

(b)控制参数2
(b) Controller parameter 2

图2 控制器参数对刀具运动轨迹的影响

Fig. 2 The influences of the controller parameters on the tracking errors

3) 稿件经编辑加工和排版后,编辑部将把pdf校样通过E-mail发送给通信作者,请务必逐字逐句核对并在三天内返回校样。除编辑和排版错误外请勿对校样内容作进一步的修改。校样内容的修改将导致因需要请同行专家对稿件作进一步评议而延误文章发表,因此而导致的后果和发生的费用将由作者承担。

4) 在修改稿加工格式方面若有其他不明事宜,请来函来电垂询。感谢合作!

4 结论

文章的结论请勿与摘要、前言等重复。

致谢

文章作者可以在此处对支持和帮助其研究工作或本文写作的其他组织或人员表示感谢。注意,支持此研究的基金项目请直接标注在首页页脚处,不用在此再次致谢。

References

- 1 Ran B, Boyce D E. *Modeling Dynamic Transportation Network*. Berlin: Springer-Verlag, 1996. 69–83

- 2 Payton D, Estkowski R, Howard M. Compound behaviors in pheromone robotics. *Robotics and Autonomous Systems*, 2003, **44**(3): 229–240
- 3 Su Lian-Cheng, Zhu Feng. Design of a novel omnidirectional stereo vision system. *Acta Automatica Sinica*, 2006, **32**(1): 67–72
(苏连成, 朱枫. 一种新的全向立体视觉系统的设计. *自动化学报*, 2006, **32**(1): 62–72)
- 4 Roychoudhury R, Bandyopadhyay S, Paul K. A distributed mechanism for topology discovery in ad hoc wireless networks using mobile agents. In: *Proceedings of IEEE First Annual Workshop on Mobile and Ad hoc Networking and Computing*. New York, USA: IEEE, 2000. 145–146
- 5 Hryniewicz O. An evaluation of the reliability of complex systems using shadowed sets and fuzzy lifetime data. *International Journal of Automation and Computing*, to be published
- 6 Zhang W. *Reinforcement Learning for Job-Shop Scheduling* [Ph. D. dissertation], Peking University, 1996
- 7 The Math Works. *Image Processing Toolbox for Use with Matlab: User's Guide* [Online], available: <http://www.mathworks.com>, November 3, 2006
- 8 Reilly R C, Mack J L. *The Self-organization of Spatially Invariant Representations*, Technical Report PDP.CNS.92.5, Department of Psychology, Mellon University, USA, 1993
- 9 Wilkinson J P. *Nonlinear Resonant Circuit*, U.S. Patent 362412, July 1990
- 10 IEEE Criteria for Class IE Electric System, IEEE Standard 308, 1969



尚书林 中国科学院自动化研究所博士研究生。2002年获得北京师范大学信息学院电子系学士学位。主要研究方向为图像与视频压缩技术。

E-mail: aas_editor@ia.ac.cn

(**SHANG Shu-Lin** Ph. D. candidate at the Institute of Automation, Chinese Academy of Sciences. He received his bachelor degree from Beijing Normal University in 2002. His research interest covers image compression and video coding.)



左年明 中国科学院自动化研究所博士研究生。2002年获得山东大学数学学院学士学位。主要研究方向为医学图像处理, CT图像重建。

E-mail: aas_editor@ia.ac.cn

(**ZUO Nian-Ming** Ph. D. candidate at the Institute of Automation, Chinese Academy of Sciences. He received his bachelor degree from Shandong University in 2002. His research interest covers medical CT image reconstruction and medical image processing.)

陈培颖 《国际自动化与计算杂志》编辑部责任编辑.

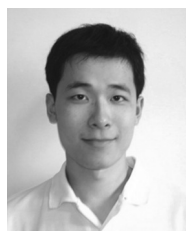
E-mail: peiying.chen@ia.ac.cn

(**CHEN Pei-Ying** Managing editor at the Editorial Office of *International Journal of Automation and Computing*.)

欧 彦 《自动化学报(英文版)》编辑部责任编辑.

E-mail: yan.ou@ia.ac.cn

(**OU Yan** Managing editor at the Editorial Office of *IEEE/CAA Journal of Automatica Sinica (JAS)*.)



张 哲 《国际自动化与计算杂志》与《自动化学报》编辑部技术编辑. 2006年获得北京工业大学电控学院自动化系学士学位. 本文通信作者.

E-mail: zhe.zhang@ia.ac.cn

(**ZHANG Zhe** Copy-editor at the Editorial Office of *International Journal of Automation and Computing*, and

Acta Automatica Sinica. He received his bachelor degree from Beijing University of Technology in 2006. Corresponding author of this paper.)