

文章编号 1004-924X(2016)01-0229-07

应用图学习算法的跨媒体相关模型图像语义标注

李 玲^{1*}, 宋莹玮¹, 杨秀华², 陈逸杰¹

(1. 吉林大学 通信工程学院, 吉林 长春 130012;

2. 吉林大学 网络中心, 吉林 长春 130012)

摘要:针对传统跨媒体相关模型(CMRM)只考虑图像的视觉信息与标注词之间的相关性,忽略标注词之间所具有的语义相关性的问题,本文提出了一种新的基于图学习算法的 CMRM 图像语义标注方法。该方法首先根据运动领域图片训练集中的标注词,建立运动领域本体来标注图像;然后采用传统的 CMRM 标注算法对训练集图像进行第一次标注,获得基于概率模型的图像标注结果;最后,根据本体概念相似度,利用图学习方法对第一次标注结果进行修正,在每幅图像的概率关系表中选择概率最大的 N 个关键词作为最终的标注结果,完成第二次标注。实验结果表明,本文提出的模型的查全率和查准率均高于传统的 CMRM 算法。

关键词:图像分析;图像语义标注;跨媒体相关模型;本体;图学习

中图分类号:TP391 **文献标识码:**A **doi:**10.3788/OPE.20162401.0229

Image semantic annotation of CMRM based on graph learning

LI Ling^{1*}, SONG Ying-wei¹, YANG Xiu-hua², CHEN Yi-jie¹

(1. College of Communication Engineering, Jilin University, Changchun 130012;

2. Network Center, Jilin University, Changchun 130012)

* Corresponding author, E-mail: liling2002@jlu.edu.cn

Abstract: The traditional Crossmedia Relevance Model(CMRM) is based on the relevance between visual information and annotation words, while ignoring the inter-word semantic relevance. Therefore, a new CMRM image semantic annotation model based on a graph learning was proposed. Firstly, the ontology of a sport field was established to label the images of the sport field according the annotation words in an image training set. Then, the traditional CMRM was adopted in the training images to complete the basic image annotations and obtain the image annotation result based on a probability model. Finally, the graph learning was used to refine the basic image annotations based on ontology concept similarity, and the top N keywords in the probability table for each image were chosen as the final annotation results. Experimental results show that the recall and precision of the proposed model are improved as compared with those of the traditional CMRMs.

Key words: image analysis; image semantic annotation; Crossmedia Relevance Model (CMRM); ontology; graph learning

收稿日期:2015-11-12;修订日期:2015-12-07.

基金项目:国家自然科学基金资助项目(No. 61371092);吉林省教育厅“十二五”科学技术研究项目(No. 2014B006)

1 引言

对图像进行语义标注^[1]是图像检索^[2-3]中非常重要且具有挑战性的问题。如何建立底层视觉特征与高层语义之间的关联,减小语义鸿沟^[4],改善图像的标注效果已成为决定图像语义概念的最重要问题。现有的经典图像标注方法主要分为基于分类的图像标注^[5]、基于概率模型的图像标注^[6]以及基于主题的图像标注^[7]。

跨媒体相关模型(Crossmedia Relevance Model, CMRM)^[8]是评价各类图像标注方法的标准方法。该模型采用基于概率模型的标注算法,从概率统计的角度,通过对训练集图像进行学习来获得图像的视觉信息与标注词的联合概率,从而预测测试图像的标注词。

通过对 CMRM 算法的研究发现,该模型只考虑了视觉特征与标注关键词之间的对应关系,忽略了关键词本身所具有的相关性,从而导致存在“语义鸿沟”,在 CMRM 算法中引入语义相关性可提高图像标注性能。

现有的针对 CMRM 的改进多是对采用离散斑点表示图像视觉特征造成的图像信息损失的问题而提出的,文献[9]提出了连续空间相关模型(Continuous-space Relevance Model, CRM)。CRM 直接采用连续特征向量的组合来表示图像视觉信息。然而,由于图像分割过程不准确,导致该模型标注性能不是很理想。文献[10]提出的模型 ICMRM,主要做了两方面的改进:一方面是提出了二值化的词汇概率;另一方面是将 TFIDF 方法引入斑点权重分配方案中,以便体现不同斑点在整个数据集中的重要程度。然而,由于标注词之间的弱相关性导致标注结果准确性不高。

针对上述存在的问题,本文提出了一种基于图学习算法的 CMRM 图像语义标注模型,通过充分考虑标注词之间的相关性信息,利用本体良好的概念层次结构和对逻辑推理的支持,根据运动领域图片的标注词在领域本体中的位置优化标注结果。

本文图像标注的过程分为 3 步:1)对运动领域的图像建立领域本体,用于标注待标注图像;2)

采用 CMRM 图像标注方法对待标注图像进行第一次标注;3)计算待标注图像初始标注词在领域本体中的概念相似度,采用图学习方法充分发掘本体概念之间的关系,以优化标注结果。

2 CMRM 图像标注算法

CMRM 算法^[8]根据斑点和标注词这两个集合之间存在的关联性进行图像标注。其中,训练图像 J 表示为 $J = \{b_1, b_2, \dots, b_m; \omega_1, \omega_2, \dots, \omega_n\}$,斑点 $\{b_1, b_2, \dots, b_m\}$ 代表图像 J 对应的视觉特征的集合, $\{\omega_1, \omega_2, \dots, \omega_n\}$ 代表图像 J 对应的标注词的集合。CMRM 算法计算图像 J 的斑点组 $\{b_1, b_2, \dots, b_m\}$ 和标注词 ω_i 之间的联合概率密度函数如下:

$$P(\omega, b_1, \dots, b_m) = \sum_{J \in \mathcal{T}} P(J) P(\omega, b_1, \dots, b_m | J). \quad (1)$$

CMRM 假设 ω 和 $\{b_1, b_2, \dots, b_m\}$ 这两个事件相互独立且具有相同分布,因此公式(1)可重写为:

$$P(\omega, b_1, \dots, b_m) = \sum_{J \in \mathcal{T}} P(J) P(\omega | J) \prod_{i=1}^m P(b_i | J). \quad (2)$$

其中: $P(\omega | J)$ 和 $P(b_i | J)$ 代表图像的文本信息和视觉信息,CMRM 认为任何一个训练图像的斑点集和标注词集都存在着对应关系,根据平滑最大似然估计方法得到:

$$P(\omega | J) = (1 - \alpha_j) \frac{\#(\omega, J)}{|J|} + \alpha_j \frac{\#(\omega, T)}{|T|}, \quad (3)$$

$$P(b | J) = (1 - \beta_j) \frac{\#(b, J)}{|J|} + \beta_j \frac{\#(b, T)}{|T|}, \quad (4)$$

其中: $\#(\omega, J)$ 代表关键词 ω 在训练图像 J 中标注的次数, $|J|$ 代表训练图像 J 包含的标注词和斑点个数。 $\#(\omega, T)$ 代表关键词 ω 在全部训练集图像中出现的总次数, $|T|$ 代表全部训练集图像中包含的关键词和斑点总个数。同样, $\#(b, J)$ 代表视觉词汇 b 在训练图像 J 中出现的次数, $\#(b, T)$ 代表视觉词汇 b 在全部训练集图像中出现的总次数。Jeon, et al^[8]证明,在 $\alpha_j = 0.1, \beta_j = 0.9$ 时,CMRM 的标注效果最好。

然而传统的 CMRM 算法没有考虑标注词之

间所具有的相关性,因此如何充分利用标注词之间的关系来改善标注效果是本文要解决的问题。

3 基于图学习算法的 CMRM 图像标注

3.1 运动领域本体的构建

本体^[11]是一种共享概念模型的说明。本体可用于规范化和形式化地描述多个领域内的概念及概念之间的关系^[11],因此可用于解决图像语义描述中存在的语义鸿沟问题。在图像标注领域中,采用基于本体的图像标注模型,即通过学习领域本体中概念之间的关系完成图像自动语义学

习,能够解决 CMRM 图像标注中标注词之间语义相关性缺失的问题。

图像对应的领域本体是图像的一种描述方式,任何一幅图像都是其相应领域本体的一个实例,并且可以采用本体中的概念对其进行形式化描述。图 1 是本文采用的运动领域本体概念层次模型。用有向无环图(Directed Acyclic Graph, DAG)来表示,其中节点表示语义概念,概念之间的关系通过节点之间连线表示^[12]。基本的关系有以下 3 种:is-a, instance-of 和 part-of^[12]。其中:is-a 表示概念之间的继承关系;instance-of 表示概念的实例和概念之间的关系;part-of 表示概念之间部分与整体的关系。此外,概念之间还存在同义词、近义词等关系。

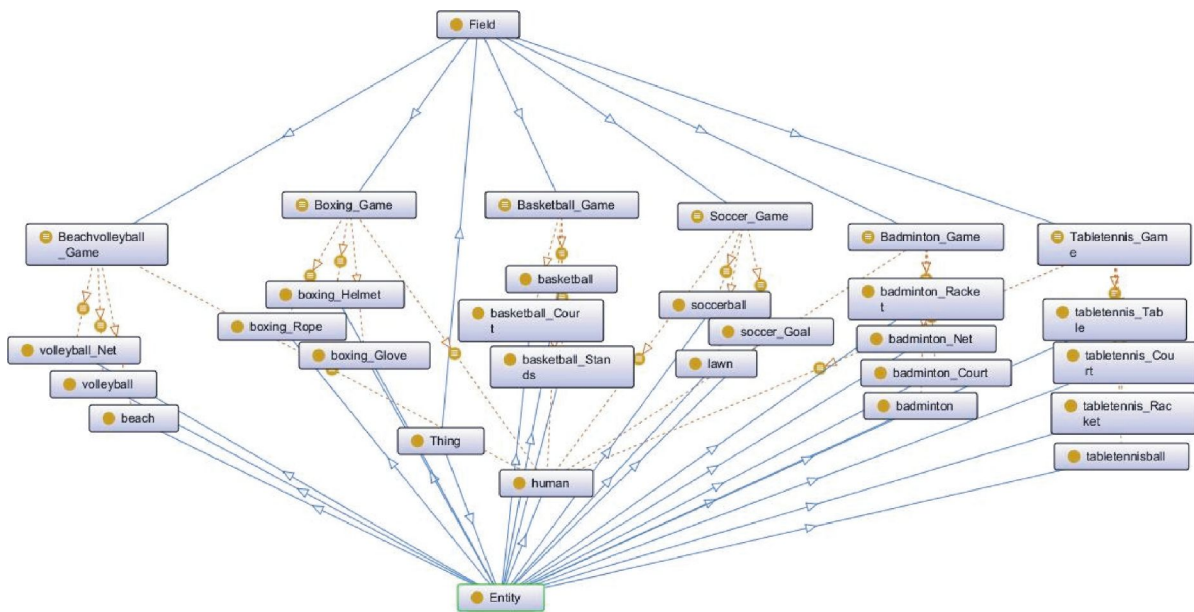


图 1 运动领域本体概念层次模型

Fig. 1 Ontology concept hierarchy model of sports field

3.2 基于图学习算法的 CMRM 图像语义标注

利用标注词之间的语义相关性,采用图学习方法将基于本体的概念相似度引入到 CMRM 算法中,利用本体中标注词间的结构化语义信息计算词汇关系,改善标注结果。

3.2.1 本体概念相似度的图学习算法

该算法以 CMRM 模型得到的图像初始标注词为节点,以标注词在本体中的相似性为边建立

图,通过图学习算法计算已标注关键词之间的关系。同时根据本体中各标注词之间的层次关系,将标注信息从已标注关键词传递到未标注关键词,从而得到最终优化的图像标注结果。

本文的基于本体概念相似度的图学习算法过程如下:

(1)计算标注词在本体中的语义相似度,得到任意两个标注词 ω_i, ω_j 的相似度矩阵 Sim_{ij} 。

$$Sim_{ij} = \begin{cases} \left(1 - \frac{\alpha'}{\text{dep}(R(\omega_i, \omega_j)) + 1}\right) \times \frac{\beta'}{d(\omega_i, \omega_j)} \times \frac{\text{son}(\omega_j)}{\text{son}(\omega_i)}; d(\omega_i, \omega_j) \neq 0, \omega_i, \omega_j \text{ 为同支概念} \\ \left(1 - \frac{\alpha'}{\text{dep}(R(\omega_i, \omega_j)) + 1}\right) \times \frac{\beta'}{d(\omega_i, \omega_j)} \times \frac{\text{son}(\omega_i) + \text{son}(\omega_j)}{\text{son}(C)}; d(\omega_i, \omega_j) \neq 0, \omega_i, \omega_j \text{ 为异支概念} \\ 1; d(\omega_i, \omega_j) = 0 \end{cases} \quad (5)$$

其中: $\text{dep}(R(\omega_i, \omega_j))$ 为 A 和 B 的最近根概念深度, $d(\omega_i, \omega_j)$ 为 ω_i 和 ω_j 之间的距离, 取值均为非负整数, $\text{son}(C)$ 表示本体概念树中以 C 为根的子树的所有节点数。参数 α' 、 β' 用于调整 $\text{dep}(R(\omega_i, \omega_j))$ 和 $d(\omega_i, \omega_j)$ 的权重, 取值为 $(0, 1]$, Sim_{ij} 的取值为 $[0, 1]$ 。图 1 中, beach 与 boxing_Rope 为异支概念, beach 与 volleyball_Net 为同支概念。

(2) 构造矩阵 $D^{-1/2} Sim D^{-1/2}$ 。

D 是对角阵, 对角线元素等于对应 Sim 的行元素之和, 即: $D_{ii} = \sum_j Sim_{ji}$ 。

(3) 算法迭代, 判断收敛性。

根据公式 $T(t+1) = \alpha \times S \times T(t) + (1 - \alpha) \times T(0)$ 进行迭代运算, 直到 $T(t+1) = T(t)$ 则算法收敛, 其中 t 为迭代次数, $\alpha \in [0, 1]$ 为传播系数, $T(0)$ 为初始概率表。

(4) 输出收敛结果 $T^* = (1 - \alpha)(1 - \alpha S)^{-1} T(0)$ 为最终的图像与标注词的概率表。

3.2.2 基于图学习的图像标注算法

结合本体概念相似度的图学习算法, 本文提出基于本体概念相似度的 CMRM 图像语义标注算法的流程图如图 2 所示, 算法的流程如下:

(1) 根据训练集图片得到基本数据信息 $\#(\omega, J)$ 、 $\#(\omega, T)$ 、 $\#(b, J)$ 、 $\#(b, T)$ 、 $|J|$ 、 $|T|$ 。

(2) 计算标注词 ω 的条件概率 $p(\omega|J)$ 和视觉词汇 b 的条件概率 $p(b|J)$, 并计算标注词和视觉词汇的联合概率, 从而得到初始概率表 $T(0)$, 完成图像的第一次标注。

(3) 根据初始概率表 $T(0)$ 与已得到的构造矩阵 S , 进行基于本体概念相似度的图运算, 得到最终的图像与标注词的概率表。

(4) 在每幅图像的概率关系表中选择概率最大的 N 个关键词作为最终的标注结果。

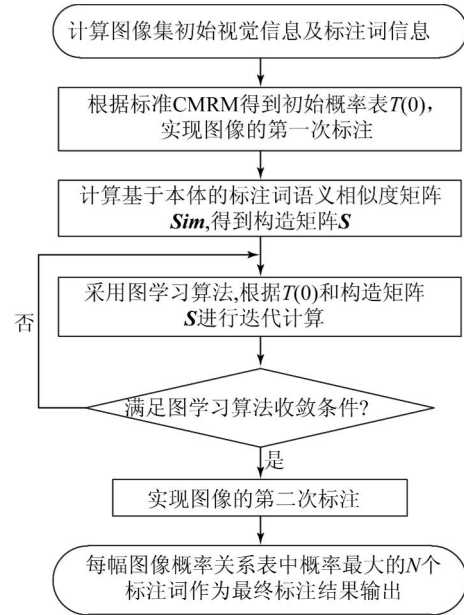


图 2 基于图学习算法的 CMRM 图像语义标注流程图

Fig. 2 Flowchart of CMRM image semantic annotation based on ontology concept similarity

4 实验结果

本文利用奥林匹克官方网站和 Google 图像搜索引擎收集了 3 000 张运动领域图片, 包含以下 6 类体育项目: Boxing_Game, Beachvolleyball_Game, Tabletennis_Game, Badminton_Game, Soccer_Game, Basketball_Game。其中每类图片 500 张, 随机抽取 2 400 张(每类抽取 400 张)作为训练集, 600 张作为测试集。

首先将训练集图像分割成斑点, 对斑点进行特征提取^[13], 采用 k-means 聚类算法^[14]对所有斑点聚类, 由聚类结果得到视觉词元; 然后采用得到的视觉词元的概率信息对测试集图片进行图片的 CMRM 标注; 最后根据图学习算法传播本体概念间相似度, 得到更新的标注结果。实验中参数设置为: $\alpha_j = 0.1$, $\beta_j = 0.9$; $\alpha' = 0.2$, $\beta' = 0.6$; $\alpha = 0.25$ 。





4.1 标注结果示例

部分图像的标注结果示例如表 1 所示。表 1 中显示了针对运动领域图像 CMRM 标注、改进算法 ICMRM 以及本文提出的结合图学习算法的

图像标注的比较结果。由表 1 可知,与传统 CMRM、ICMRM 算法相比,本文算法增加了图像标注的关键词,并且新添加的关键词是合理的。

表 1 部分标注结果示例

Tab. 1 Examples of partial annotation results

测试图像				
人工标注	human, beach, volleyball, volleyball Net	human, pingpong Ball, pingpong Table, pingpong Paddles, pingpong Court	human, badminton, badminton Racket, badminton Net, badminton Court	human, soccerball, lawn
CMRM 标注	human, volleyball, beach, basketball	human, pingpong Paddles, pingpong Ball, boxing Helmet, boxing Glove	human, badminton Racket, soccer Ball, badminton, badminton Court	human, soccerball, boxing Glove
ICMRM 标注	human, volleyball, basketball, lawn	pingpong Paddles, pingpong Ball, badminton Racket, basketball	human, badminton Racket, badminton, boxingHelmet, boxing Glove	human, soccer Ball, lawn
本文算法标注	human, volleyball, beach, volleyball Net	pingpong Paddles, pingpong Ball, pingpong Table, boxing Glove	badminton Racket, badminton, badminton Court, badminton Net	human, soccerball, lawn

4.2 图像标注性能评价

通常用于评价图像标注性能的指标是查准率 (precision) 和查全率 (recall)。

查准率反映图像查询结果的精准度,查全率代表对单个词查询所得查询结果的完整性。查准率和查全率的计算公式如下:

$$precision = \frac{C}{B}, \quad (6)$$

$$recall = \frac{C}{A}. \quad (7)$$

对于给定的检索词 ω ,在手动标注的图片中,包含 ω 的图像数为 A ,在自动标注的测试集图片中,包含 ω 的图像数为 B ,其中参照手动标注数

据,有 C 个是正确的查询结果。

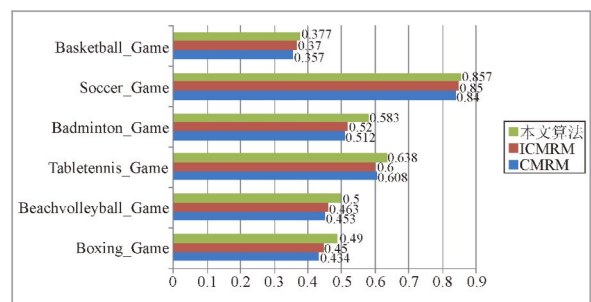


图 3 本文算法与 CMRM、ICMRM 算法查准率对比
Fig. 3 Comparison of precision between CMRM、ICMRM with proposed method

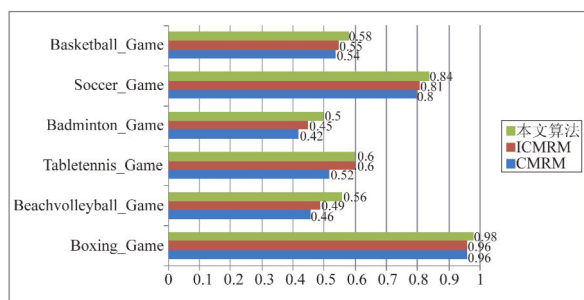


图 4 本文算法与 CMRM、ICMRM 算法查全率对比

Fig. 4 Comparison of recall between CMRM, ICMRM with proposed method

采用 CMRM 以及本文提出的算法对于每一类体育比赛图片得到的平均查全率和平均查准率结果如图 3 和图 4 所示。与 CMRM 算法相比,由于本文采用图学习算法将本体概念间相似度引入到标注结果中,对图像的标注结果进行了改善,在

测试的 6 类体育领域图像中,查准率提高了约 6%,查全率提高了约 10%。

5 结 论

本文提出了一种基于本体概念相似度的 CMRM 图像语义标注方法,该方法解决了 CMRM 算法标注词之间语义信息缺失的问题。本文主要在两方面对 CMRM 算法进行改进:第一方面是根据原始的 CMRM 算法得到初始的标注词,第二方面计算标注词在本体概念模型中的相关关系,利用图学习的方法,将本体概念相似度矩阵引入到初始标注矩阵中,实验结果表明:本文所提方法的查全率和查准率均高于传统的图像标注算法。在今后的研究工作中,需将该模型添加到多种图像标注模型中。

参考文献:

- [1] IM D H, PARK G D. Linked tag: image annotation using semantic relationships between image tags [J]. *Multimedia Tools Appl*, 2015, 74 (7): 2273-2287.
- [2] 吴晓雨,何彦,杨磊,等. 基于改进形状上下文特征的二值图像检索[J]. *光学精密工程*, 2015, 23(1): 302-309.
WU X Y, HE Y, YANG L, et al.. Binary image retrieval based on improved shape context algorithm [J]. *Opt. Precision Eng.*, 2015, 23(1): 302-309. (in Chinese)
- [3] 汪华章,何小海,宰文姣. 基于局部和全局特征融合的图像检索[J]. *光学精密工程*, 2008, 16(6): 1098-1104.
WANG H ZH, HE X H, ZAI W J. Image retrieval based on combining local and global features [J]. *Opt. Precision Eng.*, 2008, 16(6): 1098-1104. (in Chinese)
- [4] WEI W, GAO G L. Image annotation with nearest neighbor based on semantic information [C]. *Proceedings of the 2015 Chinese Intelligent Automation Conference*, 2015, 336: 345-352.
- [5] ZHANG D S, ISLAM M M, LU G J. A review on automatic image annotation techniques [J]. *Pattern Recognition*, 2012, 45(1): 346-362.
- [6] BLEI D M. Probabilistic topic models [J]. *Communications of the ACM*, 2012, 55(4): 77-84.
- [7] 刘杰,杜军平. 基于潜在主题融合的跨媒体图像语义标注[J]. *电子学报*, 2014, 42(5): 987-992.
LIU J, DU J P. Latent topic fusion-based cross-media image semantic annotation [J]. *Acta Electronica Sinica*, 2014, 42(5): 987-992. (in Chinese)
- [8] JEON J, LAVRENKO V, MAMMATHA R. Automatic image annotation and retrieval using cross-media relevance model [C]. *in Proceedings of the Twenty-Sixth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003: 119-126.
- [9] LAVRENKO V, MANMATHA R, JEON J. A model for learning the semantics of pictures [C]. *17th Annual Conference on Neural Information Processing system (NIPS)*, 2003.
- [10] SONG H Y, LI X F, WANG P J. Automatic image annotation based on improved relevance model [C]. *Asia-Pacific Conference on Information Processing*, 2009, 2: 59-62.
- [11] BANNOUR H, HUDELLOT C. Building and using fuzzy multimedia ontologies for semantic image annotation [J]. *Multimedia Tools and Application*, 2014, 72(3): 2107-2141.
- [12] ESPINOZA-MOLINA D, NIKOLAOU C, DUMITRU CO, et al.. Very-high-resolution SAR images and linked open data analytics based on ontologies [J]. *IEEE Journal of Selected Topics*

- in *Applied Earth Observations and Remote Sensing*, 2015, 8(4): 1696-1708.
- [13] 王睿,朱正丹. 融合全局-颜色信息的尺度不变特征变换[J]. *光学精密工程*, 2015, 23(1): 295-301.
WANG R, ZHU ZH D. SIFT matching with color invariant characteristics and global context [J]. *Opt. Precision Eng.*, 2015, 23(1): 295-301. (in Chinese)
- [14] 赵东,赵宏伟,于繁华. 动态多目标优化的运动物体图像分割[J]. *光学精密工程*, 2015, 23(7): 2109-2116.
ZHAO D, ZHAO H W, YU F H. Moving object image segmentation by dynamic multi-objective optimization [J]. *Opt. Precision Eng.*, 2015, 23(7): 2109-2116. (in Chinese)

作者简介:



李 玲(1965—),女,黑龙江齐齐哈尔人,硕士,副教授,硕士生导师,1987年于长春邮电学院获得工学学士学位,1994年于吉林工业大学获得工学硕士学位,主要从事数据挖掘算法研究、分布计算、移动计算和计算机网络协议分析与设计研究。E-mail: liling2002@jlu.edu.cn



宋莹玮(1991—),女,吉林四平人,硕士研究生,2013年于吉林大学获得学士学位,主要从事图像大数据挖掘、语义网研究。E-mail: ruohuhuakai@163.com