

文章编号: 2095-2163(2019)03-0100-03

中图分类号: TP181

文献标志码: A

基于启发式算法的 CP-nets 学习研究

仲兆琳, 信统昌

(烟台大学 计算机与控制工程学院, 山东 烟台 264000)

摘要: CP-nets(条件偏好网)是定性表达偏好关系的一种图形工具,作为一种表达能力的工具,CP-nets 功能强大,能直观、自然地表达用户的偏好信息。但是对于 CP-nets 学习的研究还不够深入,在实际应用中,由于用户行为或者观测误差的随机性,可能导致数据集中存在噪声数据,使得许多传统的学习方法无法得到最优的 CP-nets 结构。本文提出基于启发式算法的学习方法来解决 CP-nets 的结构学习问题。与传统方法中直接学习 CP-nets 结构不同,本文将 CP-nets 的结构学习问题转化为寻找最短路径问题,利用启发式算法的能力来寻找最优的 CP-nets。

关键词: 条件偏好网(CP-nets); 启发式算法; 结构学习

Research on learning conditional preference networks using heuristic algorithm

ZHONG Zhaolin, XIN Tongchang

(School of Computer and Control Engineering, Yantai University, Yantai Shandong 264000, China)

[Abstract] CP-nets (Conditional preference networks) is a graphical tool for qualitative expression of preference relations. As a expressive tool, CP-nets is powerful and can express user preferences intuitively and naturally. However, the study of CP-nets learning has not advanced sufficiently. In practical applications, due to the randomness of user's behavior or the observation error, it may lead to noisy data in data sets, which makes many traditional learning methods fail to get the optimal CP-nets. In this paper, a learning method based on heuristic algorithm is proposed to solve the structural learning problem of CP-nets. Compared with the direct learning CP-nets in traditional method, this paper transforms the structural learning problem of CP-nets into finding the shortest path problem, using heuristic algorithm to find the optimal CP-nets.

[Key words] Conditional preference networks(CP-nets); heuristic algorithm; structural learning

0 引言

传统的商业研究是一个通过人工分析数据来探索和定义的各种因素之间关系的过程,但是,即使有强大的计算机和统计软件,许多隐藏的和潜在的有用信息还是无法被分析师挖掘出来。如今,这样的问题更加尖锐,因为许多企业会在短时间内产生大量的数据,面对数据的爆炸性增长,需要一种有效的方法来挖掘数据库中有用的知识。通过数据挖掘,可以发现各种因素之间的复杂关系。提取出的有用的知识,可以提高决策的效率和质量。

条件偏好网(CP-nets)^[1-4]是一种图模型,能够表示变量之间的关系。CP-nets 的学习问题是通过观察用户的查询记录来提取偏好结构和多个偏好参数。准确地提取和排序这些偏好是困难的,不仅是对于无环 CP-nets^[5-6],对二进制 CP-nets 和可分离 CP-nets 也很困难。这一过程涉及到 CP-nets 的学习方法,例如利用假设检验的方法学习 CP-nets^[7]、基于 CP-nets 诱导图的学习^[8]等。CP-nets 能够发

现决策变量中的定性偏好或者是变量之间的依赖关系,使得决策工具能够快速发展,并且具有高效的决策能力,这使 CP-nets 在决策论^[9-11]中大量使用。

1 问题描述

要得到 CP-nets 准确的图模型是比较困难的,存在 2 个主要问题:维度问题和噪声数据。首先,随着 CP-nets 中的结点数目的增加,CP-nets 的数量呈指数型增加,所以 CP-nets 学习属于 NP-Complete 问题。其次,在处理数据集时经常出现的一个问题是噪音数据。有时收集数据时会存在误差导致噪声数据的产生,从而对学习算法产生影响,因此,有必要对学习算法的鲁棒性进行评估。针对以上问题,本文采用启发式算法来开展研究。在状态空间搜索图中,找到一条最短路径,路径上的结点中包含的变量关系就是最佳的 CP-nets。因此,找到最短路径,就能得到最佳 CP-nets 结构。

2 CP-nets 学习的相关概念

CP-nets 是一个关于顶点 V 的带有有向边的图

作者简介: 仲兆琳(1993-),男,硕士研究生,主要研究方向:人工智能领域的 CP-nets 学习。

收稿日期: 2019-02-27

模型,包括2部分:图 G 和条件偏好表(CPT)。其中,图 G 包括变量集 $V = \{X_1, X_2, \dots, X_n\}$ 和有向边集 CE ,变量之间的有向边代表变量间的依赖关系。每个变量 X_i 对应一个条件偏好表 $CPT(X_i)$,用来表示偏好关系。条件偏好表用来表示变量 X_i 在 $Pare(X_i)$ 的不同取值下,用户对 $Dom(X_i)$ 集合中的偏好。 $Pare(X_i)$ 的取值不同,用户对 $Dom(X_i)$ 集合中的偏好也会发生变化。

用于学习CP-nets的状态空间搜索图是属性域集合的所有子集的哈斯图。具有4个变量的状态空间搜索图如图1所示,在第一层为空集的结点为初始结点,第 $n + 1$ 层包含所有变量的结点为目标结点,其中 n 为变量总数。

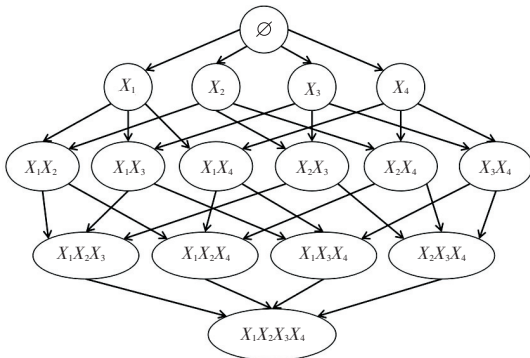


图1 具有4个变量的状态空间搜索图

Fig. 1 State space search graph with four variables

目前,针对解决最短路径问题的研究方法有很多,在本文中,考虑运用启发式算法中的动态规划算法、遗传算法和A*算法来解决这一问题。

3 基于启发式算法学习CP-nets

3.1 基于动态规划方法学习CP-nets

采用动态规划方法进行CP-nets学习,动态规划方法从状态空间图的第一层开始扫描,并对层与层之间的每条路径进行评估,选出上下两层之间的最佳路径,为搜索图中的每个结点包含的变量找到最佳子网。扫描到最后一层结点的时候,一定能找到一条最佳路径对应于一个最佳CP-nets。但是,随着变量数目的增加,算法的运行时间将呈指数型增长,动态规划方法变得不可行。

3.2 基于遗传算法学习CP-nets

采用遗传算法进行CP-nets学习,遗传算法是一种搜索启发式算法,在人工智能领域中,用于解决最优化问题。本文要求解最佳CP-nets结构,因此可以通过遗传算法来解决这一问题。通过进行选择、交叉和变异的一系列遗传操作,找到最短路径,

进而得到最佳CP-nets结构。

遗传算法的核心是遗传操作和适应度函数(打分),遗传操作即选择算子、交叉算子和变异算子。在遗传算法学习CP-nets过程中,产生一个由随机路径组成的初始种群,通过适应度函数得出所有路径的得分。然后进行遗传操作,运用选择算子,选择出其中一部分路径作为父本和母本;运用交叉算子,对选择出来的路径(父本、母本)进行交叉操作,产生一个后代种群,对后代种群中的路径进行适应度计算得到分数;运用变异算子,对后代种群中的路径随机发生变异,并对变异后的路径进行适应度计算得到分数;在初始种群和后代种群中,根据选择算子选择出得分高的“精英”路径组成下一次循环迭代的种群,依次循环,最终得到最优的路径,即最短路径,进而得到最佳CP-nets结构。

3.3 基于A*算法学习CP-nets

采用A*算法进行CP-nets学习,A*算法是一种典型的启发式搜索算法,A*算法运用的思路是:定义一个评价函数 $f, f(n) = g(n) + h(n)$,对当前的搜索状态进行评估,找出一个最有希望的结点来扩展。定义评价函数的参考原则有:一个结点在最佳路径上的可能性;根据格局或状态的特点来打分。

A*算法有一个OPEN集合来存储搜索边界,还有一个CLOSED集合来存储扩展结点。算法运行之初,OPEN集合中只有初始空间结点,CLOSED集合中是空的。在每个搜索步骤中,从OPEN集合中找出最小 f 值的空间结点为 U ,选择该节点用于扩展以生成其后继结点。但是,在扩展 U 之前,需要首先检查该节点是否是目标节点。如果是,说明找到了到达目标的最短路径,可以从此路径构造一个CP-nets并终止搜索;如果 U 不是目标结点,将其扩展生成后继节点后,在每个后继结点 S 中加入一个新的变量 X, S 作为现有的网络变量 U 的一个子结点,即 $S = U \cup \{X\}$ 。检查CLOSED列表中是否已经存在该结点,如果存在的话,将进一步检查该结点是否具有比 S 更小的 g 值,若该结点具有比 S 更小的 g 值,那就删除 S ,因为这时的 S 代表路径不是最优的;若该结点具有的 g 值没有比 S 的更小,那么该结点将会从CLOSED集合中删掉,并且将 S 放入OPEN集合中。如果在CLOSED集合中不存在该结点,接下来直接扫描OPEN集合,若该结点不在OPEN集合中,只需将 S 添加到OPEN集合中;若该结点在OPEN集合中,将该结点的 g 值与 S 的 g 值相比较,如果该结点的 g 值小,将OPEN列表中的 S 删除;如

果 S 的 g 值小,该结点将被 S 替换掉。在生成所有的后继结点之后,将结点 U 放入 CLOSED 集合中,说明该结点已经扩展完成。此时,从起始状态到目标状态的最短路径被找到。从目标结点开始,追踪最短路径直至到达起始结点,从而得到新的 CP-nets 图模型。路径上的每个结点都存储一个变量,这些变量组成了 CP-nets 中的顶点集。

4 结束语

在人工智能的研究中,对偏好学习的研究在实践中具有重要的价值。本文提出学习方法不直接学习 CP-nets,而是将学习问题转化为最短路径发现问题。使用状态空间图表示学习问题的求解空间,利用启发式算法寻找求解空间中的最短路径。在启发式函数的指导下,启发式算法通过搜索求解空间中最优部分来构建最优的 CP-nets。

本文中的学习方法仍有局限性,当顶点数目过多时,求解空间将呈指数倍增大,导致运行时间增加。今后的工作将致力于学习算法的改进,进一步提高学习方法的性能。

参考文献

[1] BOUTILIER C, BRAFMAN R I, DOMSHLAK C, et al. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements[J]. Journal of Artificial Intelligence Research, 2011, 21(1): 135-191.

[2] LIU Jinglei. Research on CP-nets and its expressive power[J]. Acta Automatica Sinica, 2011, 37(3): 290-302.

[3] LIU Jinglei, LIAO Shizhong, Zhang Wei. On the completeness and consistency for CP-nets[J]. Journal of Software, 2012, 23(6): 1531-1541.

[4] GOLDSMITH J, LANG J, TRUSZCZY NSKI M, et al. The computational complexity of dominance and consistency in CP-Nets[J]. Journal of Artificial Intelligence Research, 2008, 33(1): 403-432.

[5] ALANAZI E, MOUHOU M, ZILLES S. The complexity of learning acyclic CP-nets [C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016). Palo Alto, CA, USA: AAAI, 2016: 1361-1367.

[6] CORNELIO C, GRANDI U, GOLDSMITH J, et al. Voting with CP-nets using a probabilistic preference structure [C]// Proceedings of COMSOC. Pittsburgh, Pennsylvania: IEEE, 2014: 1-15.

[7] LIU Juntao, YAO Zhijun, XIONG Yi, et al. Learning conditional preference network from noisy samples using hypothesis testing [J]. Knowledge-Based Systems, 2013, 40(1): 7-16.

[8] LIU Juntao, XIONG Yi, WU Caihua, et al. Learning conditional preference networks from inconsistent examples [J]. IEEE Transactions on Knowledge & Data Engineering, 2013, 26(2): 376-390.

[9] XU Zeshui, ZHAO Na. Information fusion for intuitionistic fuzzy decision making: An overview[J]. Information Fusion, 2015, 28(C): 10-23.

[10] ALLEN T E. CP-nets: From theory to practice [C]// 4th International Conference on Algorithmic Decision Theory. Lexington, KY, USA: Springer International Publishing, 2015: 555-560.

[11] CONITZER V. Making decisions based on the preferences of multiple agents[J]. Communications of the ACM, 2010, 53(3): 84-94.

(上接第99页)

4 结束语

本文提出了角点检测和图论相结合的 ROI 定位方法,在一定程度上降低了背景、尺度、角度对图像分类的影响,使所提取的 SIFT 特征点集中在物体上,则能生成更具有代表性的视觉词汇。同时引入模糊隶属度函数,在图像直方图表示生成时融合空间信息,提高图像的分类精度。在 Caltech-101 数据集上的实验结果表明,本算法较传统词袋模型算法有更高的分类准确度。

参考文献

[1] WANG Chong, HUANG Kaiqi. How to use Bag-of-Words model better for image classification[J]. Image and Vision Computing, 2015, 38: 65-74.

[2] LI Feifei, FERGUS R, PERONA P. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories[J]. Computer Vision and

Image Understanding, 2007, 106(1): 59-70.

[3] GREGORY G, ALEX H, PIETRO P. Caltech-256 object category dataset [D]. California: California Institute of Technology, 2007.

[4] EVERINGHAM M, Van GOOL L, WILLIAMS C K, et al. The pascal visual object classes (voc) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[5] WANG Junqiu, CIPOLLA R, ZHA Hongbin. Vision-based global localization using a visual vocabulary [C] // Proceedings of the 2015 IEEE International Conference on Robotics and Automation. Barcelona: IEEE, 2015: 4230-4235.

[6] PHILBIN J, CHUM O, ISARD M, et al. Lost in quantization: Improving particular object retrieval in large scale image databases [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2008(CVPR 2008). Anchorage, AK: IEEE, 2008: 1-8.

[7] SHI J, TOMASI C. Good feature to track [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle: IEEE Press, 1994: 593-600.

[8] JIANG Wen, GUO Fei, LIU Zheng. A graph theory method for determination of cryo-EM image focuses[J]. Journal of Structural Biology, 2012, 180(2): 343-351.

[9] KUMAR M, STOLL N, THUROW K, et al. Fuzzy memberships descriptors for images[J]. IEEE Transactions on Fuzzy Systems, 2016, 24(1): 195-207.