贝叶斯学习

蒋良孝 | 中国地质大学（武汉）

CUG-Miner 机器学习与数据挖掘团队

ljiang@cug.edu.cn

http://www.escience.cn/people/jlx/

本章内容

一、贝叶斯学习基础知识

二、贝叶斯最优分类器

三、朴素贝叶斯分类器

四、朴素贝叶斯分类器改进

一、贝叶斯学习基础知识

- 用P(h)表示在没有观察到训练数据之前假设h拥有的初始概率，P(h)被称为假设h的先验概率。
- 先验概率反映了关于假设h是一正确假设的机会的背景知识；如果没有这一先验知识，可以简单地将每一候选假设赋予相同的先验概率。
- 类似地，P(D)表示训练数据D的先验概率，那么P(D|h)就表示假设h成立时D的概率。
- 在分类问题中，我们关心的是P(h|D)，即给定D时h的成立的概率，称为h的后验概率。

一、贝叶斯学习基础知识

- 交换规则：P(A, B)=P(B, A)
- 乘法规则：P(A, B)=P(A|B)P(B)=P(B|A)P(A)=P(B, A)
- 贝叶斯定理：P(h|D)=P(D|h)P(h)/P(D)
- 全概率法则：如果事件$A_1 ... A_n$互斥，且满足：

$$\sum_{i=1}^{n} P(A_i) = 1\text{，则 } P(B) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i)$$

一、贝叶斯学习基础知识

- 贝叶斯定理提供了从先验概率P(h)、P(D)以及P(D|h)计算后验概率P(h|D)的方法

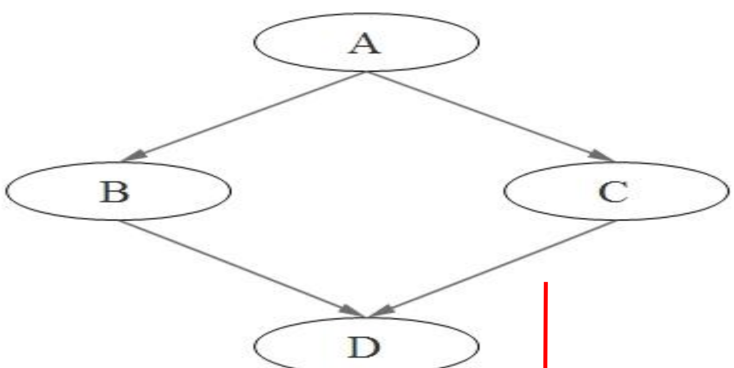
$$P(h|D)=\frac{P(D|h)P(h)}{P(D)}$$

- P(h|D)随着P(h)和P(D|h)的增长而增长，而随着P(D)的增长而减少。这是很合理的，因为如果D独立于h时被观察到的可能性越大，那么D对h的支持度就越小。

一、贝叶斯学习基础知识

- 贝叶斯网络，也叫贝叶斯信念网，是一种用来表示变量间连续概率的有向无环图模型，图中的节点表示变量，有向边表示变量间的依赖关系，依赖关系的强弱用标识在边旁边的条件概率来表示。
- 贝叶斯网络表示一组变量的联合概率分布。

$$P(A,B,C,D) = P(A|B,C,D)P(B|C,D)P(C|D)P(D)$$



贝叶斯网络学习 $\left\{\begin{array}{l}\text{结构学习}\\\text{参数学习}\end{array}\right.$

$$P(A,B,C,D) = P(A)P(B|A)P(C|A)P(D|B,C)$$

二、贝叶斯最优分类器

m

n

Instance ID	A_1	A_2	…	A_m	C
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
…					
n					

a_1	a_2	…	a_m	c=?

$$c(x) = \arg\max_{c \in C} P(c|a_1, a_2, \cdots, a_m)$$

贝叶斯定理

$$c(x) = \arg\max_{c \in C} \frac{P(a_1, a_2, \cdots, a_m|c)P(c)}{P(a_1, a_2, \cdots, a_m)}$$

全概率法则

$$c(x) = \arg\max_{c \in C} \frac{P(a_1, a_2, \cdots, a_m|c)P(c)}{\sum_c P(a_1, a_2, \cdots, a_m|c)P(c)}$$

NP-Hard Problem

(Chickering, 1996)

三、朴素贝叶斯分类器

m

Instance ID	A_1	A_2	…	A_m	C
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
…					
n					

n

a_1	a_2	…	a_m	c=?

属性条件独立

$$P(a_1, a_2, \cdots, a_m|c) = \prod_{i=1}^{m} P(a_i|c)$$

朴素贝叶斯分类器

$$c(x) = \arg\max_{c\in C} \frac{P(c)\prod_{i=1}^{m} P(a_i|c)}{\sum_c P(c)\prod_{i=1}^{m} P(a_i|c)}$$

分类：去分母

$$c(x) = \arg\max_{c\in C} P(c)\prod_{i=1}^{m} P(a_i|c)$$

三、朴素贝叶斯分类器

m

Instance ID	A_1	A_2	…	A_m	C
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
…					
n					

n

a_1	a_2	…	a_m	c=?

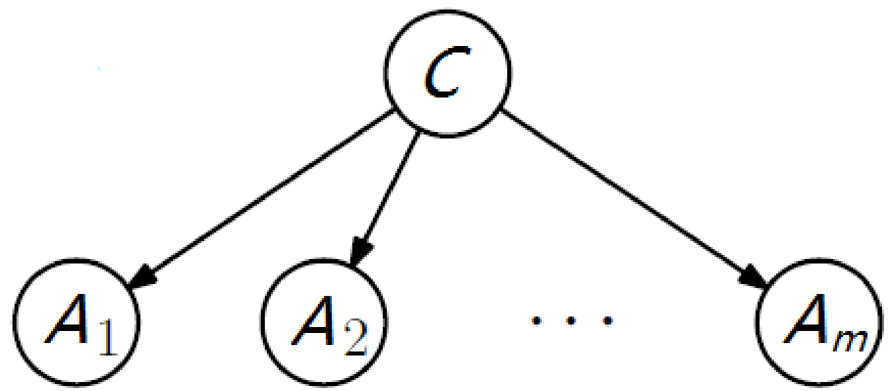
朴素贝叶斯分类器

Naive Bayes (NB)

$$c(x) = \arg\max_{c \in C} P(c) \prod_{i=1}^{m} P(a_i|c)$$

$$P(c) = \frac{\sum_{j=1}^{n} \delta(c_j, c) + 1}{n + n_c}$$

$$P(a_i|c) = \frac{\sum_{j=1}^{n} \delta(a_{ji}, a_i)\delta(c_j, c) + 1}{\sum_{j=1}^{n} \delta(c_j, c) + n_i}$$



三、朴素贝叶斯分类器

Day	**Outlook**	**Temperature**	**Humidity**	**Wind**	**PlayTennis**
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

15	**Sunny**	**Cool**	**High**	**Strong**	?

三、朴素贝叶斯分类器

$$c(x) = \arg\max_{c \in \{yes,no\}} P(c)P(sunny|c)P(cool|c)P(high|c)P(strong|c)$$

- 根据数据库，可以计算出上式需要的各项概率值
  - P(yes)=9/14=0.64
  - P(no)=5/14=0.36
  - P(strong|yes)=3/9=0.33
  - P(strong|no)=3/5=0.60
  - ...
- 求c(x)
  - P(yes)P(sunny|yes)P(cool|yes)P(high|yes)P(strong|yes)=0.0053
  - P(no)P(sunny|no)P(cool|no)P(high|no)P(strong|no)=0.0206
  - c(x) =no

三、朴素贝叶斯分类器

Knowl Inf Syst (2008) 14:1–37
DOI 10.1007/s10115-007-0114-2

SURVEY PAPER

Top 10 algorithms in data mining

Xindong Wu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang · Hiroshi Motoda · Geoffrey J. McLachlan · Angus Ng · Bing Liu · Philip S. Yu · Zhi-Hua Zhou · Michael Steinbach · David J. Hand · Dan Steinberg

Received: 9 July 2007 / Revised: 28 September 2007 / Accepted: 8 October 2007
Published online: 4 December 2007

Abstract This paper presents the top 10 data mining algorithms identified by the IEEE International Conference on Data Mining (ICDM) in December 2006: C4.5, *k*-Means, SVM, Apriori, EM, PageRank, AdaBoost, *k*NN, Naive Bayes, and CART. These top 10 algorithms are among the most influential data mining algorithms in the research community. With each algorithm, we provide a description of the algorithm, discuss the impact of the algorithm, and review current and further research on the algorithm. These 10 algorithms cover classification, clustering, statistical learning, association analysis, and link mining, which are all among the most important topics in data mining research and development.

三、朴素贝叶斯分类器

Machine Learning, 29, 131–163 (1997)

Bayesian Network Classifiers*

NIR FRIEDMAN nir@cs.berkeley.edu
Computer Science Division, 387 Soda Hall, University of California, Berkeley, CA 94720

DAN GEIGER dang@cs.technion.ac.il
Computer Science Department, Technion, Haifa, Israel, 32000

MOISES GOLDSZMIDT moises@erg.sri.com
SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025

Editor: G. Provan, P. Langley, and P. Smyth

Abstract. Recent work in supervised learning has shown that a surprisingly simple Bayesian classifier with strong assumptions of independence among features, called *naive Bayes*, is competitive with state-of-the-art classifiers such as C4.5. This fact raises the question of whether a classifier with less restrictive assumptions can perform even better. In this paper we evaluate approaches for inducing classifiers from data, based on the theory of learning *Bayesian networks*. These networks are factored representations of probability distributions that generalize the

四、朴素贝叶斯分类器改进

m

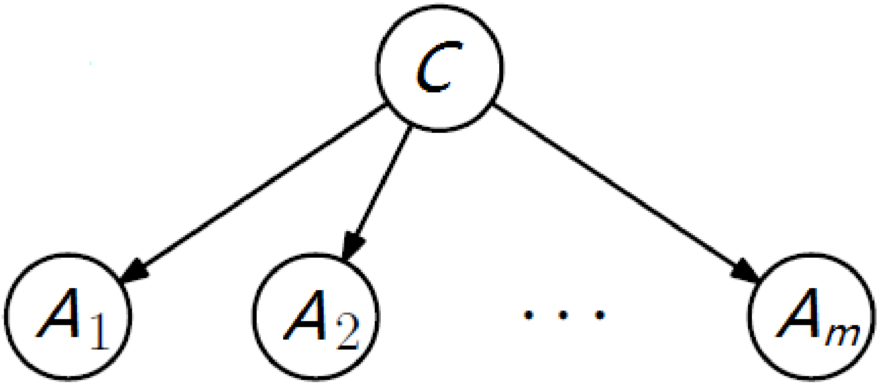
Instance ID	A_1	A_2	…	A_m	C
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
…					
n					
	a_1	a_2	…	a_m	c=?

n

朴素贝叶斯分类器

Naive Bayes (NB)

$$c(x) = \arg\max_{c \in C} P(c) \prod_{i=1}^{m} P(a_i|c)$$



1）处理算法：结构扩展

2）处理数据：面向特征（特征选择、特征加权）

面向实例（实例选择、实例加权）

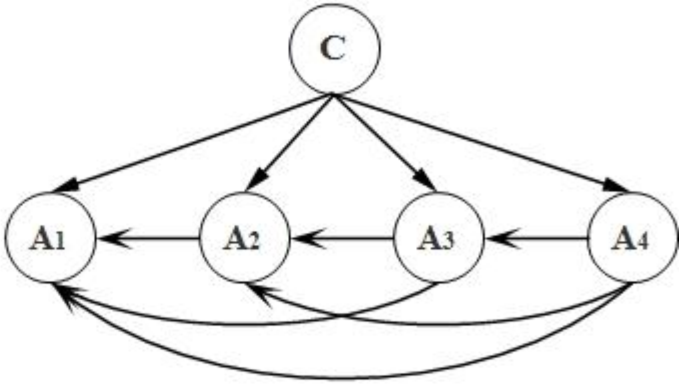
四、朴素贝叶斯分类器改进

结构扩展

$$c(x) = \arg\max_{c \in C} P(c) \prod_{i=1}^{m} P(a_i | \Pi_{a_i}, c)$$

$$P(c) = \frac{\sum_{j=1}^{n} \delta(c_j, c) + 1}{n + n_c}$$

$$P(a_i | \Pi_{a_i}, c) = \frac{\sum_{j=1}^{n} \delta(a_{ji}, a_i) \delta(\Pi_{a_{ji}}, \Pi_{a_i}) \delta(c_j, c) + 1}{\sum_{j=1}^{n} \delta(\Pi_{a_{ji}}, \Pi_{a_i}) \delta(c_j, c) + n_i}$$

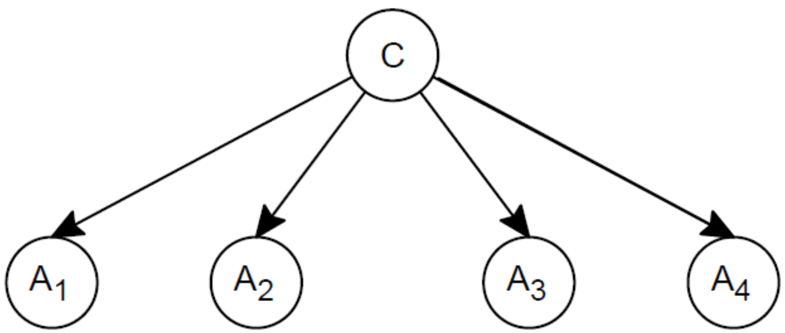


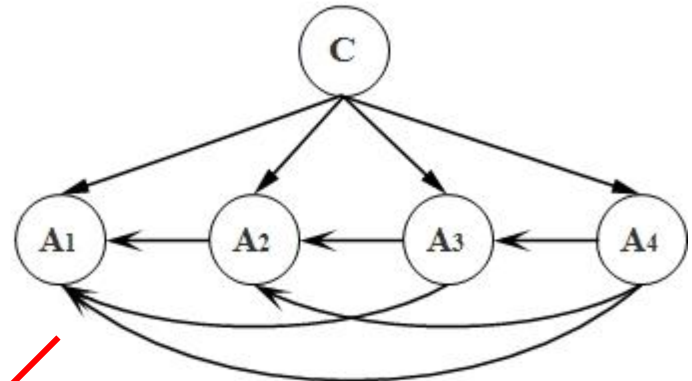
BNC

NP-Hard Problem
(Chickering, 1996)
Unrestricted not always Better
(Friedman et al., 1997)

四、朴素贝叶斯分类器改进

结构扩展



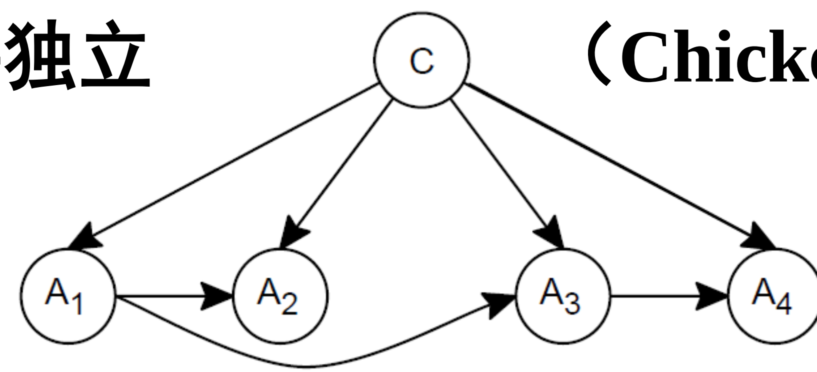


NB

属性条件独立

BNC

（Chickering, 1996）

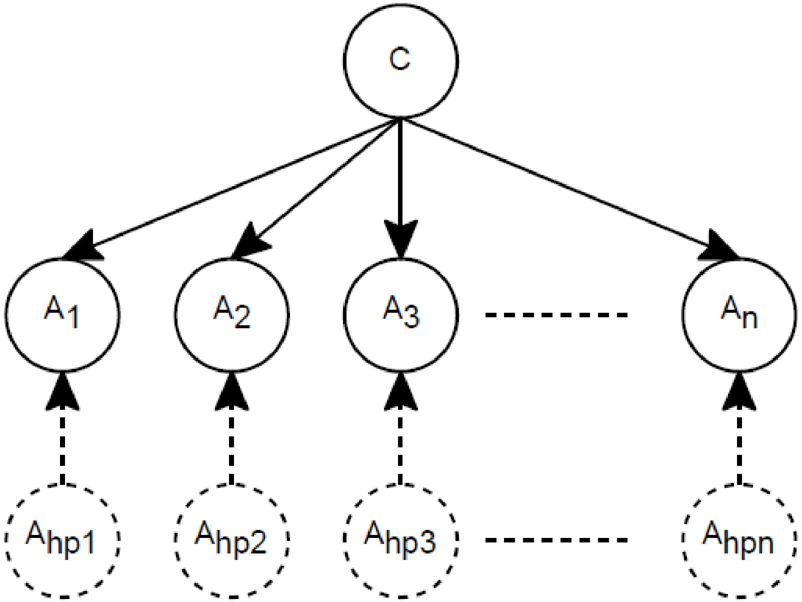


TAN (Friedman et al., 1997); SP (Keogh & Pazzani,1999)

FAN (Jiang et al., 2005);

四、朴素贝叶斯分类器改进

结构扩展



$$c(x) = \arg\max_{c \in C} P(c) \prod_{i=1}^{m} P(a_i | a_{hp_i}, c)$$

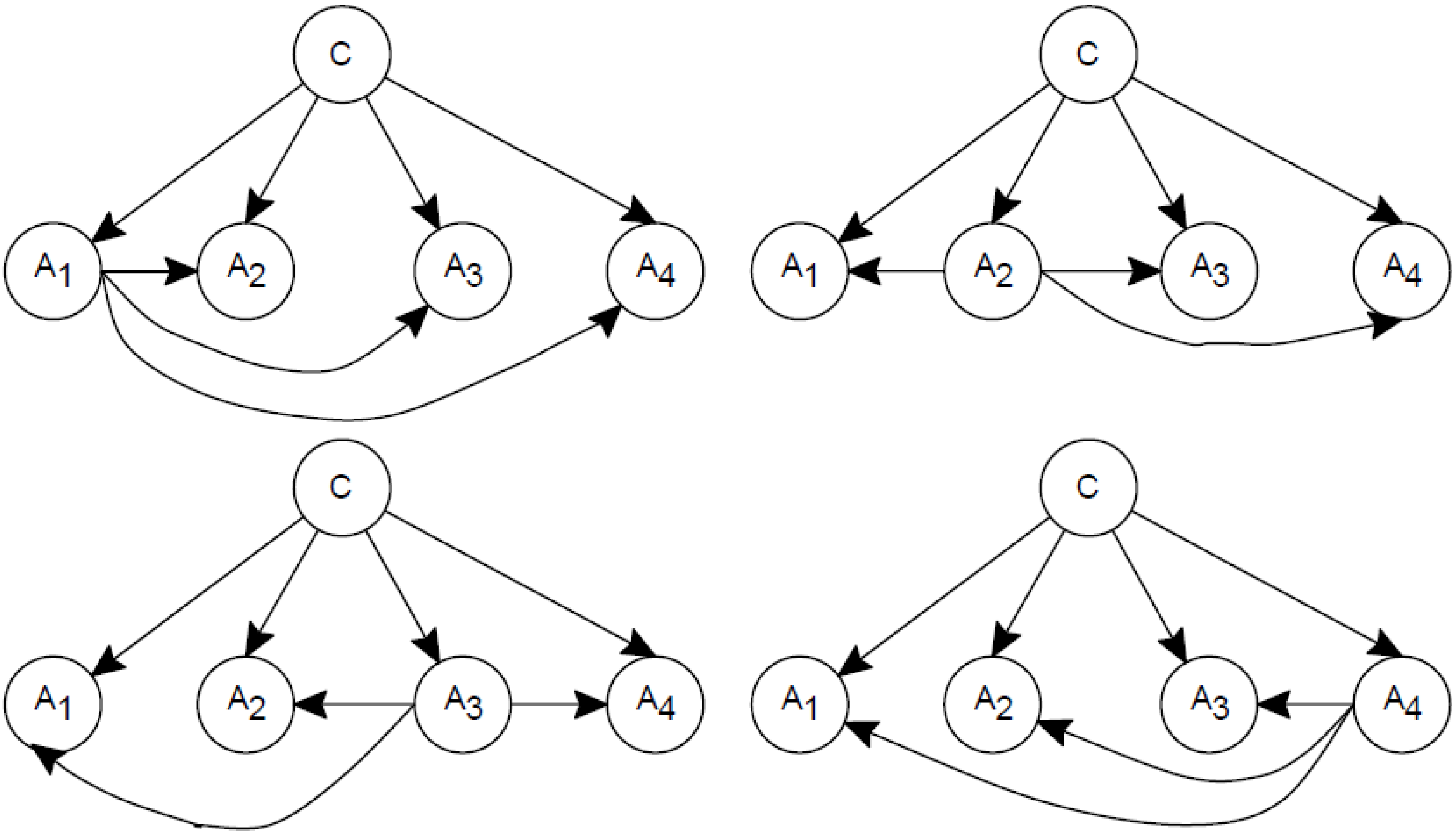
$$P(a_i | a_{hp_i}, c) = \sum_{j=1, j \neq i}^{m} W_{ij} * P(a_i | a_j, c)$$

$$W_{ij} = \frac{I_P(A_i; A_j | C)}{\sum_{j=1, j \neq i}^{m} I_P(A_i; A_j | C)}$$

HNB (Zhang & Jiang, 2005; Jiang et al., 2009)

四、朴素贝叶斯分类器改进

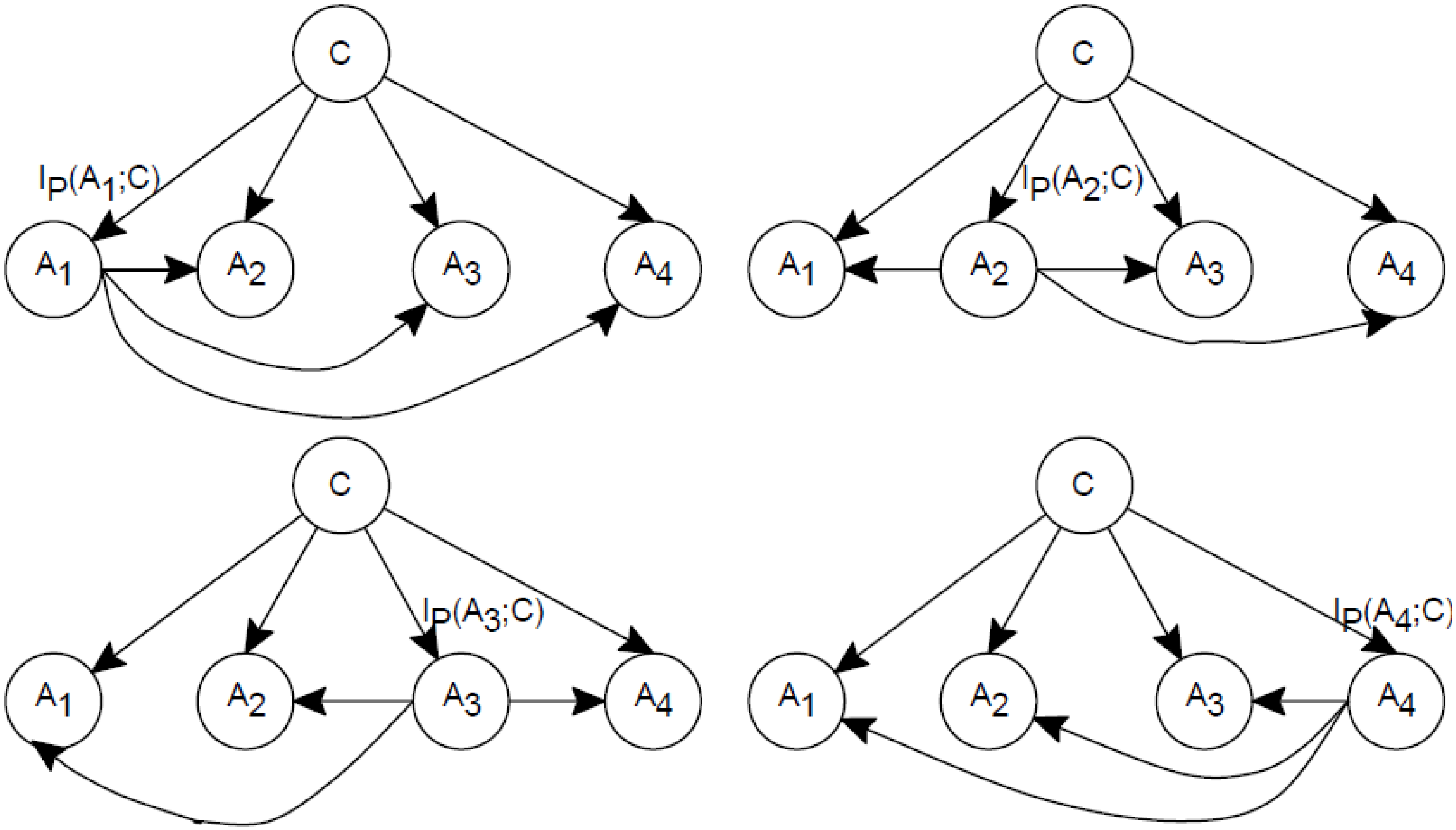
结构扩展



AODE (Webb et al., 2005)

四、朴素贝叶斯分类器改进

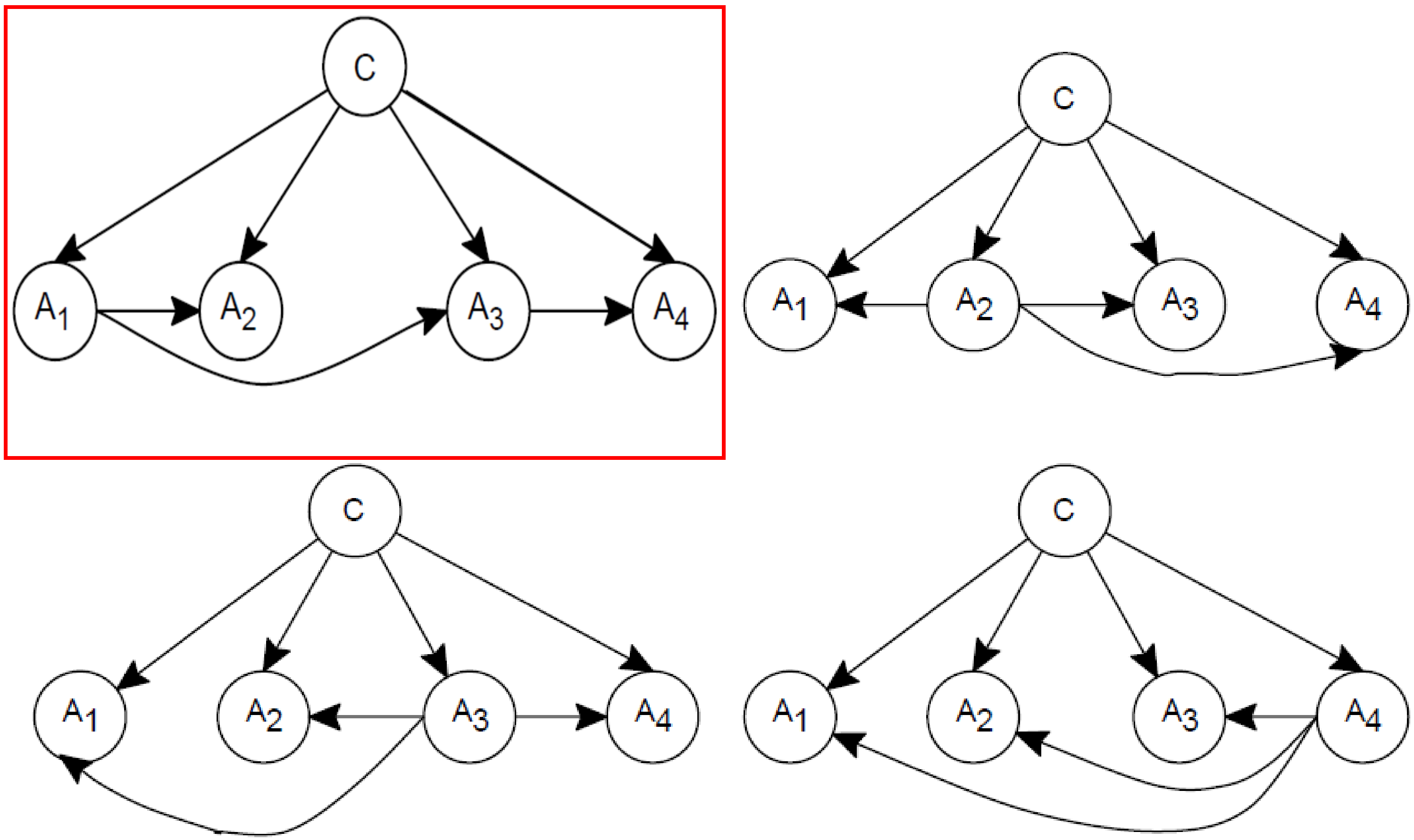
结构扩展



WADOE (Jiang et al., 2006； Jiang et al., 2012)

四、朴素贝叶斯分类器改进

结构扩展



ATAN (Jiang et al., 2012)

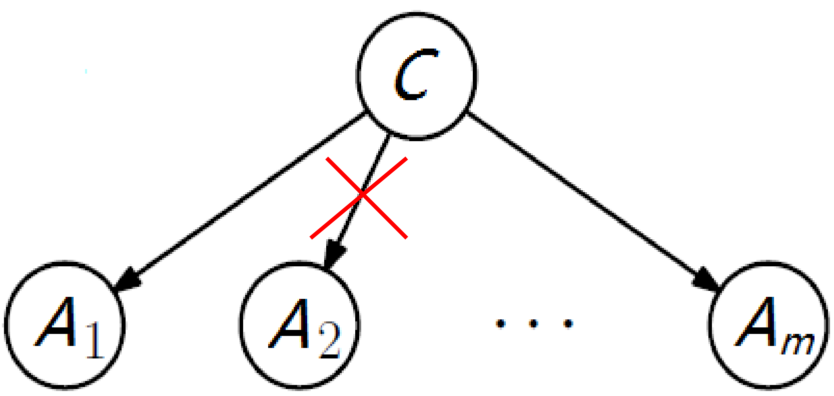
四、朴素贝叶斯分类器改进

特征选择

$$c(x) = \arg\max_{c \in C} P(c) \prod_{i=1}^{s} P(a_i|c)$$

$$P(c) = \frac{\sum_{j=1}^{n} \delta(c_j, c) + 1}{n + n_c}$$

$$P(a_i|c) = \frac{\sum_{j=1}^{n} \delta(a_{ji}, a_i)\delta(c_j, c) + 1}{\sum_{j=1}^{n} \delta(c_j, c) + n_i}$$



SBC (Ratanamahatana, 2003)
SB (Langley & Sage, 1994)
ENB (Jiang et al., 2005)
RSNB (Jiang et al., 2012)
TCSNB (Jiang et al., 2016)

四、朴素贝叶斯分类器改进

特征加权

$$c(x) = \arg\max_{c \in C} P(c) \prod_{i=1}^{m} P(a_i|c)^{w_i}$$

$$P(c) = \frac{\sum_{j=1}^{n} \delta(c_j, c) + 1}{n + n_c}$$

$$P(a_i|c) = \frac{\sum_{j=1}^{n} \delta(a_{ji}, a_i)\delta(c_j, c) + 1}{\sum_{j=1}^{n} \delta(c_j, c) + n_i}$$

GRWNB (Zhang & Sheng, 2004)
DTWNB (Hall, 2007)
KLMWNB (Lee et al., 2011)
DFWNB (Jiang et al., 2016)
CFWNB (Jiang et al., 2018)
DEWNB (Wu & Cai, 2011)
CLLFWNB, MSEFWNB (Zaidi et al., 2013)

四、朴素贝叶斯分类器改进

实例选择
局部学习

$$c(x) = \arg\max_{c \in C} P(c) \prod_{i=1}^{m} P(a_i|c)$$

$$P(c) = \frac{\sum_{j=1}^{s} \delta(c_j, c) + 1}{s + n_c}$$

$$P(a_i|c) = \frac{\sum_{j=1}^{s} \delta(a_{ji}, a_i)\delta(c_j, c) + 1}{\sum_{j=1}^{s} \delta(c_j, c) + n_i}$$

NBTree (Kohavi, 1996)
LBR (Zheng & Webb, 2002)
LWNB (Frank et al., 2003)
SNNB (Xie et al., 2002)
ICLNB(Jiang et al., 2005；Jiang et al., 2008)
DLNB (Jiang et al., 2009); CNNB (Jiang et al., 2010)

四、朴素贝叶斯分类器改进

实例加权

$$c(x) = \arg\max_{c \in C} P(c) \prod_{i=1}^{m} P(a_i|c)$$

$$P(c) = \frac{\sum_{j=1}^{n} w_j \delta(c_j, c) + 1}{\sum_{j=1}^{n} w_j + n_c}$$

$$P(a_i|c) = \frac{\sum_{j=1}^{n} w_j \delta(a_{ji}, a_i) \delta(c_j, c) + 1}{\sum_{j=1}^{n} w_j \delta(c_j, c) + n_i}$$

IWNB (Jiang et al., 2010)

CSNB (Jiang et al., 2014)

NB-MCT (Jiang et al., 2015)

BNB(Elkan, 1997)

DWNB (Jiang et al., 2012)

资助
教育部新世纪优秀人才支持计划（NCET-12-0953）
武汉市青年科技晨光计划（2015070404010202）
国家青年科学基金项目（61203287）

贝叶斯网络分类器：算法与应用

BEIYESI WANGLUO FENLEIQI:
SUANFA YU YINGYONG

蒋良孝 李超群 编著

中国地质大学出版社
ZHONGGUO DIZHI DAXUE CHUBANSHE

图书在版编目(CIP)数据

贝叶斯网络分类器:算法与应用/蒋良孝,李超群编著. 一武汉:中国地质大学出版社,2015.12

ISBN 978-7-5625-3780-9

Ⅰ.①贝…
Ⅱ.①蒋… ②李…
Ⅲ.①贝叶斯理论-应用-数据采集
Ⅳ.①TP274

中国版本图书馆 CIP 数据核字(2015)第 307024 号

贝叶斯网络分类器:算法与应用		蒋良孝 李超群 编著
责任编辑:张 琰 党梅梅	组稿编辑:张 琰	责任校对:周 旭
出版发行:中国地质大学出版社(武汉市洪山区鲁磨路 388 号)		邮政编码:430074
电 话:(027)67883511	传 真:67883580	E-mail:cbb @ cug.edu.cn
经 销:全国新华书店		http://www.cugp.cug.edu.cn
开本:787 毫米×1092 毫米 1/16		字数:190 千字 印张:7.5
版次:2015 年 12 月第 1 版		印次:2015 年 12 月第 1 次印刷
印刷:武汉市籍缘印刷厂		印数:1—500 册
ISBN 978-7-5625-3780-9		定价:35.00 元

如有印装质量问题请与印刷厂联系调换