Presentation:

Analyzing the Impact of the Discretization Method When Comparing Bayesian Classifiers

The Twenty Third International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2010)

03/06/2010

M. Julia Flores, José A. Gámez, Ana M. Martínez and José M. Puerta

Computing Systems Department

Albacete - UCLM - Spain

Outline

1. **Motivation**
2. **Bayesian Networks Classifiers**
   - Naive Bayes
   - TAN
   - AODE
   - HAODE
3. **Discretization Methods**
4. **Experimental Methodology and Results**
   - Study in terms of accuracy
   - Study in terms of bias and variance
5. **Conclusions and Future Work**

Ana M. Martínez

Outline

1. **Motivation**

2. **Bayesian Networks Classifiers**
   - Naive Bayes
   - TAN
   - AODE
   - HAODE

3. **Discretization Methods**

4. **Experimental Methodology and Results**
   - Study in terms of accuracy
   - Study in terms of bias and variance

5. **Conclusions and Future Work**

Motivation

- Discretization is probably one of the **pre-processing techniques most broadly used** in machine learning.
- The real distribution of the data is replaced with a mixture of uniform distributions.
- **Reasons**:
  - Some methods **can only deal with discrete** variables.
  - **Improves** an algorithm's **run time**.
  - **Reduction** of the **noise** which is quite possibly present in the data.
- Many distinct techniques for discretization can be found in literature.
- *Should we worry about the discretization method applied when designing the set of experiments?*.
  - **Yes!** Number of experiments will be muliplied by k.
  - **No!** Our experiments will be k times faster.

Overview

- **Empiric analysis** of this problem:
  - Subset of classifiers based on BNs:
    - **Naive Bayes**,
    - **TAN**,
    - **AODE** and
    - **HAODE** [Flores et al., 2009].
  - Discretization methods:
    - **Supervised discretization methods:** (Fayyad and Irani).
    - **Unsupervised** (equal frequency and width, 5 and 10 bins and optimizing the number of bins based on the entropy).
- **Aim:** analyzing a set of discretization methods and check if the result obtained by the classifiers is sensitive to the discretization method.

Outline

1 Motivation

2 **Bayesian Networks Classifiers**
- Naive Bayes
- TAN
- AODE
- HAODE

3 Discretization Methods

4 Experimental Methodology and Results
- Study in terms of accuracy
- Study in terms of bias and variance

5 Conclusions and Future Work

Outline

1 Motivation

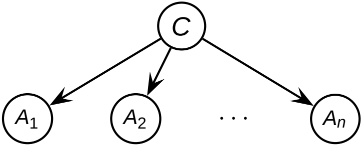
2 **Bayesian Networks Classifiers**
- Naive Bayes
- TAN
- AODE
- HAODE

3 Discretization Methods

4 Experimental Methodology and Results
- Study in terms of accuracy
- Study in terms of bias and variance

5 Conclusions and Future Work

NB classifier (Naive Bayes)



- The attributes are conditionally independent given the class value $I(A_i, A_j | C)$.

$$c_{MAP} = argmax_{c \in \Omega_C} \; p(c) \prod_{i=1}^{n} p(a_i | c)$$

- **Time complexity**: linear in training and classification time.
- Drawbacks:
  - ✕ : It does not work properly in certain datasets.
  - ✕ : Dependencies between attributes reduce, unavoidably, the prediction capability of NB.
  - ✕ : Not only interesting to be right in the classification in certain applications.

Outline

1 Motivation

2 Bayesian Networks Classifiers
- Naive Bayes
- TAN
- AODE
- HAODE

3 Discretization Methods

4 Experimental Methodology and Results
- Study in terms of accuracy
- Study in terms of bias and variance

5 Conclusions and Future Work

TAN classifier (Tree Augmented Naive Bayes)

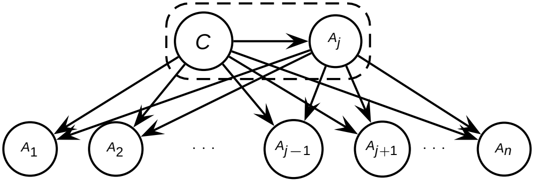
- TAN releases the conditional independence restriction without a large increase in the complexity of the construction process.
- Learns a **maximum weighted spanning tree** based on the **conditional mutual information** between two attributes given the class label.
- Chooses a **variable as root** and completing the model by adding a link from the class to each attribute.
- Considered a fair **trade-off between model complexity and model accuracy**.

Outline

1. Motivation
2. **Bayesian Networks Classifiers**
   - Naive Bayes
   - TAN
   - AODE
   - HAODE
3. Discretization Methods
4. Experimental Methodology and Results
   - Study in terms of accuracy
   - Study in terms of bias and variance
5. Conclusions and Future Work

AODE classifier I (Averaging One-Dependence Estimators)

- **AODE** is significantly better in terms of error reduction compared to the rest of semi-naive techniques [Zheng and Webb, 2005].



- **Training and classification**: Quadratic

× : **Only discrete variables.**

Outline

1 Motivation

2 **Bayesian Networks Classifiers**
- Naive Bayes
- TAN
- AODE
- HAODE

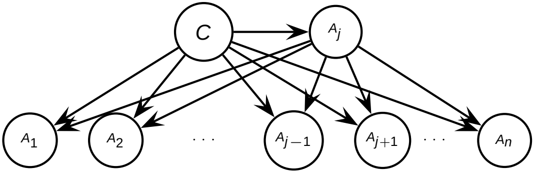
3 Discretization Methods

4 Experimental Methodology and Results
- Study in terms of accuracy
- Study in terms of bias and variance

5 Conclusions and Future Work

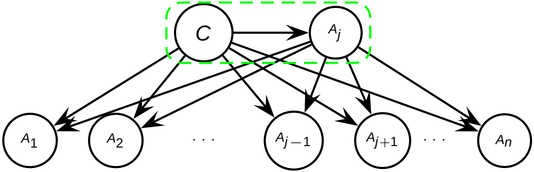
HAODE Classifier (Hybrid AODE)

- **Discrete** super-parent (A_j) in every model.



HAODE Classifier (Hybrid AODE)

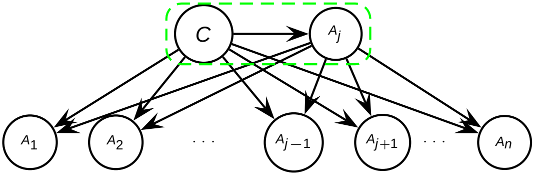
- **Discrete** super-parent ($\boldsymbol{A_j}$) in every model.



HAODE Classifier (Hybrid AODE)

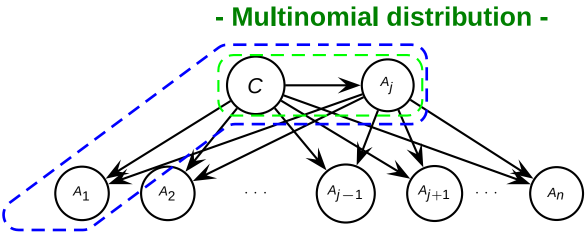
- **Discrete** super-parent (A_j) in every model.

- Multinomial distribution -



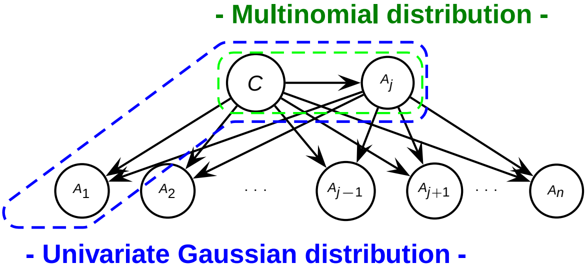
HAODE Classifier (Hybrid AODE)

- **Discrete** super-parent (A_j) in every model.



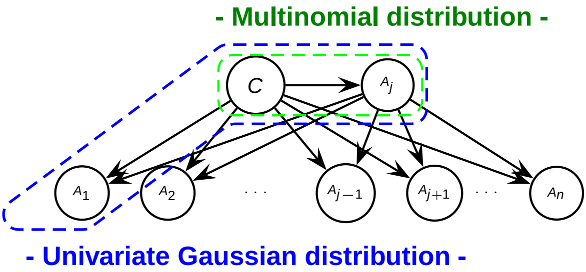
HAODE Classifier (Hybrid AODE)

- **Discrete** super-parent ($\boldsymbol{A_j}$) in every model.



HAODE Classifier (Hybrid AODE)

- **Discrete** super-parent ($\boldsymbol{A_j}$) in every model.



√ Able to deal with **hybrid datasets** too.

Outline

1. Motivation
2. Bayesian Networks Classifiers
   - Naive Bayes
   - TAN
   - AODE
   - HAODE
3. **Discretization Methods**
4. Experimental Methodology and Results
   - Study in terms of accuracy
   - Study in terms of bias and variance
5. Conclusions and Future Work

Evaluated Discretization Methods

- **Equal-width discretization** (unsupervised) [**EW5 and EW10**]
  - Divides the range of the attribute into *b* bins with **same width**.
  - Usual to set this value to **5** or **10 bins**.
  - **Search of the most appropriate** value for *b* by **minimizing the entropy** of the partition [**EWE**].
- **Equal-depth (or frequency) discretization** (unsupervised) [**EF5 and EF10**]
  - Divides into *b* bins so that they contain approximately the same number of training instances (t/b).
- **Minimum entropy-based discretization by Fayyad & Irani** (supervised) [**FeI**]
  - A binary discretization is performed in that candidate cut-off point which **minimizes the entropy**.
  - Recursive process by applying the **MDL criterion** to decide when to stop.

Outline

1. Motivation
2. Bayesian Networks Classifiers
   - Naive Bayes
   - TAN
   - AODE
   - HAODE
3. Discretization Methods
4. **Experimental Methodology and Results**
   - Study in terms of accuracy
   - Study in terms of bias and variance
5. Conclusions and Future Work

Experimental Frame

- Experiments over the **26 numeric datasets** (Weka home page and UCI repository).

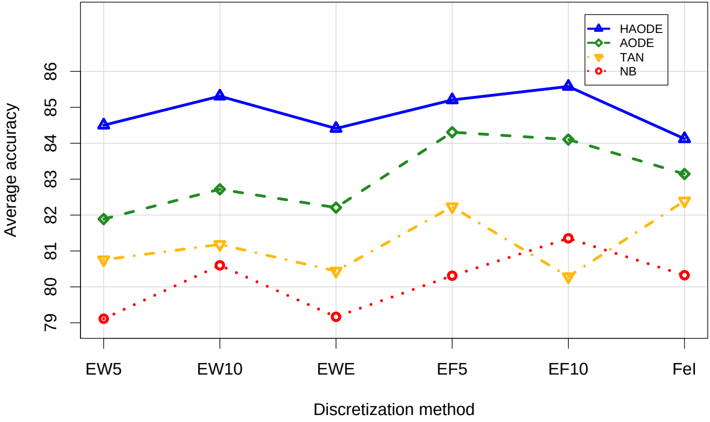
Table: Main characteristics of the datasets: number of predictive variables (*n*), number of classes (*k*), and number of instances (*t*).

Id	Dataset	*n*	*k*	*t*	Id	Dataset	*n*	*k*	*t*
1	balance-scale	4	3	625	14	mfeat-fourier	76	10	2000
2	breast-w	9	2	699	15	mfeat-karh	64	10	2000
3	diabetes	8	2	768	16	mfeat-morph	6	10	2000
4	ecoli	7	8	336	17	mfeat-zernike	47	10	2000
5	glass	9	7	214	18	optdigits	64	9	5620
6	hayes-roth	4	4	160	19	page-blocks	10	5	5473
7	heart-statlog	13	2	270	20	pendigits	16	9	10992
8	ionosphere	34	2	351	21	segment	19	7	2310
9	iris	4	3	150	22	sonar	60	2	208
10	kdd-JapanV	14	9	9961	23	spambase	57	2	4601
11	letter	16	26	20000	24	vehicle	18	4	946
12	liver-disorders	6	2	345	25	waveform-5000	40	3	5000
13	mfeat-factors	216	10	2000	26	wine	13	3	178

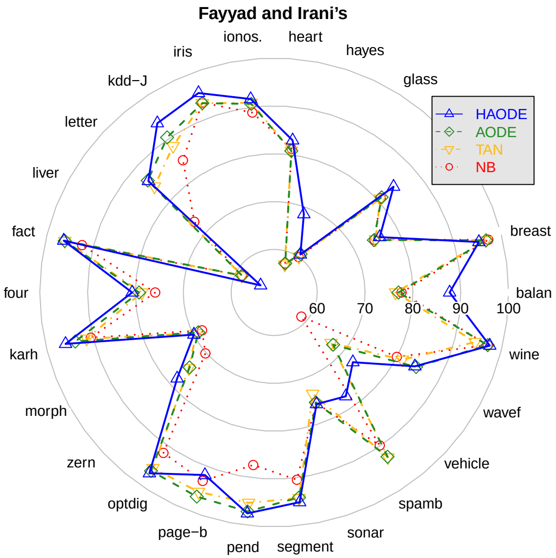
- 5x2*cv* for the evaluation process.

Study in terms of accuracy

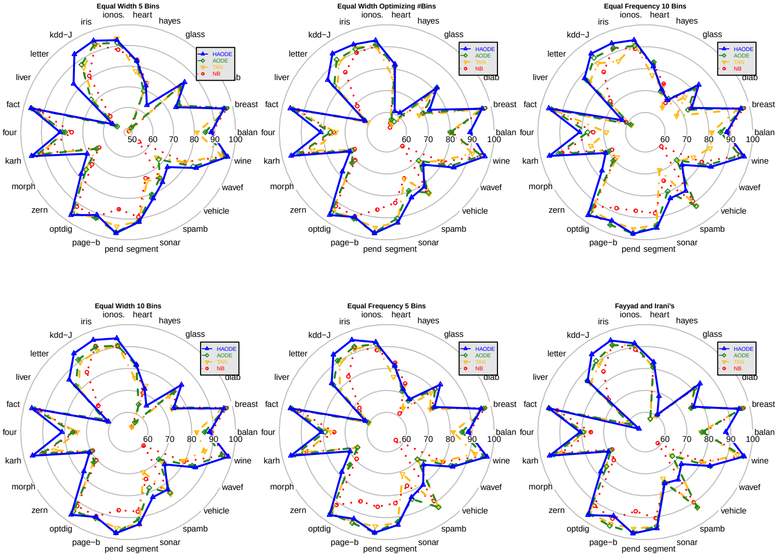
- Percentage of correctly predicted instances in the test dataset.



Study in terms of accuracy



Study in terms of accuracy



Comparisons between discretization methods:

- **Friedman's tests** to perform the multiple comparison of the different discretization methods for each classifier and **post-hoc tests**.
- [Demšar, 2006, García and Herrera, 2009] guidelines.

	FRIEDMAN	IMAN-DAV.	NEMENYI
NB	Reject H_0 (0.034)	Not necessary	• None
TAN	Reject H_0 (0.006)	Not necessary	• FEI vs (FEW&EF10)
AODE	Accept H_0 (0.069)	Accept H_0 (0.065)	• None
HAODE	Accept H_0 (0.052)	Reject H_0 (0.049)	• None

Comparisons between discretization methods:

- **Friedman's tests** to perform the multiple comparison of the different discretization methods for each classifier and **post-hoc tests**.
- [Demšar, 2006, García and Herrera, 2009] guidelines.

	FRIEDMAN	IMAN-DAV.	NEMENYI
NB	Reject H_0 (0.034)	Not necessary	• None
TAN	Reject H_0 (0.006)	Not necessary	• FEI vs (FEW&EF10)
AODE	Accept H_0 (0.069)	Accept H_0 (0.065)	• None
HAODE	Accept H_0 (0.052)	Reject H_0 (0.049)	• None

Comparisons between discretization methods:

- **Friedman's tests** to perform the multiple comparison of the different discretization methods for each classifier and **post-hoc tests**.
- [Demšar, 2006, García and Herrera, 2009] guidelines.

	FRIEDMAN	IMAN-DAV.	NEMENYI
NB	Reject H_0 (0.034)	Not necessary	• None
TAN	Reject H_0 (0.006)	Not necessary	• FEI vs (FEW&EF10)
AODE	Accept H_0 (0.069)	Accept H_0 (0.065)	• None
HAODE	Accept H_0 (0.052)	Reject H_0 (0.049)	• None

Comparisons between discretization methods:

- **Friedman's tests** to perform the multiple comparison of the different discretization methods for each classifier and **post-hoc tests**.
- [Demšar, 2006, García and Herrera, 2009] guidelines.

	FRIEDMAN	IMAN-DAV.	NEMENYI
NB	Reject H_0 (0.034)	Not necessary	• None
TAN	Reject H_0 (0.006)	Not necessary	• FEI vs (FEW&EF10)
AODE	Accept H_0 (0.069)	Accept H_0 (0.065)	• None
HAODE	Accept H_0 (0.052)	Reject H_0 (0.049)	• None

Comparisons between discretization methods:

- **Friedman's tests** to perform the multiple comparison of the different discretization methods for each classifier and **post-hoc tests**.
- [Demšar, 2006, García and Herrera, 2009] guidelines.

	FRIEDMAN	IMAN-DAV.	NEMENYI
NB	Reject H_0 (0.034)	Not necessary	• None
TAN	Reject H_0 (0.006)	Not necessary	• FEI vs (FEW&EF10)
AODE	Accept H_0 (0.069)	Accept H_0 (0.065)	• None
HAODE	Accept H_0 (0.052)	Reject H_0 (0.049)	• None

Comparisons between discretization methods:

- **Friedman's tests** to perform the multiple comparison of the different discretization methods for each classifier and **post-hoc tests**.
- [Demšar, 2006, García and Herrera, 2009] guidelines.

	FRIEDMAN	IMAN-DAV.	NEMENYI
NB	Reject H_0 (0.034)	Not necessary	• None
TAN	Reject H_0 (0.006)	Not necessary	• FEI vs (FEW&EF10)
AODE	Accept H_0 (0.069)	Accept H_0 (0.065)	• None
HAODE	Accept H_0 (0.052)	Reject H_0 (0.049)	• None

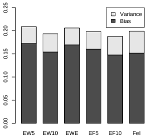
- The **null hypothesis** (H_0) states that **there is not difference between the algorithms**.
- $\alpha = \mathbf{0.05}$. for all the cases.
- (In brackets the p-value obtained).

Comparisons between classifiers:

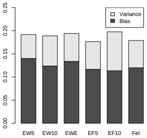
- **Friedman test**: statistical **difference in all cases.**
- **Nemenyi tests**:
  - **Nemenyi**: **HADOE is significantly better than NB and TAN** in all cases. Also states that AODE is better than NB when EW5 is used and TAN for EF10. .
- In all cases HAODE is placed in first position and AODE in the second one by the **ranking** performed by the Friedman test.

Study in terms of bias and variance

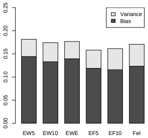
- **Error component**:
  - **Bias**: **systematic error** when learning the algorithm.
  - **Variance**: random variation existing on the training data and from the random behavior when learning the algorithm (**sensitiveness**).
  - **Irreducible term**:error existing in an optimal algorithm (**noise level in data**).



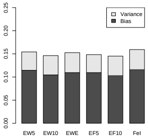
(a) NB



(b) TAN



(c) AODE



(d) HAODE

Outline

1 Motivation

2 Bayesian Networks Classifiers
   - Naive Bayes
   - TAN
   - AODE
   - HAODE

3 Discretization Methods

4 Experimental Methodology and Results
   - Study in terms of accuracy
   - Study in terms of bias and variance

5 **Conclusions and Future Work**

Conclusions and Future Work

- Study the effect in terms of **accuracy**, **bias** and **variance** obtained when applying some of the most common **discretization** methods to **NB, TAN, AODE and HAODE**.
- **Goals:**
  1. Comparison between AODE and HAODE to see if results in [Flores et al., 2009] could be generalized.
     - **HAODE** obtains **better** performance than AODE **in all cases**.
     - HAODE is less sensitive to the discretization method.
  2. Should we add a new parameter to our experiments?
     - If the **set of datasets is large enough**, the **discretization method** applied becomes **irrelevant** in the comparison.
     - As a side conclusion, for some concrete domains the discretization method matters.
- **Future work:** extend this work by adding **more sophisticated discretization techniques**, e.g. proportional discretization or equal size discretizations [Yang and Webb, 2009].

Thanks for your attention

Questions? Suggestions?

Referencias I

[Demšar, 2006] Demšar, J. 2006.
Statistical Comparisons of Classifiers over Multiple Data Sets.
J. Mach. Learn. Res., 7:1–30.

[Flores et al., 2009] Flores, M. J., Gámez, J. A., Martínez, A. M., and Puerta, J. M. 2009.
GAODE and HAODE: two proposals based on AODE to deal with continuous variables.
In: Danyluk, A. P., Bottou, L., and Littman, M. L. (Editors), ICML. ACM, 382:40.

[García and Herrera, 2009] García, S. and Herrera, F. 2009.
An Extension on “Statistical Comparisons of Classifiers over Multiple Data Sets” for all Pairwise Comparisons.
J. Mach. Learn. Res., 9:2677–2694.

[Yang and Webb, 2009] Yang, Y. and Webb, G. I. 2009.
Discretization for Naive-Bayes Learning: Managing Discretization Bias and Variance.
Mach. Learn., 74(1):39–74.

Referencias II

[Zheng and Webb, 2005] Zheng, F. and Webb, G. 2005.
A Comparative Study of Semi-naive Bayes Methods in Classification Learning.
In: Proc. of the 4th Australasian Data Mining Conf. (AusDM05), Sydney. University of Technology, pages 141–156.