ExpertBayes: Automatically Refining Manually Built Bayesian Networks

ICMLA 2014– December 4th 2014 – Detroit, USA

CRACS INESCTEC
FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO
U.PORTO
WISCONSIN UNIVERSITY OF WISCONSIN-MADISON
FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Ezilda Almeida
Pedro Ferreira
Tiago T. V. Vinhoza
Inês Dutra
Paulo Borges
Yirong Wu
Elizabeth Burnside

CRACS INESCTEC

Outline

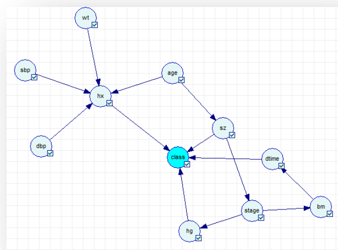
- Objectives
- Dataset
- Methodology and Tools
- Results and Analysis
- *ExpertBayes* (graphical user interface)
- Conclusions and Future Work

Outline

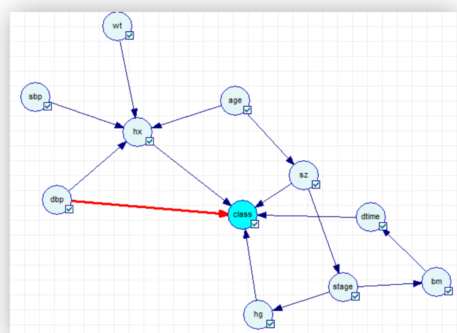
- Objectives
- Dataset
- Methodology and Tools
- Results and Analysis
- *ExpertBayes* (graphical user interface)
- Conclusions and Future Work

Objectives

Network constructed manually



ExpertBayes



New network with better score

Outline

- Objectives
- Dataset
- Methodology and Tools
- Results and Analysis
- *ExpertBayes* (graphical user interface)
- Conclusions and Future Work

Dataset

- Prostate Cancer:
  - ➢496 cases
  - ➢Each case refers to the clinical history of each patient
- Breast Cancer (1) :
  - ➢100 cases
  - ➢Each case refers to a breast nodule from mammography results
- Breast Cancer (2) :
  - ➢241 cases
  - ➢Each case refers to a breast nodule from mammography results

Attributes

- Prostate Cancer

11 Attributes

Age (age)
Weight (wt)
Family history of cancer (hx)
Systolic blood pressure (Sbp)
Diastolic blood pressure (Dbp)
Hmoglobins (hg)
Clinical stage (stage)
Doubling time PSA (Dtime)
Size of the prostate (size)
Bony metastases (bm)
Status (status)

351 Dead

(+)

145 Alive

(-)

Attributes

- Breast Cancer(1)

33 Attributes

Age
Disease
BreastDensity
MassesShape
MassesDensity
MassesSize
PostOpChange
MassesStability
Calc_Milk
…
BinaryDx

45 Benign

(-)

55 Malignant

(+)

Attributes

- Breast Cancer(2)

8 Attributes

Age
Mass_Shape
Mass_Margins
Depth
Size
Overall_Breast_Composition
Retro_Density
Biopsy_Outcome

153 Benign

(-)

88 Malignant

(+)

Outline

- Objectives
- Dataset
- Methodology and Tools
- Results and Analysis
- *ExpertBayes* (graphical user interface)
- Conclusions and Future Work

Methodology and Tools

- eclipse to develop ExpertBayes using Java language
- 

- **5-fold cross-validation** to train and test our models
- **t-test** was used to validate the results
  - **Significance level: 0.05**

CRACS INESCTEC

Outline

- Objectives
- Dataset
- Methodology and Tools
- Results and Analysis
- *ExpertBayes* (graphical user interface)
- Conclusions and Future Work

CRACS INESCTEC

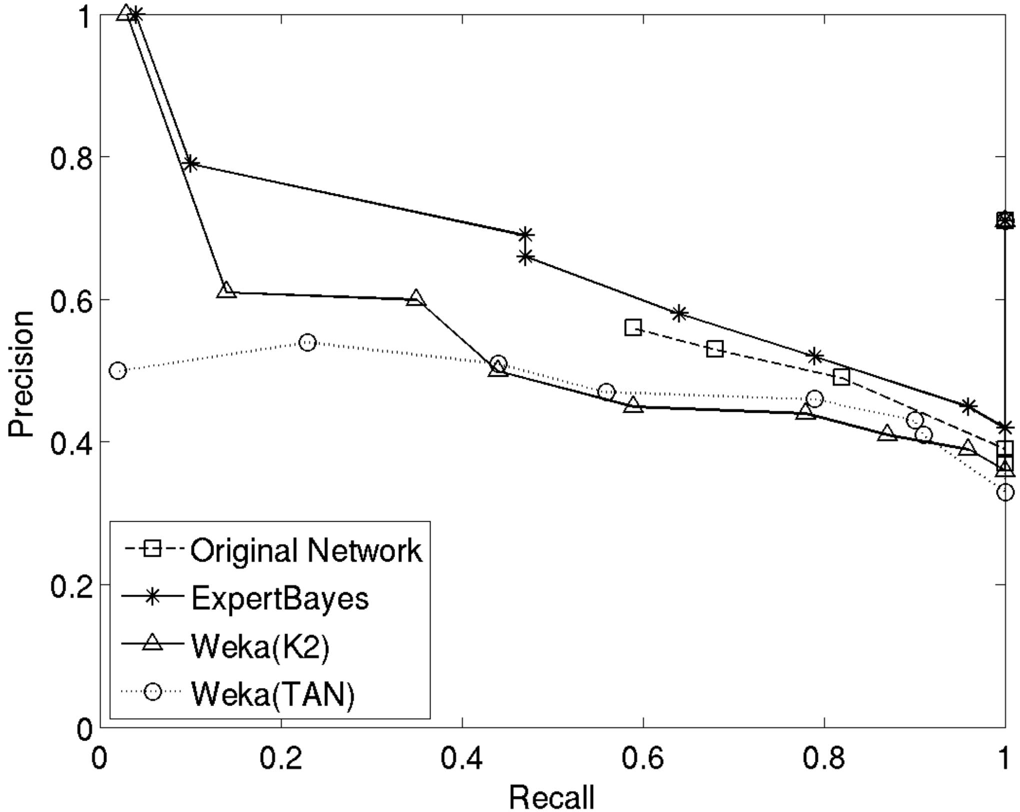
Results and Analysis

- CCI(%) test set - averaged across 5-folds

Dataset	Original	ExpertBayes	WEKA-K2	WEKA-TAN
Prostate Cancer	74	76	74	71
Breast Cancer (1)	49	63	59	57
Breast Cancer (2)	49	64	80	79

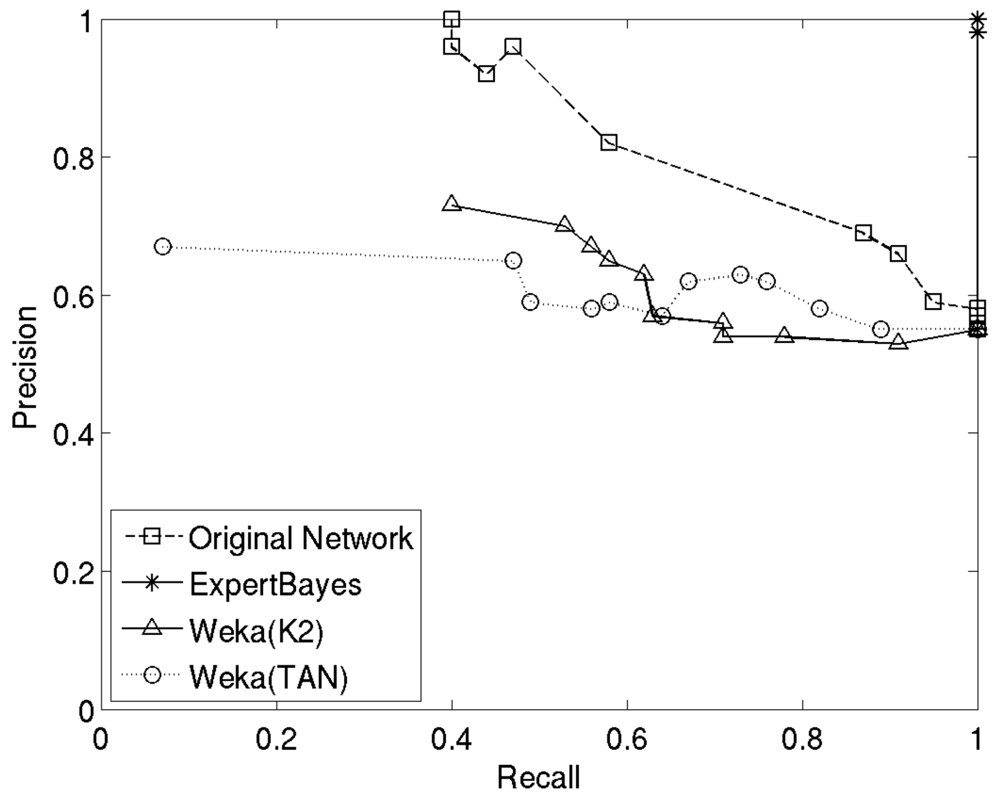
Results and Analysis

- Precision-Recall Curves for various thresholds
  - Prostate



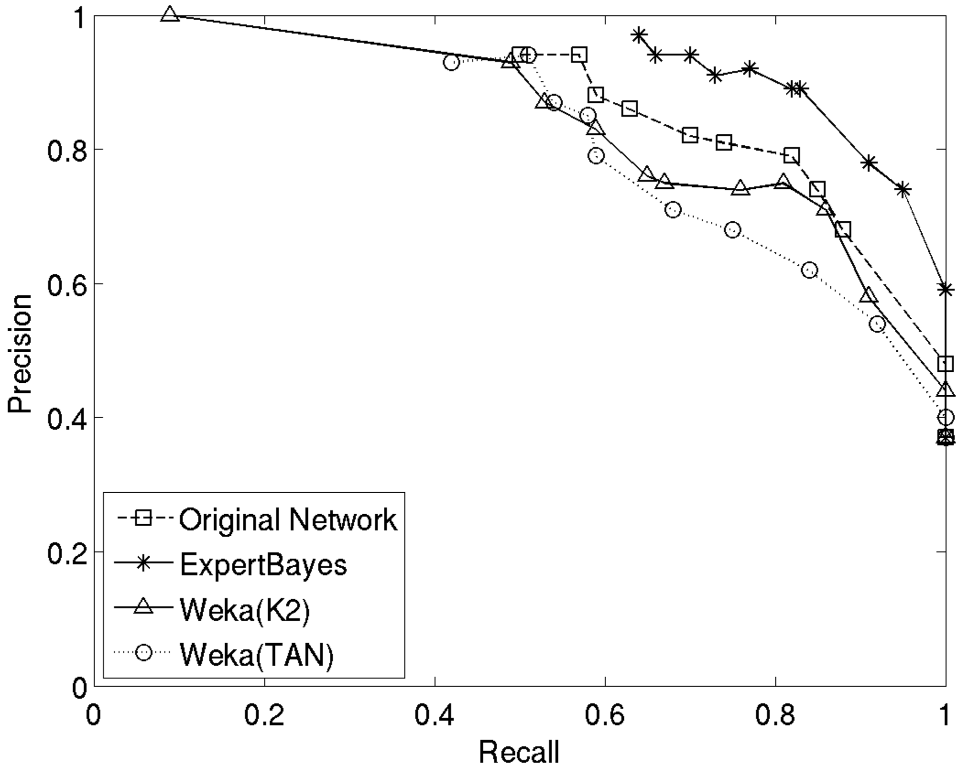
Results and Analysis

- Precision-Recall Curves for various thresholds
  - Breast Cancer (1)



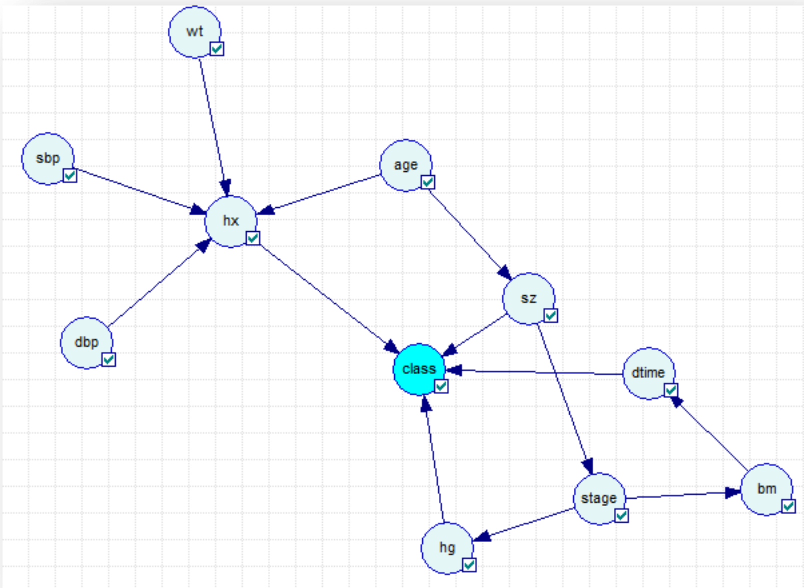
Results and Analysis

- Precision-Recall Curves for various thresholds
  - Breast Cancer (2)

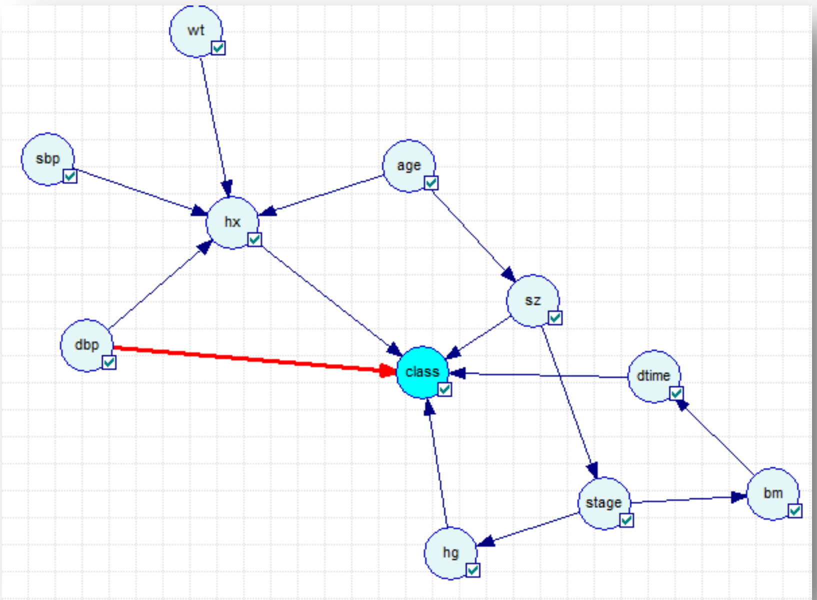


Results and Analysis

Original Network



ExpertBayes

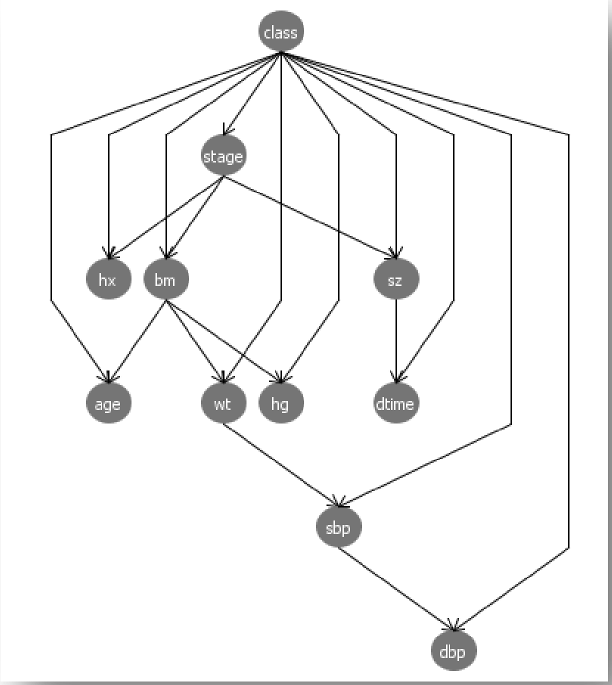


CCI :74%

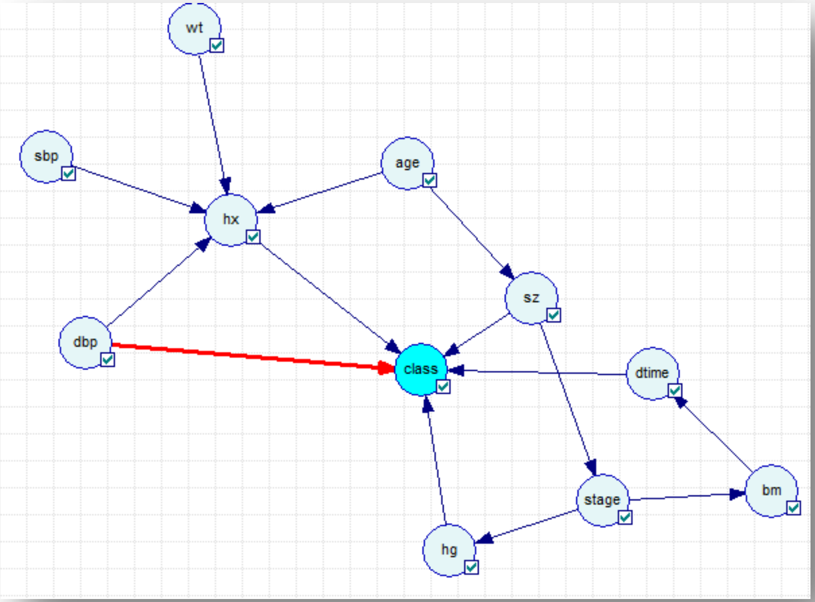
CCI :76%

Results and Analysis

Weka TAN



ExpertBayes



CCI :71%

CCI :76%

Outline

- Objectives
- Dataset
- Methodology and Tools
- Results and Analysis
- *ExpertBayes* (graphical user interface)
- Conclusions and Future Work

ExpertBayes

- Allow the user :
  - Load new Network;
  - Load new data;
  - Load new tables of conditional probabilities;
  - Save the network;
  - Add / Remove vertex;
  - Add / Remove edge;
  - Return edge;
  - Visualize the score, confusion matrix, CPT of an node, precision-recall curve and ROC curve;

- Graphical user interface

Outline

- Objectives
- Dataset
- Methodology and Tools
- Results and Analysis
- *ExpertBayes* (graphical user interface)
- Conclusions and Future Work

Conclusions and Future Work

- ExpertBayes produces better results than the original model and better results than models learned with other tools.

- ExpertBayes also provides a graphical user interface (GUI) where users can play with their models thus exploring new structures that give rise to a search for other models.

Conclusions and Future Work

- Improve the algorithm in order to have better prediction performance.
- Using more (and quality) data, different search and parameter learning methods.

Thank you!

ezildacv@gmail.com
pedroferreira@dcc.fc.up.pt
tiago.vinhoza@gmail.com
ines@dcc.fc.up.pt
pauloraborges@gmail.com
eburnside@uwhealth.org

CRACS INESCTEC
FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO
U.PORTO
WISCONSIN UNIVERSITY OF WISCONSIN-MADISON
FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Appendices

State of the Art

- Previous works considered as initial network a naive Bayes or empty network [9], [4]:

  - [9] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: **The weka data mining software: an update**. SIGKDD Explor. Newsl. 11, 10–18 (Nov. 2009), 1656274.1656278

  - [4] Chan, H., Darwiche, A.: **Sensitivity analysis in bayesian networks: From single to multiple parameters.** In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence. pp. 67–75. UAI ’04, AUAI Press, Arlington, Virginia, United States (2004),id=1036843.1036852

State of the Art

- The R packages deal [2] and bnlearn [11], [13] can refine any input network. However, deal and bnlearn refine input networks by successive refinements instead of performing the refinement only over the original network:

  - [2] Bottcher, S.G., Dethlefsen, C.: Deal: **A package for learning bayesian networks**. Journal of Statistical Software 8, 200–3 (2003)
  - [11] Nagarajan, R., Scutari, M., Lebre, S.: **Bayesian Networks in R with Applications in Systems Biology.** Springer, New York (2013), iSBN 978-1461464457
  - [13] Scutari, M.: **Learning bayesian networks with the bnlearn R package**. Journal of Statistical Software 35(3), 1–22 (2010), http://www.jstatsoft.org/v35/i03/

State of the Art

- WEKA, whose bayesian algorithms apply successive refinements to the newly built models:
  - [6] Cooper, G.F., Herskovits, E.: **A bayesian method for the induction of probabilistic networks from data**. Machine Learning 9(4), 309–347 (1992), BF00994110
  - [8] Friedman, N., Geiger, D., Goldszmidt, M.: **Bayesian network classifiers. In: Machine Learning**. vol. 29, pp. 131–163 (1997)

Methodology

WEKA :

- K2 is a greedy algorithm that, given an upper bound to the number of parents for a node, tries to find a set of parents that maximizes the likelihood of the class variable [6].

- TAN (Tree Augmented Naive Bayes) generates a tree over naive Bayes structure, where each node has at most two parents, being one of them the class variable [8].

Data Distribution

Dataset	Number of Instances	Number of Variables	Pos.	Neg.
Prostate Cancer	496	11	352	144
Breast Cancer (1)	100	34	55	45
Breast Cancer (2)	241	8	88	153

The pseudo-code for ExpertBayes

```
Data:
OriginalNet, // initial network structure;
Train // training set;
Test // test set
Result:
scoreTrain // scores in the training set for BestNet
scoreTest // scores in the test set for BestNet
BestNet // best scored network on Train

Read OriginalNet;
Read Train and Test sets;
BestNet = OriginalNet;
Learn parameters for OriginalNet from training set;
repeat
    Randomly choose a pair of nodes N1 and N2;
    if there exists an edge between N1 and N2 then
        randomly choose: revert or remove
    else
        choose add operation;
        randomly choose edge direction
    end
    Apply operation to OriginalNet obtaining NewNet;
    Rebuild necessary CPT entries, if necessary;
    Compute scoreTrain of the NewNet;
    if scoreTrain NewNet > scoreTrain BestNet then
        BestNet = NewNet
    end
until N iterations using OriginalNet and Train;
Apply BestNet to Test and compute scoreTest;
```

CRACS INESCTEC

ExpertBayes Advantages

- Reduces the computational costs;
- Embed knowledge of an expert in the newly built network;
- Allows the construction of fresh new networks, through its graphical interface.