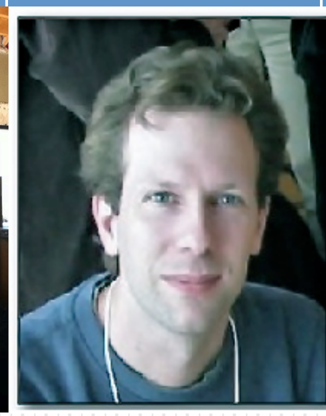
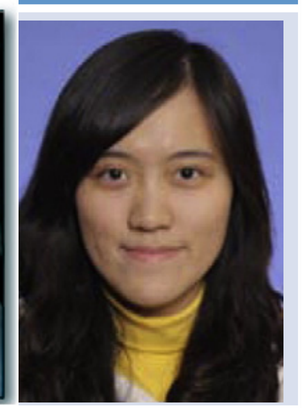
School of Computing Science
Simon Fraser University
Vancouver, Canada

A Hierarchy of Independence Assumptions for Multi-Relational Bayes Net Classifiers

Oliver Schulte | Bahareh Bina | Branden Crawford | Derek Bingham | Yi Xiong



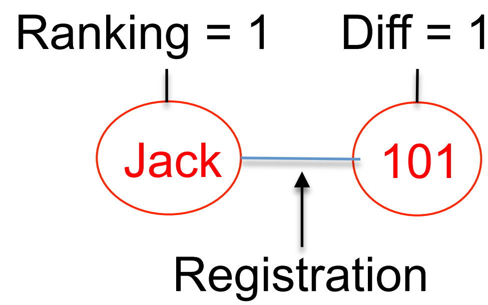




Outline

- Multi-Relational Classifiers
- Multi-Relational Independence Assumptions
- Classification Formulas
- Bayes Nets
- Evaluation

Database Tables

- Tables for Entities, Relationships
- Can visualize as network



Course		
c-id	Rating	Difficulty
101	3	1
102	2	2

Student		
s-id	Intelligence	Ranking
Jack	???	1
Kim	2	1
Paul	1	2

Professor		
p-id	Popularity	Teaching-a
Oliver	3	1
Jim	2	1

Registration			
s-id	**c.id**	**Grade**	**Satisfaction**
Jack	101	A	1
Jack	102	B	2
Kim	102	A	1
Paul	101	B	1

Link-based Classification

Target table: Student

Target entity: Jack

Target attribute (class): Intelligence

Extended Database Tables

Registration			
s-id	**c.id**	**Grade**	**Satisfaction**
Jack	101	A	1
Jack	102	B	2
Kim	102	A	1
Paul	101	B	1

Student		
s-id	Intelligence	Ranking
Jack	???	1
Kim	2	1
Paul	1	2

Course		
c-id	Rating	Difficulty
101	3	1
102	2	2

s-id	c.id	Grade	Satisfaction	Intelligence	Ranking	Rating	Difficulty
Jack	101	A	1	???	1	3	1
Jack	102	B	2	???	1	2	2
Kim	102	A	1	2	1	2	2
Paul	101	B	1	1	2	3	1

Multi-Relational Classifiers

Count relational features

Aggregate relational features

Log-Linear Models
Example:
1. use number of A,s number of Bs,…
2. $\ln(P(class)) = \Sigma\, x_i\, w_i - Z$

Disadvantage: slow learning

Propositionalization
Example: use average grade
Disadvantages:
• loses information
• slow to learn (up to several CPU days)

+ Independence Assumptions

Log-Linear Models With Independencies
+ Fast to learn
−Independence Assumptions may be only approximately true

Independence Assumptions

Independence Assumptions: Naïve Bayes

s-id	c.id	Grade	Satisfaction	Intelligence	Ranking	Rating	Difficulty
Jack	101	A	1	???	1	3	1
Jack	102	B	2	???	1	2	2
Kim	102	A	1	2	1	2	2
Paul	101	B	1	1	2	3	1

Naive Bayes:
non-class attributes are independent of each other, given the target class label.

Legend: Given the blue information, the yellow columns are independent.

Path Independence

s-id	c.id	Grade	Satisfaction	Intelligence	Ranking	Rating	Difficulty
Jack	101	A	1	???	1	3	1
Jack	102	B	2	???	1	2	2
Kim	102	A	1	2	1	2	2
Paul	101	B	1	1	2	3	1

Naive Bayes:
non-class attributes are independent of each other, given the target class label.

Path Independence:
Links/paths are independent of each other, given the attributes of the linked entities.

Legend: Given the blue information, the yellow rows are independent.

Influence Independence

s-id	c.id	Grade	Satisfaction	Intelligence	Ranking	Rating	Difficulty
Jack	101	A	1	???	1	3	1
Jack	102	B	2	???	1	2	2
Kim	102	A	1	2	1	2	2
Paul	101	B	1	1	2	3	1

Legend: Given the blue information,
the yellow columns
are independent from
the orange columns

Naive Bayes:
non-class attributes are independent of each other, given the target class label.

Path Independence:
Links/paths are independent of each other, given the attributes of the linked entities.

Influence Independence:
Attributes of the target entity are independent of attributes of related entities, given the target class label.

Path-Class Independence:
the existence of a link/path is independent of the class label.

Classification Formulas

- Can *rigorously derive log-linear prediction formulas from independence assumptions*.
- Path Independence:
  predict max class for: log(P(class | target attributes)) +
  sum over each table, each row:
  [log(P(class | information in row)) – log(P(class | target atts))]
- PI + Influence Independence:
  predict max class for: log(P(class | target attributes)) +
  sum over each table, each row:
  [log(P(class | information in row)) – log(prior P(class))]

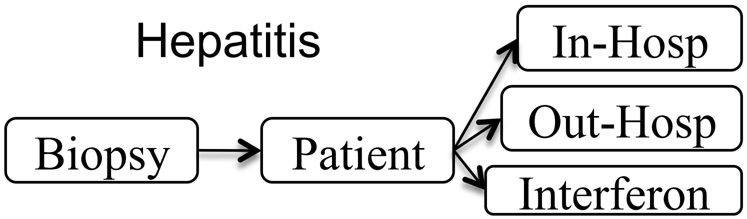
Relationship to Previous Formulas

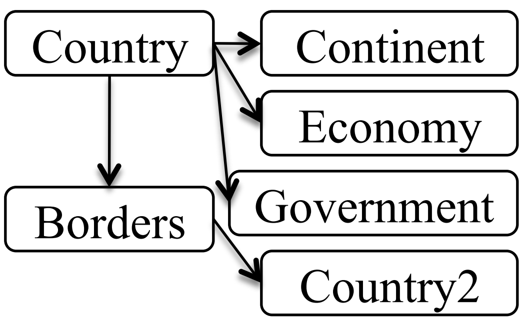
Assumption	Previous Work with Classification Formula
Path Independence	none; our new model.
PI + Influence Independence	Heterogeneous Naive Bayes Classifier Manjunath et al. ICPR 2010.
PI + II + Naive Bayes	Exists + Naive Bayes (single relation only) Getoor, Segal, Taskar, Koller 2001
PI + II + NB + Path-Class	Multi-relational Bayesian Classifier Chen, Han et al. Decision Support Systems 2009

Evaluation

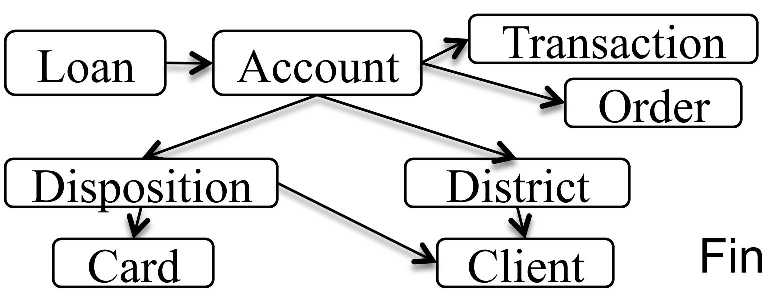
Data Sets and Base Classifier

Hepatitis





Mondial



Financial

- **Standard Databases** KDD Cup, UC Irvine
- MovieLens not shown.

Classifier

- Can plug in any single-table probabilistic *base classifier* with classification formula .
- We use **Bayes nets**.

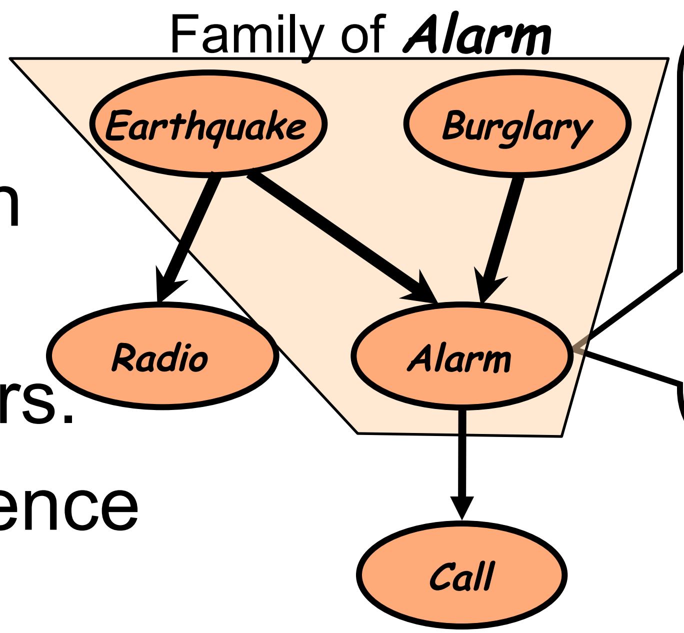
What is a Bayes net?

Compact representation of joint probability distributions via conditional independence

Qualitative part:

Directed acyclic graph (DAG)

- Nodes - random vars.
- Edges - direct influence



E	B	$P(A \mid E,B)$	
e	b	0.9	0.1
e	$\bar{b}$	0.2	0.8
$\bar{e}$	b	0.9	0.1
$\bar{e}$	$\bar{b}$	0.01	0.99

Together:
Define a unique distribution in a factored form

Quantitative part:
Set of conditional probability distributions

$$P(B,E,A,C,R) = P(B)P(E)P(A \mid B,E)P(R \mid E)P(C \mid A)$$

Figure from N. Friedman

Independence-Based Learning is Fast

weakest assumption

strongest assumption

	Bayes Net Classifiers				Other Methods	
Dataset	PIC	HNBC	E-NB	MRNBC	MLN	Tilde
Hepatitis	7.43	7.01	**2.07**	**2.07**	3902	853
Financial	28.31	23.21	**15.01**	**15.01**	NT	2429
MovieLens	25.32	17.67	**5.31**	**5.31**	960	1100
Mondial	5.41	5.08	1.89	1.89	5.44	**0.3**

Training Time in seconds

Independence-Based Models are Accurate

weakest assumption → strongest assumption

Accuracy	Bayes Net Classifiers				Reference Methods	
Dataset	PIC	HNBC	E-NB	MRNBC	MLN	Tilde
Hepatitis	**0.80**	0.78	0.78	0.74	0.77	0.61
Financial	**0.91**	0.90	0.89	0.81	NT	0.89
MovieLens	**0.66**	0.57	0.53	0.50	0.484	0.48
Mondial	**0.85**	0.82	0.78	0.82	0.76	0.71

- Similar results for F-measure, Area Under Curve

Conclusion

- Several plausible independence assumptions/classification formulas investigated in previous work.
  - Organized in *unifying hierarchy*.
- New assumption: <u>multi-relational path independence</u>.
  - *most general*, implicit in other models.
- Big advantage: Fast scalable simple learning.
  - Plug in single-table probabilistic classifier.
- Limitation: no pruning or weighting of different tables. Can use logistic regression to learn weights (Bina, Schulte et al. 2013).

Bina, B.; Schulte, O.; Crawford, B.; Qian, Z. & Xiong, Y.
"Simple decision forests for multi-relational classification", *Decision Support Systems*, **2013**

Thank you!

- Any questions?