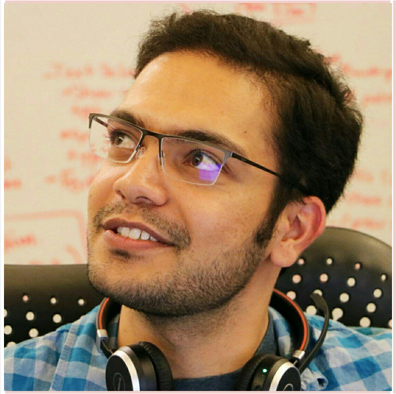
School of Computing Science
Simon Fraser University
Vancouver, Canada

Model Selection Scores for Multi-Relational Bayesian Networks

Sajjad Gholami	Oliver Schulte



Outline

- Brief Summary of Main Paper: Learning first-order Bayesian Networks for Multi-Relational Data
- What is a first-order BN good for?
  - Answering first-order probabilistic queries.
  - Probabilistic inference over links and attributes
  - Extracting Features: Extract, Transform, Load
  - Classification
  - Exception Mining and Anomaly Detection

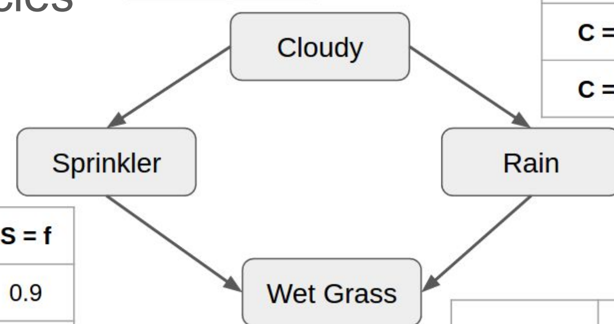
Main Paper Summary

Locally Consistent Bayesian Network Scores for Multi-Relational Data

Bayesian Networks

- Probabilistic Graphical Model
- Nodes: Random variables
- Edges: Conditional dependencies
- Directed Acyclic Graph (DAG)

C = t	C = f
0.5	0.5



	R = t	R = f
C = t	0.8	0.2
C = f	0.2	0.8

	S = t	S = f
C = t	0.1	0.9
C = f	0.5	0.5

	W = t	W = f
S = t, R = t	0.99	0.01
S = t, R = f	0.9	0.1
S = f, R = t	0.9	0.1
S = f, R = f	0	0

Multi-Relational Data

- k-ary functor: maps a tuple of k individuals to a value
- I.I.D. data: unary (k=1) functors
- Multi-Relational: k>1

User

User_id	Age	Gender
3	0	M
5	1	F
7	2	M
...		

Rating

User_id	Movie_id	Rating
3	The Dictator	1
5	Thor	4
5	The Dictator	3
7	BraveHeart	5
...		

Movie

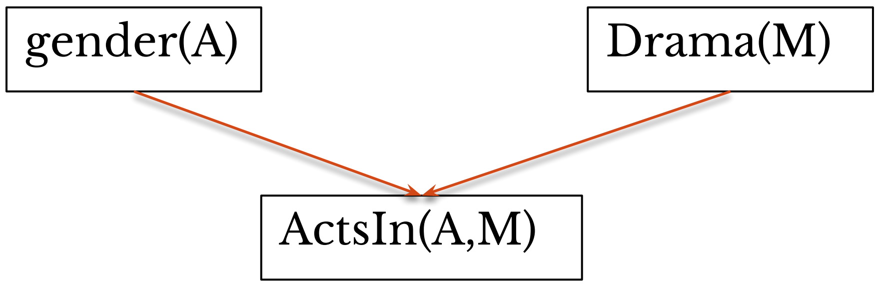
Movie_id	Action	Drama	Horror
The Dictator	0	0	0
Thor	1	0	0
BraveHeart	1	1	1
...			

First-Order Bayesian Networks

- A first-order Bayesian network is a Bayesian network whose nodes are first-order terms

  (Wang et al. 2008)

- AKA parametrized Bayesian network
  (Poole 2003, Kimmig et al. 2014)



First-Order Bayesian Networks represent a Database Distribution

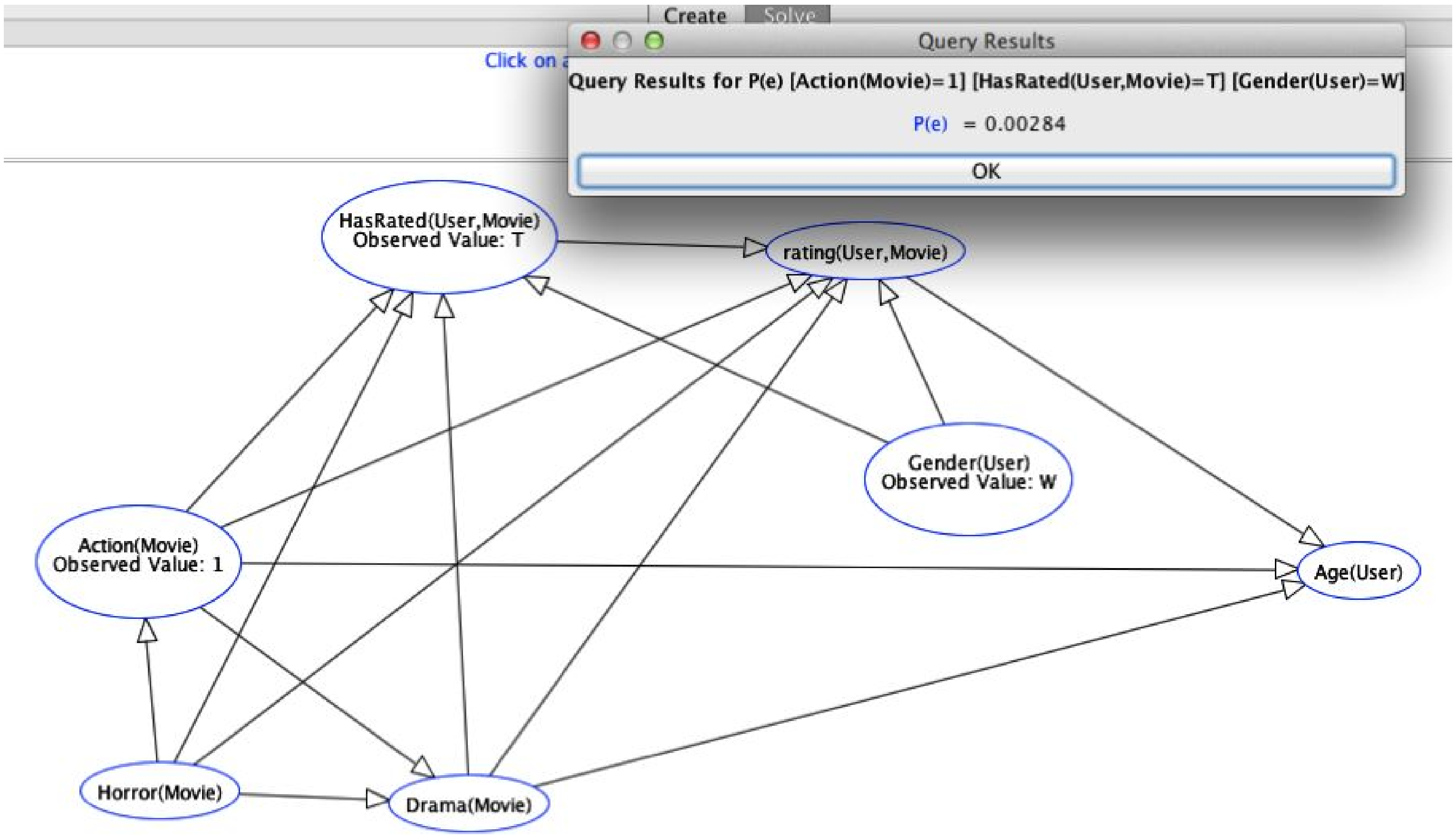
Wang, D. Z.; Michelakis, E.; Garofalakis, M. & Hellerstein, J. M. (2008), BayesStore: managing large, uncertain data repositories with probabilistic graphical models, in , VLDB Endowment, , pp. 340--351.
Kimmig, A.; Mihalkova, L. & Getoor, L. (2014), 'Lifted graphical models: a survey', *Machine Learning, 1--45.*

Probabilistic First-Order Queries

- Database probability of a first-order formula = #satisfying instantiations/#possible instantiations

- Examples:
  - $P_D(gender(Actor) = W) = 0.5$
  - $P_D(gender(Actor) = W, ActsIn(Actor,Movie) = T) = 1/4$

DEMO: Bayes Net Query



http://www.aispace.org/

Data Query

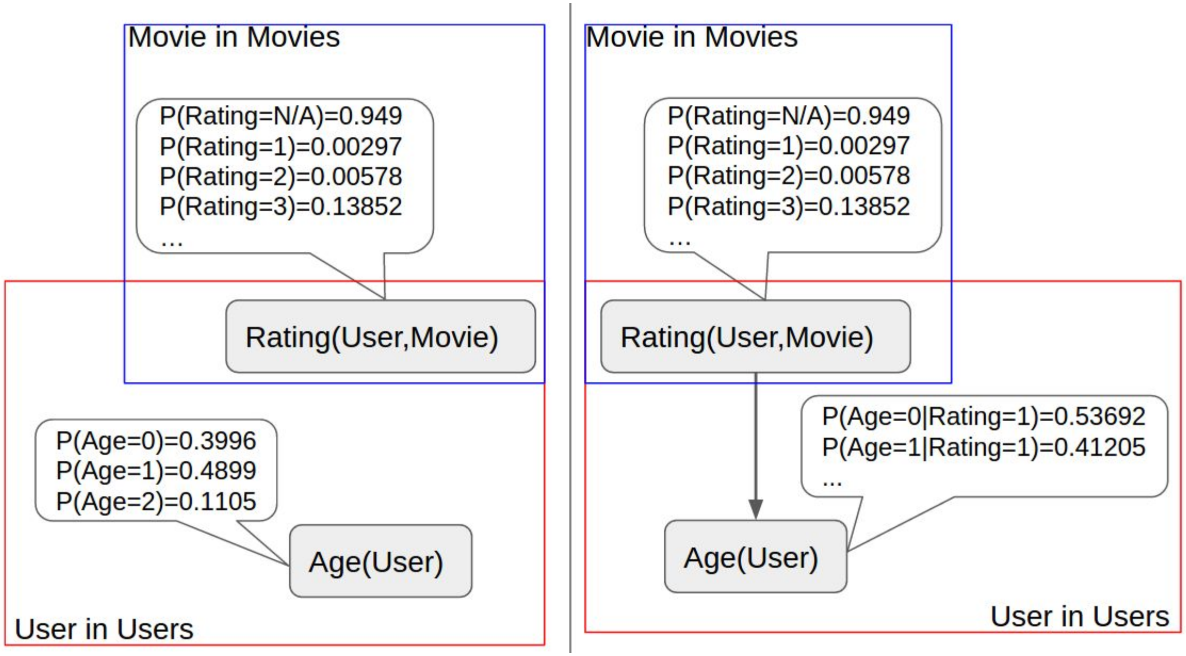
Num Movies	**3883**
Num Users	6039
Num Movie-User Pairs	3883 x 6039 = 23449437

movie-user pairs with action movie, woman user

Action(Movie) =T, HasRated(User,Movie) =T, gender(User) =W	**66642**
Frequency	66642/23449437= 0.0028

First-Order Bayesian Networks

Nodes = First-order terms

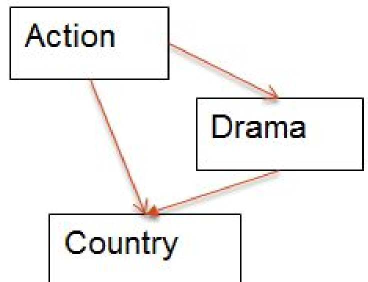


Upgrading Bayesian Network Scores

data table

Title	Drama	Action	Country
Fargo	T	T	U.S.
Kill_Bill	F	T	U.S.

Bayes net graph

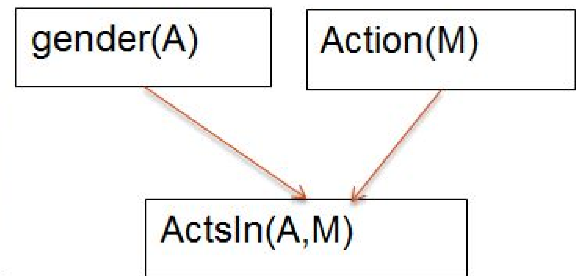


BN Score for data table, e.g. -3.5

score upgrade

relational database

Parametrized Bayes net graph



BN Score for database, e.g. -3.5

Upgrading Bayesian Network Scores

Desired Properties:

- **Generalization**:
  - Model comparison for i.i.d special case of the upgraded one
- **Preserving Local Consistency**:
  - Model comparison consistent for i.i.d. then consistent for multi-relational data.

Statistical Consistency

- We provide a general method for generalizing non-relational model selection scores to multi-relational scores

- A model selection score is **consistent** if BNs that maximize the score are guaranteed to represent the database distribution if the domain sizes are arbitrarily large.

Intuitions

- The main issue: Event counts depend on the data **and** on the networks considered.
- Solution:
  1. Rescale counts to be compatible for current and alternative graphs.
  2. Normalize scores to be compatible across alternative graphs.

Bayesian Network Scores

Standard scores have this form:

Score = LL - Penalty

LL: Log Likelihood of data under model

Penalty: A function of number of parameters and sample size

- Log-Likelihood:
  - Sum over all parent-child value assignments.
- AIC: LL - f(#Parameters)
- BIC: LL - f(#Parameters) x log(SampleSize)

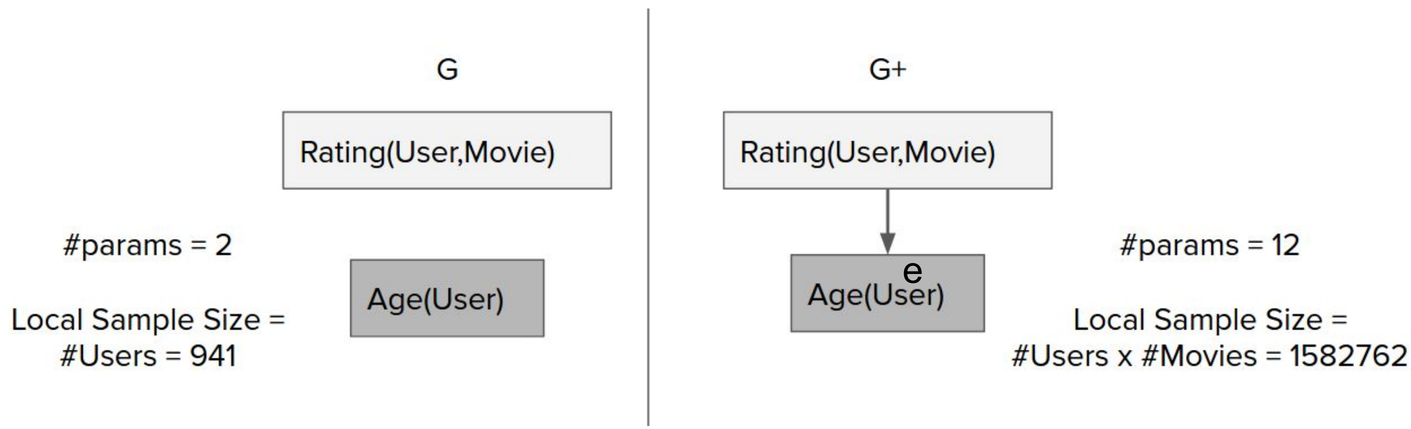
Normalized Gain Upgrade Method

To score an edge e:

A = Normalized Likelihood differential

B = Normalized penalty term differential by the **larger** local sample size

Normalized gain = A – B



Applications of First-Order Bayesian Networks

Tutorial:
https://oschulte.github.io/srl-tutorial-slides/

Github:
https://github.com/sfu-cl-lab/FactorBase

Example: IMDB toy data

ETL: Feature Vectors

$$\begin{tikzcd} \text{ActsIn(A,M)} \arrow[dr] & & \text{Drama(M)} \arrow[dl] \\ & \text{gender(A)} & \end{tikzcd}$$

- Learned Bayesian network →family conjunctions
- Feature = proportion of family conjunction in data
- Works well for classification and anomaly detection

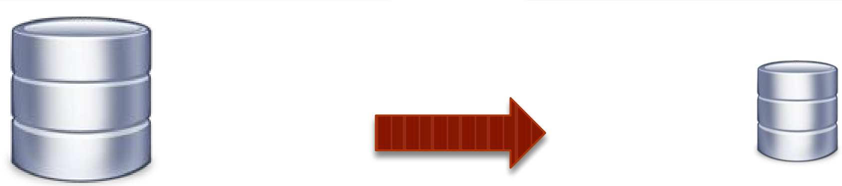
gender(A)	ActsIn(A,M)	Drama(M)				
M	T	T	0	0	1/2	0
M	T	F	0	0	0	0
M	F	T	1/2	0	0	0
M	F	F	1/2	0	1/2	0
W	T	T	0	0	0	0
W	T	F	0	1/2	0	1/2
W	F	T	0	1/2	0	1/2
W	F	F	0	0	0	0

Native Relational Applications

- Via ETL, a Bayesian network can leverage traditional machine learning classification and anomaly/exception mining.
- Purely relational methods work even better (see refs).
- For exception mining, compare Kullback-Leibler divergence (KLD) between data distribution for a *random average* individual and data distribution for a specific individual.

Population Database
e.g. IMDB

Individual Database
e.g. Brad Pitt's movies



Schulte, O.; Qian, Z.; Kirkpatrick, A. E.; Yin, X. & Sun, Y. (2016), 'Fast learning of relational dependency networks', Machine Learning, 1--30.
Riahi, F. & Schulte, O. (2015), Model-Based Outlier Detection for Object-Relational Data, *in 'Computational Intelligence, 2015 IEEE Symposium Series on', pp. 1590--1598.*

Case Study: Movies

- Brave Heart is the most unusual drama
- Its most unusual feature is Actor Quality
  - 93% of Brave Heart's actors have the highest actor quality
  - A random average drama has only 42% actors of highest quality

MovieTitle	Genre	KLD Rank	KLD Max Node	KLD Max feature Value	Individual Probability	Class Probability
Brave Heart	Drama	1	Actor_Quality	a_quality=4	0.93	0.42
Austin Powers	Comedy	2	Cast_position	cast_num=3	0.78	0.49
Blue Brothers	Comedy	3	Cast_position	cast_num=3	0.88	0.49

Case Study: Soccer Strikers

Player Name	Position	KLD Rank	KLD Max Node	KLD Max Value	Individual Probability	Class Probability
Edin Dzeko	Striker	1	Dribble Efficiency	DE = Low	0.16	0.50
Paul Robinson	Goalie	2	SavesMade	SM = Medium	0.30	0.04
Michel Vorm	Goalie	3	SavesMade	SM = Medium	0.37	0.04

Conclusions

- Relational Bayes net learning has a solid theoretical foundation.
- Scales to millions of records.
- Supports various applications, e.g.
  - database frequency estimation
  - joint inference over attributes and relationships
  - link-based classification
  - exception mining

Thank you!

Tutorial:
https://oschulte.github.io/srl-tutorial-slides/

Github:
https://github.com/sfu-cl-lab/FactorBase

Formulas

- A (conjunctive) formula is a **joint assignment** $term_1 = value_1, \ldots, term_n = value_n$
  - e.g., ActsIn(Actor, Movie) = T, gender(Actor) = W

- A *ground* formula contains only constants
  - e.g., ActsIn(UmaThurman, KillBill) = T, gender(UmaThurman) = W

Learned Bayes Net for IMDb