School of Computing Science
Simon Fraser University
Vancouver, Canada

SFU

Model-based Exception Mining for Relational Data

Oliver Schulte | Fatemeh Riahi

- Extended from 2015 IEEE Symposium Series on Computational Intelligence
- Best Student Paper Award

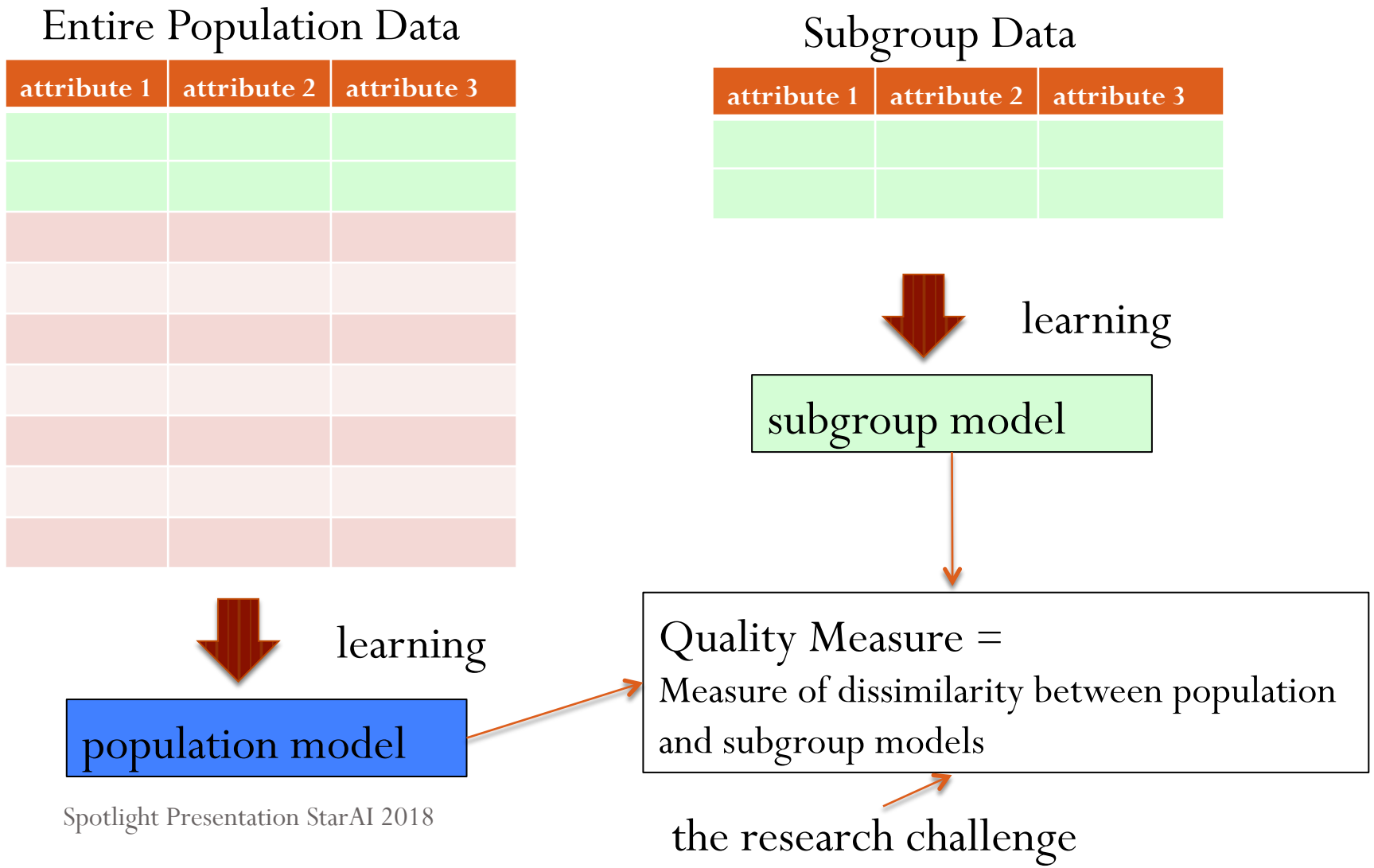
Relational Exception Mining

- The task: *identify exceptional individuals* (entities, nodes, objects) in relational data
- Approach: apply the Exceptional Model Mining (EMM) framework (Duivesteijn, **Knobbe** et al. 2016)

Model/parameter learning method → EMM → exception mining method

Duivesteijn, W.; Feelders, A. J. & Knobbe, A. (2016), 'Exceptional model mining', *Data Mining and Knowledge Discovery* ***30(1), 47—98.***

Exceptional Model Mining: I.I.D Single-Table Data



EMM: Multi-relational Data

Entire Observed Network

relational learning

Subnetwork Centered on Individual
aka egonet, interpretation

learning

individual SRL model

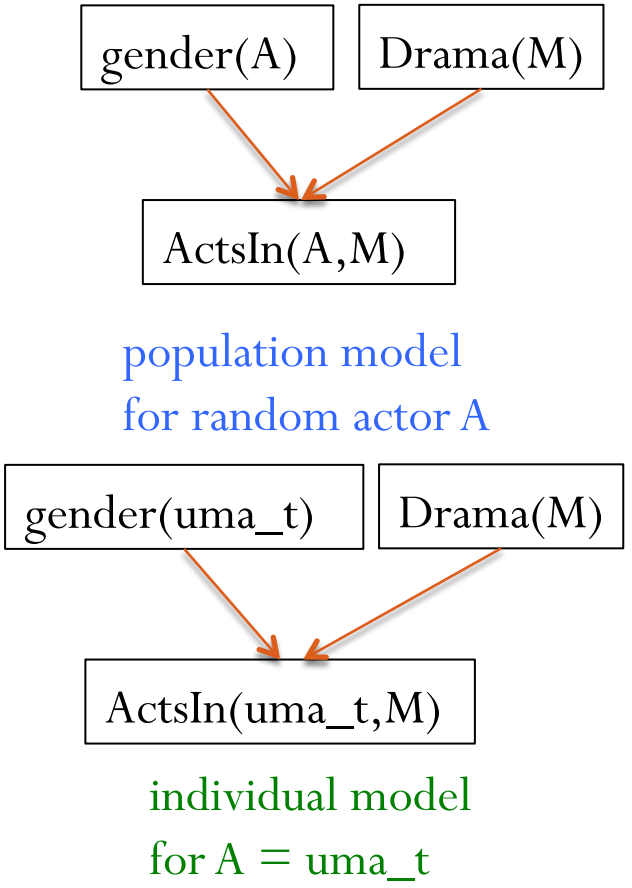
statistical-relational population model

Outlierness Metric (quality measure)
= Measure of dissimilarity between population and individual SRL models

we investigate several metrics based on log-linear likelihood functions

Model Type and Outlierness Metrics

- We use first-order Bayesian networks with a log-linear likelihood function (Wang et al. 2008, Schulte 2011).
- Outlierness metrics = variants of Kullback-Leibler divergence (some are novel)
- log-linear + KLD ➔ total outlier score = sum of feature-wise differences
- works for other log-linear models, e.g. Markov Logic Networks



Wang, D. Z.; Michelakis, E.; Garofalakis, M. & Hellerstein, J. M. (2008), BayesStore: managing large, uncertain data repositories with probabilistic graphical models, *in 'Proceedings VLDB', VLDB Endowment, , pp. 340—351.*
Schulte, O. (2011), A tractable pseudo-likelihood function for Bayes Nets applied to relational data, *in 'SIAM SDM', pp. 462-473.*

Case Study: Strikers and Movies

Player Name	Position	ELD Rank	ELD Max Node	Max Value	Individual Probability	Ref. Class Probability
Edin Dzeko	Striker	1	Dribble Efficiency	DE = Low	0.16	0.5
Paul Robinson	Goalie	2	SavesMade	SM = Medium	0.3	0.04
Michel Vorm	Goalie	3	SavesMade	SM = Medium	0.37	0.04

MovieTitle	Genre	ELD Rank	ELD Max Node	Max Value	Individual Probability	Ref. 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