Aggregation and Concept Classes in Relational Modeling

Claudia Perlich & Foster Provost
New York University

SIGKDD, August 2003

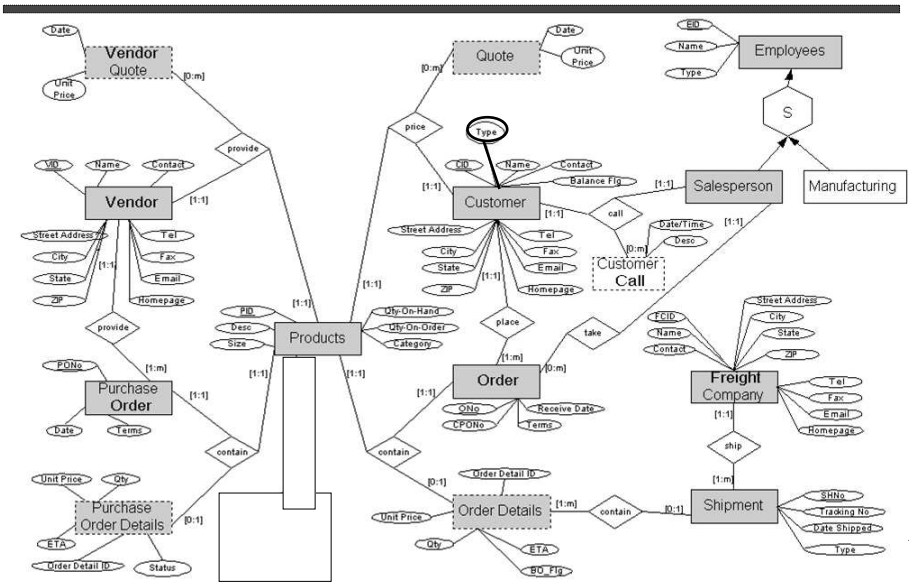
NYU STERN

Data – Data Mining View

```
b,41.50,1.54,u,g,i,bb,3.5,f,f,0,f,g,00216,0,0
b,23.92,0.665,u,g,c,v,0.165,f,f,0,f,g,00100,0,1
a,25.75,0.5,u,g,c,h,0.875,t,f,0,t,g,00491,0,1
b,26.00,1,u,g,q,v,1.75,t,f,0,t,g,00280,0,0
b,37.42,2.04,u,g,w,v,0.04,t,f,0,t,g,00400,5800,0
b,34.92,2.5,u,g,w,v,0,t,f,0,t,g,00239,200,0
b,34.25,3,u,g,cc,h,7.415,t,f,0,t,g,00000,0,1
b,23.33,11.625,y,p,w,v,0.835,t,f,0,t,g,00160,300,0
b,23.17,0,u,g,cc,v,0.085,t,f,0,f,g,00000,0,1
b,44.33,0.5,u,g,i,h,5,t,f,0,t,g,00320,0,0
b,35.17,4.5,u,g,x,h,5.75,f,f,0,t,s,00711,0,0
b,43.25,3,u,g,q,h,6,t,t,11,f,g,00080,0,1
b,56.75,12.25,u,g,m,v,1.25,t,t,04,t,g,00200,0,0
b,31.67,16.165,u,g,d,v,3,t,t,09,f,g,00250,730,1
a,23.42,0.79,y,p,q,v,1.5,t,t,02,t,g,00080,400,1
a,20.42,0.835,u,g,q,v,1.585,t,t,01,f,g,00000,0,1
b,26.67,4.25,u,g,cc,v,4.29,t,t,01,t,g,00120,0,0
b,34.17,1.54,u,g,cc,v,1.54,t,t,01,t,g,00520,50000,1
a,36.00,1,u,g,c,v,2,t,t,11,f,g,00000,456,1
b,25.50,0.375,u,g,m,v,0.25,t,t,03,f,g,00260,15108,1
b,19.42,6.5,u,g,w,h,1.46,t,t,07,f,g,00080,2954,0
b,35.17,25.125,u,g,x,h,1.625,t,t,01,t,g,00515,500,1
```

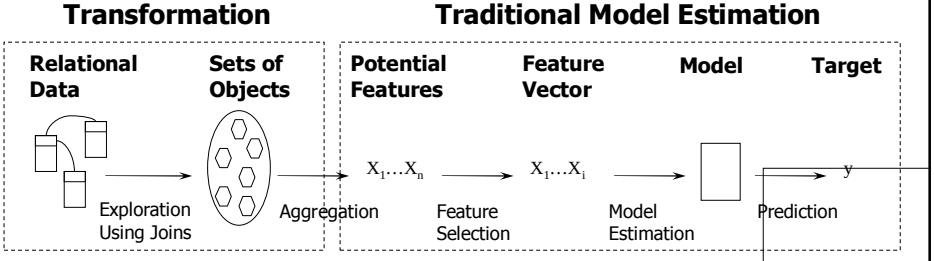
NYU STERN

Data – Application View

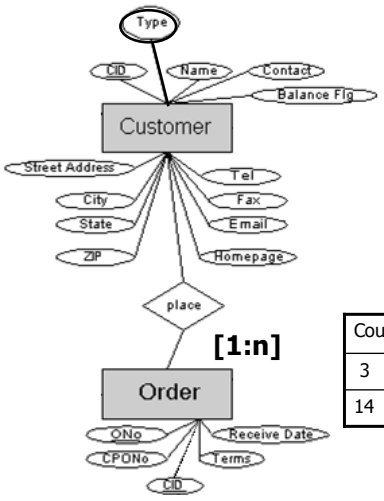


Current Approaches

- Logic-Based Approaches (e.g., ILP)
  - Customer(X) AND Transaction(X,Y,Z) AND Z="book"
- Transformation-Based Relational Learning
  - 'Upgrades' of existing learners using simple aggregates (counts, mode, min, max, average) for feature construction



Transformation Challenge



Customer

CID	Type
Smith	1
Peters	0

Order

CID	P
Smith	
Smith	
Smith	

+

Count	Sum	Mode	...	Type
3	30.87	CD		1
14	171.67	Book		0

Order

CID	Price	Product
Peters	11.99	CD
Peters	21.99	CD
Peters	25.99	DVD
Peters	4.99	Book
Peters	18.99	DVD
Peters	13.99	Book
Peters	6.99	CD
Peters	9.99	Book
Peters	12.89	Book
Peters	17.99	DVD
Peters	8.99	Book
Peters	14.99	Book
Peters	8.99	CD
Peters	12.89	Book

Towards a Theory of Aggregation: Assumptions

Hierarchy of Concept Complexity

High

Complexity

Low

5 Global dependencies

'High Reputation in Social Networks'

4 Dependencies across sets of related objects

'Average price of retuned items'

3 Dependence of attributes

'Average book price'

2 Independence of attributes

'Most common product type'

1 Propositional (1-1 or n-1 relations)

'Average Household Income'

NYU STERN

Novel Approach to Aggregation

ObjectID	*Class*
ID 1	0
ID 2	1
ID 3	0
...	...

ObjectID	*Price*	*Type*
ID 1	5	Book
ID 1	10	Book
ID 1	14	CD
ID 2	49	Book
ID 2	18	DVD
ID 3	12	CD
...	...	...

(5, Book)
(10, Book)
(14, CD)
(12, CD)
...
$\sim D_{\text{Class 0}}$

(49, Book)
(18,DVD)
...
$\sim D_{\text{Class 1}}$

1. Sample statistics (e.g., count) are different
2. Probability densities of related objects are different

Aggregation ≡ ~~Density Estimation~~

Categorical Density Estimation

ObjectID	*Class*
ID1	0
ID2	1
ID3	0
...	...

ObjectID	*Type*
ID1	Book
ID1	Book
ID1	CD
ID2	Book
ID2	DVD
ID3	CD

(Book)
(Book)
(CD)
(CD)
...
$\sim D_{\text{Class 0}}$

(Book)
(DVD)
...
$\sim D_{\text{Class 1}}$

1. Reference Vectors (RV): Estimates of $D_{\text{Class 0}}$ and $D_{\text{Class 1}}$
2. Case Vectors (CV): Estimates of the density of objects related to specific cases ID1, ID2, ID3, ...
3. Vector Distances (d): L1, L2, cosine, Mahalanobis between case and reference vectors as features

NYU STERN

Example: Density Estimation

Id1	0
Id2	1
Id3	1
Id4	0

Id1	B
Id2	A
Id2	A
Id2	B
Id3	A
Id4	B
Id4	B
Id4	B
Id4	A

1: Reference Vectors:

RV	**A**	**B**
$RV_{Class\ 1}$	0.75	0.25
$RV_{Class\ 0}$	0.2	0.8

2: Case Vectors:

CV	**A**	**B**
ID1	0	1
ID2	0.66	0.33
ID3	1	0
ID4	0.25	0.27

3: L2 Distances for ID2:

$L2(ID2, RV_{Class\ 1}) = 0.12$

$L2(ID2, RV_{Class\ 0}) = 0.65$

NYU STERN

Additional Features

Class-Conditional Densities Estimates:

$RV_{Class\ 0}$	
Book	0.21
DVD	0.28
CD	0.36
VHS	0.09
VCR	0.06

$RV_{Class\ 1}$	
Book	0.01
DVD	0.33
CD	0.31
VHS	0.22
VCR	0.13

Counts for particular categorical Values:

- Most common (MC): CD
- Most common class 1 (MC1): DVD
- Most common class 0 (MC0): CD
- Most discriminative (MD): Book

NYU STERN

Summary: Feature Complexity

High

Complexity

Low

2.3: Discriminative class-conditional features

- MD, $d(IDn,D_{Class\ 1})$-$d(IDn,D_{Class\ 0})$, ...

2.2: Class-conditional features

- MC1, MC0, $d(IDn,D_{Class\ 1})$, $d(IDn,D_{Class\ 0})$, ...

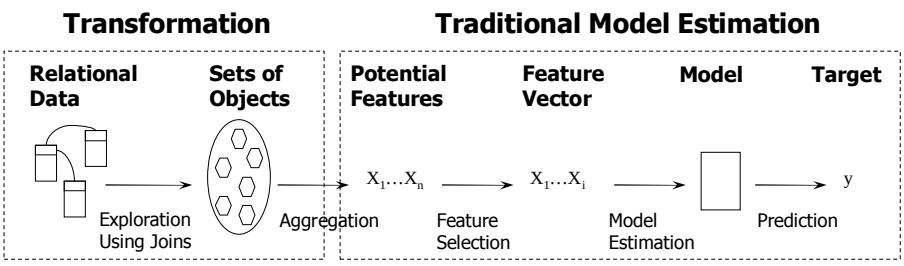
2.1: Unconditional features

- MC, Count
- $\cos(IDn,RV_{All})$, $L2(IDn,RV_{All})$, $Mah(IDn,RV_{All})$, ...

1 : No Relational Features

NYU STERN

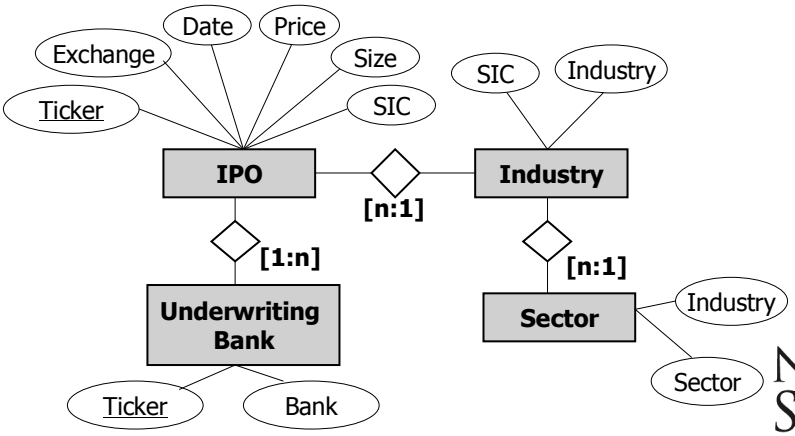
ACORA: Automated Construction of Relational Attributes



- Exploration of relational structure: Breadth-first search
- Aggregation: 50% of the training data
- Feature Selection: 10-fold weighted random sampling
- Model Estimation: Logistic regression and C4.5 on 50%
- Prediction: Average over 10 models on holdout data

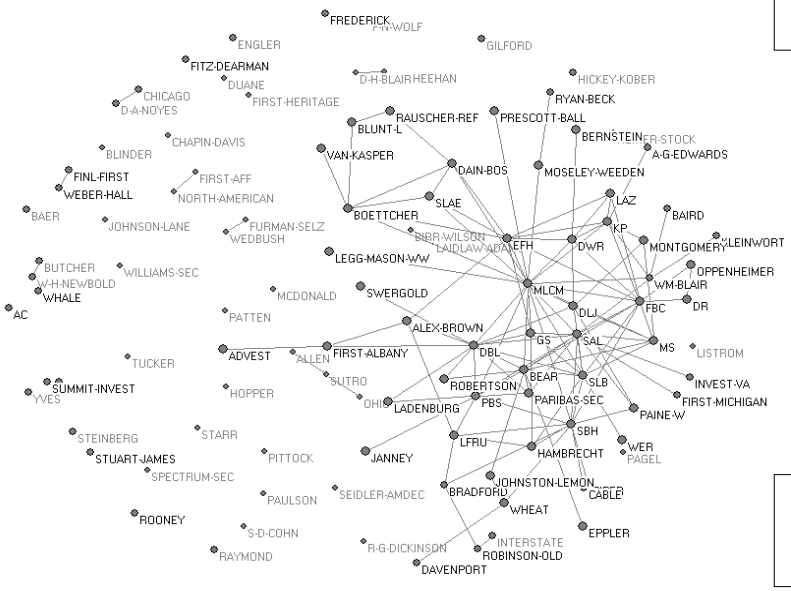
Exchange Identification: IPO Domain

- Task: Was the exchange NASDAQ?
- Data: 2700 IPOs + 3 background relations



NYU STERN

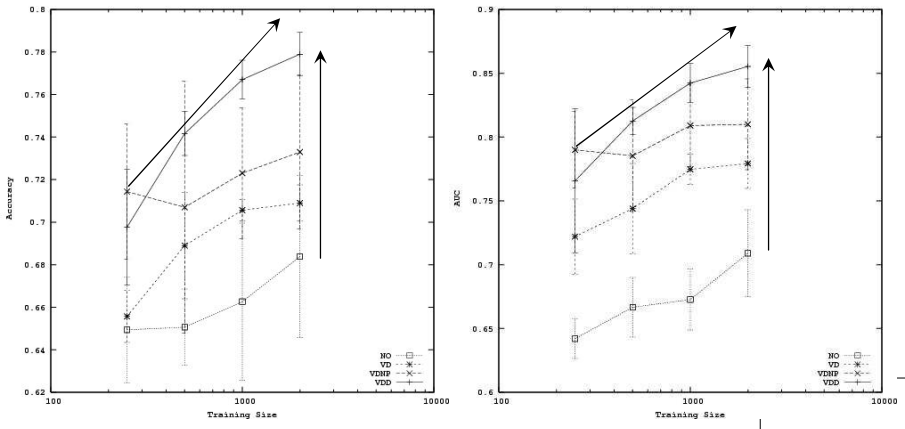
'Social' Network of IPO Underwriters



Aggregation Complexity and Performance

As complexity increases performance increases

As training size increases performance increases



Accuracy AUC

Model Choice for Accuracy: Logistic Regression vs. Decision Trees

Size	Log	C4	Log	C4	Log	C4
250	0.63	0.62	0.70	0.71	0.69	0.70
500	0.64	0.64	0.71	0.71	0.73	0.74
1000	0.65	0.67	0.71	0.71	0.74	0.76
2000	0.65	0.70	0.73	0.73	0.74	0.76
	Unconditional Features		**Conditional Features**		**Discriminative Conditional Features**	

- Aggregation has stronger impact on performance than model type does

NYU STERN

Specific Values vs. Density Distances

Low → Complexity Level → High

	Unconditional Features				Conditional Features			Discriminative Features		
Size	NO	MOC	VD	MVD	MPN	VDPN	MVDPN	MD	VDD	MVDD
250: 6	0.642	0.697	0.717	0.691	0.672	0.748	0.716	0.68	0.729	0.734
250: 9	0.642	0.707	0.711	0.74	0.725	0.756	0.761	0.749	0.75	**0.764**
250:12	0.642	0.729	0.722	*0.755*	0.715	*0.79*	0.74	0.713	0.763	0.76
500: 6	0.666	0.702	0.738	0.741	0.72	0.746	0.739	0.75	0.774	0.79
500: 9	0.666	0.775	0.753	0.757	0.758	0.77	*0.802*	0.796	0.775	**0.821**
500:12	0.666	0.741	0.744	*0.787*	0.775	0.785	0.76	0.792	0.812	0.812
1000: 6	0.672	0.743	0.754	0.749	0.735	0.793	0.797	0.767	0.788	0.802
1000: 9	0.672	0.765	0.768	0.763	0.787	0.808	*0.825*	0.797	0.818	0.826
1000:12	0.672	0.778	0.774	*0.781*	0.78	0.809	0.797	0.793	**0.842**	0.829
2000: 6	0.709	0.727	0.744	0.752	0.732	0.795	0.796	0.787	0.794	0.824
2000: 9	0.709	0.785	0.772	0.781	0.807	0.805	*0.835*	0.799	0.832	0.838
2000:12	0.709	0.791	0.779	*0.801*	0.79	0.81	0.788	0.798	**0.855**	0.836

Comparison to Logic-Based Approaches

Size	ACORA	FOIL	Tilde	Lime	Progol	Bin
250	0.703	0.645	0.646	0.568	0.594	0.59
500	0.741	0.664	0.628	0.563	0.558	0.643
1000	0.767	0.658	0.63	0.53	0.53	0.638
2000	0.779	0.671	0.65	0.51	0.541	0.641

- Logic Setup: declarative bias and look-ahead
- ACORA outperforms by a large margin all 5 logic-based methods for classification

NYU STERN

Comparison to Non-Logic Approaches

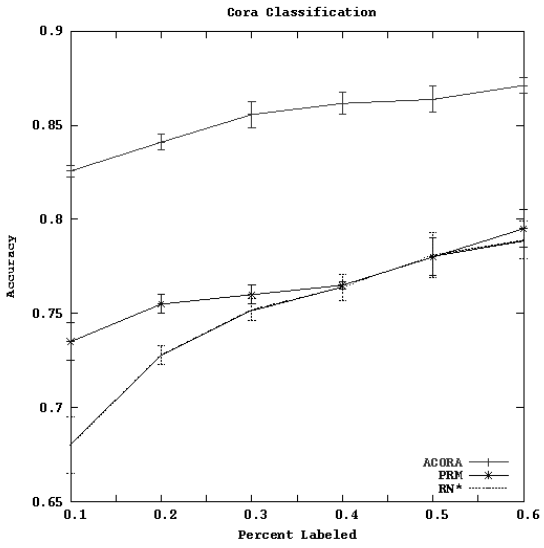
- Domain: Cora Database
  -Publications in 7 Categories in Machine Learning

- Tables:
  - Papers (PID, Category)
  - Authors (PID, Author)
  - Citations(PID, PID)

- Comparison:
  - ACORA with discriminative features
  - Relational Neighbor Classifier
    - (Macskassy and Provost MRDM Workshop Kdd 2003)
  - Probabilistic Relational Models
    - (Koller and Pfeffer 1998)
    - Results from Taskar et al. IJCAI 2001

NYU STERN

Document Classification Results



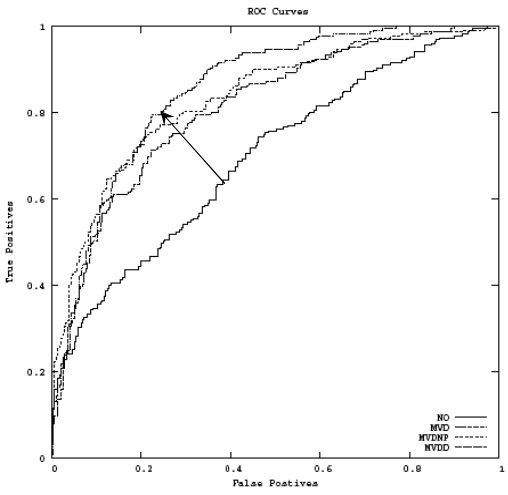
NYU STERN

Summary

- Working toward a theory of aggregation
- Examining relationship between expressive power of models and aggregation assumptions
- Increased aggregation complexity has strong positive impact on generalization performance
- Superior performance using complex, independent aggregation (density estimation) compared to alternative relational models (ILP, PRM)

NYU STERN

Lift in ROC



NYU STERN