Aggregation Based Feature Invention and Relational Concept Classes

(Claudia Perlich & Foster Provost)

Relational Learning

- Expressive
- Background Knowledge can be incorporated easily
- Aggregation

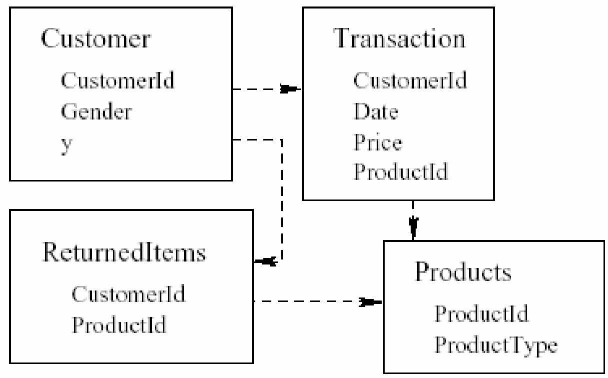


Figure 1: Transaction database

Predictive Relational Learning

- M: (t, RDB) $\longrightarrow$ y

$$y = \varphi(t, \psi(RDB)) + \varepsilon$$

- Complexity of relational concept
  1. Complexity of relationships
  2. Complexity of Aggregation Function
  3. Complexity of the function

Relational Concept Classes

- Propositional
  - Features can be concatenated
  - No aggregation
  - Example – One customer table and other demographic table
- Independent Attributes
  - 1 to n relationship requires simple aggregation
  - Mapping from a bag of zero or more attributes to a categorical or numeric value
  - Ex Sum, Average for numeric values
  - Ex Mode for categorical attributes

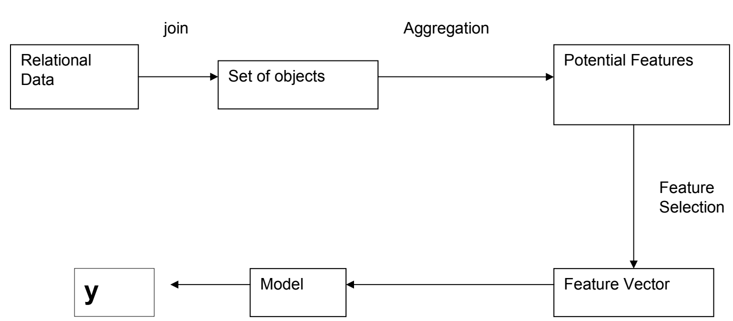
Relational Concept Classes - Contd

- Dependent Attributes within one table
  - Multi-dimensional Aggregation
  - Number of products bought on Dec 22^{nd} (conditioned on Date)
- Dependent Attributes across tables
  - More than one bag of objects of different type
  - Amount spent on items returned at a later date
  - Needs info from more than 1 table
- Global graph features
  - Transitive closure over a set possible joins
  - Customer Reputation

Methods for Relational Aggregation

- First Order Logic - ILP
- Simple Numeric Aggregation
  - Simple Aggregation operators – Mean, Min, Max, Mode
  - Cannot express above level 2
- Set Distances
  - Relational Distance metric & KNN
  - Calculates the minimum distance of all possible pairs of objects
  - Distance – Sum of squared distance (numeric values) or edit distance (categorical values)
  - Assumes attribute independence

Transformation Based Learning



Value Distributions

- Value Order: List of (Value: Index) pairs
  - Ex (watch:1, book:2,CD:3,DVD:4)
- Case Vector
  - Ex {book,CD,CD,book,DVD,book} for case t
  - $CV^{t}_{Products.ProductType} = (0,3,2,1)$
- Reference Vector – based on a condition c
  - Has at position i the sum of values CV[i] for all cases t for which c was true
  - Ex Number of CDs
- Variance Vector – $(CV[i])^2 / (N_c - 1)$
  where N_c – number of cases where c is true

Aggregation = Density Estimation

Target Dependent Individual Values

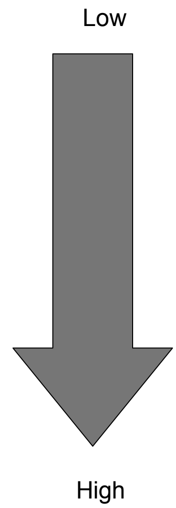
RV Class +ve	
Book	.01
CD	.31
DVD	.35
VCR	.33

RV Class -ve	
Book	.21
CD	.36
DVD	.28
VCR	.15

- Most common (MC) - CD
- Most common positive (MOP): DVD
- Most common Negative (MON): CD
- Most Discriminative (MOD): Book

Feature Complexity

1. No Relational Features
2. Unconditional Features MC, Count
3. Class Conditional Features – MOP,MON
4. Discriminative Class Conditional Features – MOD,MOM



Vector Distances

$$EDD = ED(RV^{y=1}, CV) - ED(RV^{y=0}, CV)$$
$$EUD = EU(RV^{y=1}, CV) - EU(RV^{y=0}, CV)$$
$$COSD = COS(RV^{y=1}, CV) - COS(RV^{y=0}, CV)$$
$$MAD = MA(RV^{y=1}, CV) - MA(RV^{y=0}, CV)$$

Reference Vector	**Euclidean**	**Edit**	**Cosine**	**Mahalanobis**
All	EU	ED	COS	MA
Positive Negative	EUP EUN	EDP EDN	COSP COSN	MAP MAN
Positive vs. Negative	EUD	EDD	COSD	MAD

Domain: Initial Public Offerings

- IPO(Date,Size,Price,Ticker,Exchange,SIC,Runup)
- HEAD(Ticker,Bank)
- UNDER(Ticker,Bank)
- IND(SIC,Ind2)
- IND2(Ind2,Ind)

- Goal: To predict whether the offer was made on the NASDAQ exchange

Implementation details

- Four approaches were tested
  - ILP
  - Logic Based feature construction
  - Selection of specific individual values
  - Target dependent vector aggregation

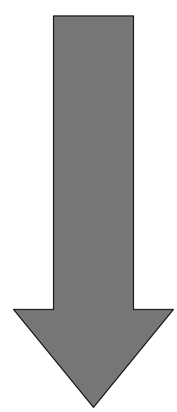
- Two features were constructed
  - One for (n:1) joins
  - Other for autocorrelation

Details (Contd)

- Exploration – To find related objects
  – Uses BFS
  – Stopping criterion – maximum number of chains
- Feature Selection – Weighted Sampling to select a subset of 10 features
- Model Estimation – Uses C4.5 to learn a tree
  – No change in results if logistic regression was used
- Logic Based Feature construction – Uses ILP to learn FOL clauses and append the binary features
- ILP – Only class labels

Aggregation approaches

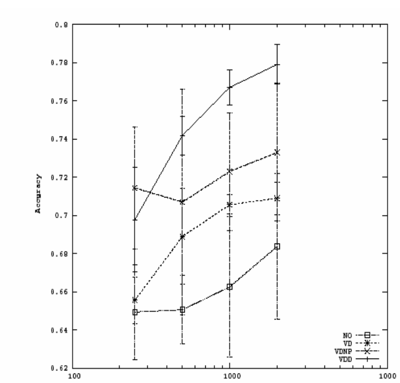
NO	No Feature Construction
MOC VD MVD	Unconditional features – Counts in IPO table
MOP MON VDPN	Class Conditional Features – Most positive and Negative categoricals and vector distances
MOD MOM MVDD	Discriminative Features – Most common categoricals and vector distances



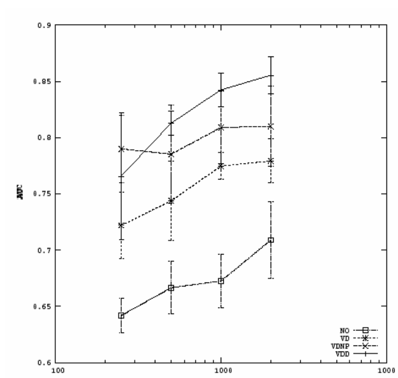
low Complexity Level high

		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

AUC values for aggregation methods grouped by complexity



Accuracy



AUC

As complexity increases, performance increases

As training size increases, performance increases

Conclusions

- Expressive power of models combined with aggregation
- Distance metric
- Complex aggregations can reduce explorations
- Focusses only upto level 2 of the hierarchy