CorClass: Correlated Association Rule Mining for Classification

Albrecht Zimmermann and Luc De Raedt
Institute of Computer Science
Machine Learning Lab
Albert-Ludwigs-University Freiburg

October 3, 2004

Outline

- Problem Setting
- Associative Classification
- Our Solution
- Interestingness Measures
- Rule Conflict Resolution
- Experiments
- Conclusions and Future Work

Classification

- Examples $\langle A_1 = v, ..., A_l = v, C = c\rangle$
- Goal: some way to predict value of C
- Multitude of approaches, among them: rule induction
- Goal: Rules $A_i = v \wedge ... \wedge A_j = v$, reliably predicting value of C
- Several approaches: Separate-and-Conquer (DTs), Divide-and-Conquer (CN2), relatively new: Associative Classification

Associative Classification

- Main Idea:
  - Mine frequent itemsets that include some $C = c$ as item
  - Keep those association rules $\bigwedge A = v \Rightarrow C = c$ (*class association rules*) with high confidence
  - Use for classification
- Advantages: uses well-known and efficient algorithms; many rules, so usage/combination should be robust
- Disadvantages: how to set *MinSupport/MinConfidence*; many rules, so usage/combination has to be smart

Associative Classification - Existing Approaches

- The Pioneer: CBA [Liu et al., 1998]
  - Mining: *Apriori*, MinSupport=1%, MinConfidence=50%
  - Post-processing: ranking, pruning w.r.t. database coverage
  - Ordered rule list used for classification

Associative Classification - Existing Approaches

- The Pioneer: CBA [Liu et al., 1998]
- The Follow-Up: CMAR [Li et al., 2001]
  - Mining: *FP-Growth*, MinSupport/MinConfidence as in CBA
  - Post-processing: ranking, pruning as CBA, removing all negative correlations
  - Unordered rule set used for classification, conflicts resolved by weighted voting (*weighted*-χ^2)

Associative Classification - Existing Approaches

- The Pioneer: CBA [Liu et al., 1998]
- The Follow-Up: CMAR [Li et al., 2001]
- The Inspiration: Classification Using Association Rules [Mutter, 2004] (Thesis)
  - Comparison of different mining and conflict resolution approaches
  - Evaluates *predictive Apriori* [Scheffer, 2001]
  - Conflict Resolution: majority vote, linear/inverse weight voting

Our Solution

- Instead of Support/Confidence, use *interestingness measures* (*Entropy Gain*, *Category Utility*, χ^2), mine k best rules
- Dynamically raise pruning threshold
- No guess-work as to support threshold necessary, low-coverage penalty built in, minimum of interestingness measure statistically better founded
- No post-processing necessary
- Expected: discriminative rules have better performance for classification
- Conflict Resolution: several possibilities

Interestingness Measures

- Interestingness measures σ quantify differences between expected and observed frequency

	K	$\neg K$	
A	y	$x - y$	x
$\neg A$	$m - y$	$n - m - (x - y)$	$n - x$
	m	$n - m$	n

- E.g. upper left cell: observed frequency of $A \wedge K$
- Comparison to expected frequency, e.g. χ^2: $\frac{(mx/n - y)^2}{mx/n}$

Interestingness Measures - cont.

- n =|Dataset|, m_i =*freq*(c_i) fixed
- For rule body b of the form $\bigwedge A = v$: x =*freq*(b), y_i =*freq*$(b \wedge c_i)$

	c_1	c_2	c_3	
b	*freq*$(b \wedge c_1) = y_1$	*freq*$(b \wedge c_2)$	*freq*$(b \wedge c_3) = x - (y_1 + y_2)$	*freq*$(b) = x$
$\neg b$	*freq*$(\neg b \wedge c_1) m_1 - y_1$	*freq*$(\neg b \wedge c_2)$	*freq*$(\neg b \wedge c_3)$	*freq*$(\neg b) = n - x$
	freq$(c_1) = m_1$	*freq*(c_2)	*freq*$(c_3) = n - (m_1 + m_2)$	n

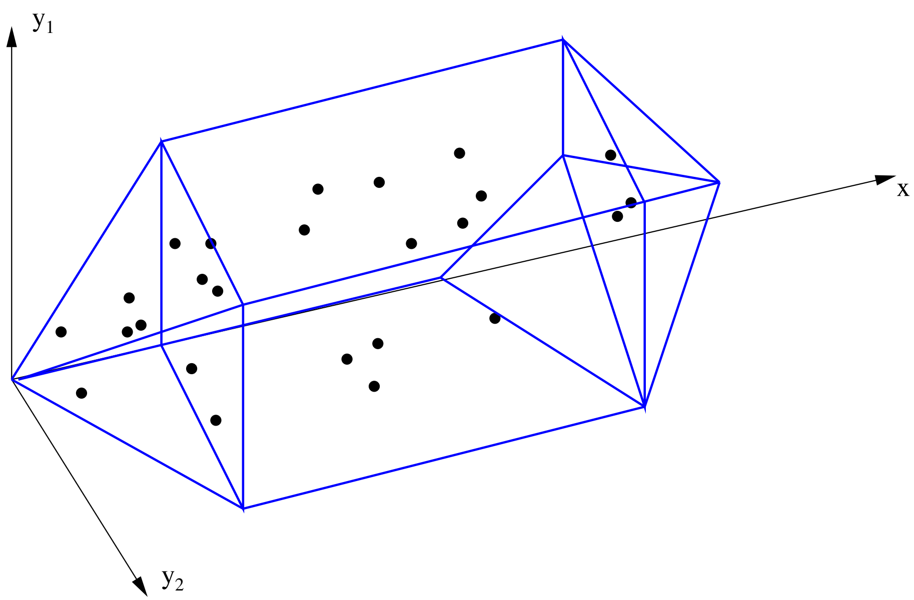
- Assume d class values $\{c_1, \ldots, c_d\}$
- Every rule body characterized by tuple $\langle x, y_1, \ldots, y_{d-1} \rangle$ (*stamp point*)

Interestingness Measures - Problem

- Goal: find k highest scoring rules without exhaustive search
- Problem: σ shows no "nice" behavior w.r.t. specializations
- Existing solutions: beam search, greedy search, similar heuristics
- Better: exploit convexity

Interestingness Measures - Convexity

- σ convex $\Rightarrow$ maximum of future values at convex hull of future stamp points
- Future stamp points unknown, some constraints known $\Rightarrow$ enumerate convex hull for possible stamp points
- Calculate σ on that convex hull $\Rightarrow$ upper bound on future values $\Rightarrow$ use for pruning [Morishita and Sese, 2000]



Algorithm

1. Pick most promising rule body (aka has highest *upper bound*) $b \in S_{cand}$
2. Specialize, count frequencies, compute value of σ, compute upper bound ub_σ
3. For each $b' \supset b$
   - **If** specialization $s \in S_{sol}$, $s \supset b'$ with same σ, stamp point: remove it
   - **If** $|S_{sol}| = k$: If $\sigma(b') >$ *kth*-best σ and no generalization $g \in S_{sol}$, $g \subset b'$ with same σ, stamp point: raise threshold τ, drop worst rule, include b'
   - **Else**: If $\sigma(b') \geq$ user-supplied threshold, include b' in solution set
4. Each b' with $ub_\sigma(b') \geq \tau$ included in candidate set S_{cand}
5. Each $s \in S_{cand}$ with $ub_\sigma(s) < \tau$ removed
6. Repeat until $|S_{cand}| = 0$

Rule Conflict Resolution

- Problem: overlap of rules probable, possibly conflicting predictions
- Solutions:
  - Decision List (highest scoring rule decides)
  - Majority Vote (each rule has weight 1)
  - Linear Weight Voting (each rule has weight $1 - \frac{rank(r)}{rank_{max}+1}$)
  - Inverse Weight Voting (each rule has weight $\frac{1}{rank(r)}$)
  - Weighted-χ^2 Voting (each rule has weight trading off actual with maximum χ^2-value)

Experimental Evaluation

- Several UCI datasets, discretized if necessary
- 10-fold cross-validation
- Mined 10,100,200,500,700,1000 rules
- All combinations of interestingness measures (*entropy gain, category utility*, χ^2) and rule conflict resolution strategies

Results

- Fewer candidate rules considered than mined by support-based approaches
- When compared to CBA and *Apriori, predictive Apriori* with different combination strategies
  - CorClass often best result
  - With exception of Tic-Tac-Toe dataset, always competitive
- Compared to CMAR, CorClass equally good, sometimes better (except Tic-Tac-Toe)
- Compared to C4.5, PART, Ripper, CorClass equally good, sometimes better (except Tic-Tac-Toe)
- Information Gain, Category Utility, χ^2 perform equally good
- Majority vote performs better than the rest, *weighted*-χ^2 improves performance on Tic-Tac-Toe significantly

Observations

- Strong Majority Vote: low-coverage rules useful, no discount necessary
- Vaunted Tic-Tac-Toe: high accuracy $\Leftrightarrow$ low coverage

  $\Rightarrow$ low-coverage penalty might be problematic

Conclusions

- Introduced efficient and effective mining algorithm based on convex interestingness measures
- No need for elaborate post-processing strategies
- Produces high-quality rules
- No best interestingness measure
- No clear-cut strategy for conflict resolution
- Weakness on datasets in which good classification synonymous with small coverage (e.g. many classes)

Future Research

- More extensive experiments
- Additional interestingness measures
- Additional resolution strategies (double/recursive induction, NB)
- Strategy for small coverage (complex concepts, more than 2 classes) ⇒ several upper bounds, negative rules

References

[Li et al., 2001] Li, W., Han, J., and Pei, J. (2001). CMAR: Accurate and efficient classification based on multiple class-association rules. In Cercone, N., Lin, T. Y., and Wu, X., editors, *Proceedings of the 2001 IEEE International Conference on Data Mining*, pages 369–376, San José, California, USA. IEEE Computer Society.

[Liu et al., 1998] Liu, B., Hsu, W., and Ma, Y. (1998). Integrating classification and association rule mining. In Agrawal, R., Stolorz, P. E., and Piatetsky-Shapiro, G., editors, *KDD*, pages 80–86, New York City, New York, USA. AAAI Press.

[Morishita and Sese, 2000] Morishita, S. and Sese, J. (2000). Traversing itemset lattices with statistical metric pruning. In *PODS*, pages 226–236, Dallas, Texas, USA. ACM.

[Mutter, 2004] Mutter, S. (2004). Classification using association rules. Master's thesis, Albert-Ludwigs-Universität Freiburg/University of Waikato, Freiburg, Germany/Hamilton, New Zealand.

[Scheffer, 2001] Scheffer, T. (2001). Finding association rules that trade support optimally against confidence. In De Raedt, L. and Siebes, A., editors, *PKDD 2001*, Lecture Notes in Computer Science, pages 424–435, Freiburg, Germany. Springer.