



Week 5 Video 3

Relationship Mining

Association Rule Mining

Association Rule Mining

- Try to automatically find simple if-then rules within the data set

Example

- Famous (and fake) example:
 - ▣ People who buy more diapers buy more beer
- If person X buys diapers,
- Person X buys beer
- Conclusion: put expensive beer next to the diapers

Interpretation #1

- Guys are sent to the grocery store to buy diapers, they want to have a drink down at the pub, but they buy beer to get drunk at home instead

Interpretation #2

- There's just no ***time*** to go to the bathroom during a major drinking bout

Serious Issue

- Association rules imply causality by their if-then nature
- But causality can go either direction

If-conditions can be more complex

- If person X buys diapers, and person X is male, and it is after 7pm, then person Y buys beer

Then-conditions can also be more complex

- If person X buys diapers, and person X is male, and it is after 7pm, then person Y buys beer and tortilla chips and salsa
- Can be harder to use, sometimes eliminated from consideration

Useful for...

- Generating hypotheses to study further
- Finding unexpected connections
 - ▣ Is there a surprisingly ineffective instructor or math problem?
 - ▣ Are there e-learning resources that tend to be selected together?

Association Rule Mining

- Find rules
- Evaluate rules

Association Rule Mining

- Find rules
- Evaluate rules

Rule Evaluation

- What would make a rule “good”?

Rule Evaluation

- Support/Coverage
- Confidence
- “Interestingness”

Support/Coverage

- Number of data points that fit the rule, divided by the total number of data points
- (Variant: just the number of data points that fit the rule)

Example

- Rule:
If a student took
Advanced Data
Mining, the student
took Intro Statistics
- Support/coverage?

Took Adv. DM	Took Intro Stat.
1	1
0	1
0	1
0	1
0	1
0	1
1	0
1	0
1	0
1	0
1	1

Example

- Rule:
If a student took
Advanced Data
Mining, the student
took Intro Statistics
- Support/coverage?
- $2/11 = 0.2222$

Took Adv. DM	Took Intro Stat.
1	1
0	1
0	1
0	1
0	1
0	1
1	0
1	0
1	0
1	0
1	1

Confidence

- Number of data points that fit the rule, divided by the number of data points that fit the rule's IF condition
- Equivalent to precision in classification
- Also referred to as accuracy, just to make things confusing
- **NOT** equivalent to accuracy in classification

Example

- Rule:
If a student took
Advanced Data
Mining, the student
took Intro Statistics
- Confidence?

Took Adv. DM	Took Intro Stat.
1	1
0	1
0	1
0	1
0	1
0	1
1	0
1	0
1	0
1	0
1	1

Example

- Rule:
If a student took
Advanced Data
Mining, the student
took Intro Statistics
- Confidence?
- $2/6 = 0.33$

Took Adv. DM	Took Intro Stat.
1	1
0	1
0	1
0	1
0	1
0	1
1	0
1	0
1	0
1	0
1	1

Important Note

- Implementations of Association Rule Mining sometimes differ based on whether the values for support and confidence (and other metrics)
- Are calculated based on exact cases
- Or some other grouping variable (sometimes called “customer” in specific packages)

For example

- Let's say you are looking at whether boredom follows frustration
- If Frustrated at time N ,
Then Bored at time $N+1$

Frustrated Time N	Bored Time $N+1$
0	0
0	0
0	0
0	0
0	0
0	1
1	1
1	1
1	1
1	0
1	1

For example

- If you just calculate it this way,
- Confidence = $4/5$

Frustrated Time N	Bored Time N+1
0	0
0	0
0	0
0	0
0	0
0	1
1	1
1	1
1	1
1	0
1	1

For example

- But if you treat student as your “customer” grouping variable
- Then whole rule applies for A, C
- And IF applies for A, B, C
- So confidence = $2/3$

Student	Frustrated Time N	Bored Time N+1
A	0	0
B	0	0
C	0	0
A	0	0
B	0	0
C	0	1
A	1	1
B	1	1
C	1	1
A	1	0
C	1	1

Arbitrary Cut-offs

- The association rule mining community differs from most other methodological communities by acknowledging that cut-offs for support and confidence are arbitrary
- Researchers typically adjust them to find a desirable number of rules to investigate, ordering from best-to-worst...
- Rather than arbitrarily saying that all rules over a certain cut-off are “good”

Other Metrics

- Support and confidence aren't enough
- Why not?

Why not?

- Possible to generate large numbers of trivial associations
 - ▣ Students who took a course took its prerequisites (AUTHORS REDACTED, 2009)
 - ▣ Students who do poorly on the exams fail the course (AUTHOR REDACTED, 2009)

Interestingness



Interestingness

- Not quite what it sounds like
- Typically defined as measures other than support and confidence
- Rather than an actual measure of the novelty or usefulness of the discovery

Potential Interestingness Measures

- Cosine

$$\frac{P(A \wedge B)}{\sqrt{P(A) * P(B)}}$$

- Measures co-occurrence

- Merceron & Yacef (2008) note that it is easy to interpret (numbers closer to 1 than 0 are better; over 0.65 is desirable)



Quiz

- If a student took Advanced Data Mining, the student took Intro Statistics
- Cosine?
 - A) 0.160
 - B) 0.309
 - C) 0.519
 - D) 0.720

Took Adv. DM	Took Intro Stat.
1	1
0	1
0	1
0	1
0	1
0	1
0	1
1	0
1	0
1	0
1	0
1	1

Potential Interestingness Measures

- Lift

$$\frac{\text{Confidence}(A \rightarrow B)}{P(B)}$$

- Measures whether data points that have both A and B are more common than data points only containing B
- Merceron & Yacef (2008) note that it is easy to interpret (lift over 1 indicates stronger association)

Quiz

- If a student took Advanced Data Mining, the student took Intro Statistics
 - Lift?
- A) 0.333
- B) 0.429
- C) 0.500
- D) 0.643

Took Adv. DM	Took Intro Stat.
1	1
0	1
0	1
0	1
0	1
0	1
1	0
1	0
1	0
1	0
1	1

Merceron & Yacef recommendation

- Rules with high cosine or high lift should be considered interesting

Other Interestingness measures

(Tan, Kumar, & Srivastava, 2002)



1	ϕ -coefficient	$\frac{P(A,B) - P(A)P(B)}{\sqrt{P(A)P(B)(1-P(A))(1-P(B))}}$
2	Goodman-Kruskal's (λ)	$\frac{\sum_j \max_k P(A_j, B_k) - \max_j P(A_j) - \max_k P(B_k)}{2 - \max_j P(A_j) - \max_k P(B_k)}$
3	Odds ratio (α)	$\frac{P(A,B)P(\bar{A},\bar{B})}{P(A,\bar{B})P(\bar{A},B)}$
4	Yule's Q	$\frac{P(A,B)P(\bar{A}\bar{B}) - P(A,\bar{B})P(\bar{A},B)}{P(A,B)P(\bar{A}\bar{B}) + P(A,\bar{B})P(\bar{A},B)} = \frac{\alpha - 1}{\alpha + 1}$
5	Yule's Y	$\frac{\sqrt{P(A,B)P(\bar{A}\bar{B})} - \sqrt{P(A,\bar{B})P(\bar{A},B)}}{\sqrt{P(A,B)P(\bar{A}\bar{B})} + \sqrt{P(A,\bar{B})P(\bar{A},B)}} = \frac{\sqrt{\alpha} - 1}{\sqrt{\alpha} + 1}$
6	Kappa (κ)	$\frac{P(A,B) + P(\bar{A},\bar{B}) - P(A)P(B) - P(\bar{A})P(\bar{B})}{1 - P(A)P(B) - P(\bar{A})P(\bar{B})}$
7	Mutual Information (M)	$\frac{\sum_i \sum_j P(A_i, B_j) \log \frac{P(A_i, B_j)}{P(A_i)P(B_j)}}{\min(-\sum_i P(A_i) \log P(A_i), -\sum_j P(B_j) \log P(B_j))}$
8	J-Measure (J)	$\max \left(P(A, B) \log \left(\frac{P(B A)}{P(B)} \right) + P(\bar{A}\bar{B}) \log \left(\frac{P(\bar{B} \bar{A})}{P(\bar{B})} \right), \right.$
9	Gini index (G)	$\left. P(A, B) \log \left(\frac{P(A B)}{P(A)} \right) + P(\bar{A}B) \log \left(\frac{P(\bar{A} B)}{P(\bar{A})} \right) \right)$ $\max \left(P(A)[P(B A)^2 + P(\bar{B} A)^2] + P(\bar{A})[P(B \bar{A})^2 + P(\bar{B} \bar{A})^2] \right.$ $\left. - P(B)^2 - P(\bar{B})^2, \right.$ $\left. P(B)[P(A B)^2 + P(\bar{A} B)^2] + P(\bar{B})[P(A \bar{B})^2 + P(\bar{A} \bar{B})^2] \right.$ $\left. - P(A)^2 - P(\bar{A})^2 \right)$
12	Laplace (L)	$\max \left(\frac{NP(A,B)+1}{NP(A)+2}, \frac{NP(A,B)+1}{NP(B)+2} \right)$
13	Conviction (V)	$\max \left(\frac{P(A)P(\bar{B})}{P(\bar{A}B)}, \frac{P(B)P(\bar{A})}{P(\bar{B}A)} \right)$
14	Interest (I)	$\frac{P(A,B)}{P(A)P(B)}$
15	cosine (IS)	$\frac{P(A,B)}{\sqrt{P(A)P(B)}}$
16	Piatetsky-Shapiro's (PS)	$P(A, B) - P(A)P(B)$
17	Certainty factor (F)	$\max \left(\frac{P(B A) - P(B)}{1 - P(B)}, \frac{P(A B) - P(A)}{1 - P(A)} \right)$
18	Added Value (AV)	$\max(P(B A) - P(B), P(A B) - P(A))$
19	Collective strength (S)	$\frac{P(A,B) + P(\bar{A}\bar{B})}{P(A)P(B) + P(\bar{A})P(\bar{B})} \times \frac{1 - P(A)P(B) - P(\bar{A})P(\bar{B})}{1 - P(A,B) - P(\bar{A}\bar{B})}$
20	Jaccard (ζ)	$\frac{P(A,B)}{P(A) + P(B) - P(A,B)}$
21	Klosgen (K)	$\sqrt{P(A,B) \max(P(B A) - P(B), P(A B) - P(A))}$

Worth drawing your attention to

- Jaccard

$$\frac{P(A \cap B)}{P(A) + P(B) - P(A \cap B)}$$

- Measures the relative degree to which having A and B together is more likely than having either A or B but not both

Other idea for selection

- Select rules based both on interestingness and based on being different from other rules already selected (e.g., involve different operators)

Alternate approach (Bazaldua et al., 2014)

- Compared “interestingness” measures to human judgments about how interesting the rules were
- They found that Jaccard and Cosine were the best single predictors
- And that Lift had predictive power independent of them
- But they also found that the correlations between [Jaccard and Cosine] and [human ratings of interestingness] were negative
 - ▣ For Cosine, opposite of prediction in Merceron & Yacef!



Open debate in the field...



Association Rule Mining

- Find rules
- Evaluate rules

The Apriori algorithm (Agrawal et al., 1996)

1. Generate frequent itemset
2. Generate rules from frequent itemset



Generate Frequent Itemset

- Generate all single items, take those with support over threshold – $\{i_1\}$
- Generate all pairs of items from items in $\{i_1\}$, take those with support over threshold – $\{i_2\}$
- Generate all triplets of items from items in $\{i_2\}$, take those with support over threshold – $\{i_3\}$
- And so on...
- Then form joint itemset of all itemsets

Generate Rules From Frequent Itemset

- Given a frequent itemset, take all items with at least two components
- Generate rules from these items
 - ▣ E.g. $\{A,B,C,D\}$ leads to $\{A,B,C\} \rightarrow D$, $\{A,B,D\} \rightarrow C$, $\{A,B\} \rightarrow \{C,D\}$, etc. etc.
- Eliminate rules with confidence below threshold

Finally

- Rank the resulting rules using your interest measures

Other Algorithms

- Typically differ primarily in terms of style of search for rules

Variant on association rules

- Negative association rules (Brin et al., 1997)
 - ▣ What ***doesn't*** go together?
(especially if probability suggests that two things should go together)
 - ▣ People who buy diapers don't buy car wax, even though 30-year old males buy both?
 - ▣ People who take advanced data mining don't take hierarchical linear models, even though everyone who takes either has advanced math?
 - ▣ Students who game the system don't go off-task?

Next lecture

- Sequential Pattern Mining