Data Mining Part 1. Introduction

1.3 Data Mining Functionalities

Spring 2010

Instructor: Dr. Masoud Yaghini

Outline

- Introduction
- Mining Associations Rules
- Classification
- Numeric Prediction
- Cluster Analysis
- Interesting Patterns
- References

Data mining tasks

Descriptive data mining

characterize the general properties of the data in the database.

Predictive data mining

perform inference on the current data in order to make predictions.

• Different views lead to different classifications of data mining functionalities:

- Data view: Kinds of data to be mined
 - e.g. numeric, categorical, mixed, & ...
- Knowledge view: Kinds of knowledge to be discovered
 - e.g. decision tree, classification rules, & ...
- Method view: Kinds of techniques utilized
 - e.g. neural networks, SVM, & ...
- Application view: Kinds of applications adapted
 - e.g. marketing, medicine, railway, & ...

Knowledge view

 Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks.

Main functionalities

- Mining Association Rules
- Classification
- Numeric Prediction
- Cluster Analysis

Mining Association Rules

Frequent Patterns

- Frequent patterns are patterns that occur frequently in data.
- The kinds of frequent patterns
 - Frequent itemsets patterns: refers to a set of items that frequently appear together in a transactional data set, such as milk and bread.
 - Frequent sequential patterns: such as the pattern that customers tend to purchase first a PC, followed by a digital camera, and then a memory card, is a (frequent) sequential pattern.
- Mining frequent patterns leads to the discovery of interesting associations and correlations within data.

- Suppose, as a marketing manager of *AllElectronics*, you would like to determine which items are frequently purchased together within the same transactions.
- An example of **association rule** from the *AllElectronics* transactional database, is:

 $buys(X, "computer") \Rightarrow buys(X, "software")$ [support = 1%, confidence = 50%]

- where X is a variable representing a customer.

- A confidence, or certainty, of 50% means that if a customer buys a *computer*, there is a 50% chance that she will buy *software* as well.
- A 1% rule support means that 1% of all of the transactions under analysis showed that computer and software were purchased together.
- This association rule involves a single attribute or predicate (i.e., *buys*) that repeats.
- Association rules that contain a single predicate are referred to as single-dimensional association rules.

• We may find association rules like: $age(X, "20...29") \land income(X, "20K...29K") \Rightarrow buys(X, "CD player")$ [support = 2%, confidence = 60%]

- The rule indicates that of the AllElectronics customers under study, 2% are 20 to 29 years of age with an income of 20,000 to 29,000 and have purchased a CD player at AllElectronics.
- There is a 60% probability that a customer in this age and income group will purchase a CD player.
- This is an association between more than one attribute (i.e., *age*, *income*, and *buys*).
- This is a multidimensional association rule.

Support and confidence of a rule

• Example: 4 cool days with normal humidity

If temperature = cool then humidity = normal

Support = 4, confidence = 100%

 Normally: minimum support and confidence prespecified, e.g. support >= 2 and confidence >= 95% for weather data

Interpreting association rules

• Interpretation is not obvious:

If windy = false and play = no then outlook = sunny and humidity = high

• is **not** the same as

If windy = false and play = no then outlook = sunny If windy = false and play = no then humidity = high

• It means that the following also holds:

```
If humidity = high and windy = false and play = no
then outlook = sunny
```

• Large number of possible associations

- Output needs to be restricted to show only the most predictive associations
- Association rules are interesting if they do satisfy both:
 - A minimum support threshold: number of instances predicted correctly
 - A minimum confidence threshold: number of correct predictions, as proportion of all instances that rule applies to

- Association learning is unsupervised
- Association rules usually involve only nonnumeric attributes
- Additional analysis can be performed to uncover interesting statistical correlations between associated attribute-value pairs.

• Examples of association rule algorithms:

- Apriori Algorithm
- FP-growth Algorithm
- Tree-Projection algorithm
- ECLAT (Equivalence CLASS Transformation) algorithm

Classification

Classification

Classification

 Construct models (functions) that describe and distinguish classes or concepts to predict the class of objects whose class label is unknown

• Example:

- In weather problem the play or don't play judgment
- In contact lenses problem the lens recommendation
- Classification learning is supervised
 - Process is provided with actual outcome

Training vs. Test Data Set

Training data set

 The derived model is based on the analysis of a set of training data (i.e., data objects whose class label is known).

Test data set

- Using an independent set of test data for which class labels are known but not made available to the machine.
- This data set is used to evaluate the success of classification learning

Classification

- Examples of classification output:
 - Decision tree
 - Classification rules
 - Neural networks

Decision trees

- "Divide and conquer" approach produces decision tree
- Nodes involve testing a particular attribute
- Usually, attribute value is compared to constant
- Other possibilities:
 - Comparing values of two attributes
 - Using a function of one or more attributes

Decision trees

Leaf nodes

- give a classification that applies to all instances that reach the leaf
- set of classifications
- probability distribution over all possible classifications
- To classify an unknown instance,
 - it is routed down the tree according to the values of the attributes tested in successive nodes, and
 - when a leaf is reached the instance is classified according to the class assigned to the leaf.

Nominal and numeric attributes

- If the attribute is nominal:
 - Number of children usually equal to the number of possible values
 - Usually attribute won't get tested more than once
- If the attribute is numeric:
 - Test whether value is greater or less than constant
 - Attribute may get tested several times
 - Other possibility: three-way split (or multi-way split)
 - Integer: less than, equal to, greater than
 - Real: below, within, above

Decision tree for the labor data



Classification rules

- Popular alternative to decision trees
- Rules include two parts:
 - Antecedent or precondition:
 - a series of tests just like the tests at the nodes of a decision tree
 - Tests are usually logically ANDed together
 - All the tests must succeed if the rule is to fire
 - Consequent or conclusion:
 - The class or set of classes or probability distribution assigned by rule
- Example: A rule from contact lens problem

If tear production rate = reduced then recommendation = none

From trees to rules

• Easy: converting a tree into a set of rules

- One rule for each leaf:
 - Antecedent contains a condition for every node on the path from the root to the leaf
 - Consequent is class assigned by the leaf
- Produces rules that are very clear
 - Doesn't matter in which order they are executed
- But: resulting rules are unnecessarily complex
 - It needs to remove redundant tests/rules

From rules to trees

• More difficult: transforming a rule set into a tree

- Tree cannot easily express disjunction between rules
- Example: rules which test different attributes

If x=1 and y=0 then class = a

If x=0 and y=1 then class = a

If x=0 and y=0 then class = b

If x=1 and y=1 then class = b

From rules to trees

The exclusive-or problem



A tree with a replicated subtree

- If it is possible to have a "default" rule that covers cases not specified by the other rules, rules are much more compact than trees
- There are four attributes, *x*, *y*, *z*, and *w*, each can be 1, 2, or 3

If
$$x=1$$
 and $y=1$ then $class = a$

If z=1 and w=1 then class = a

Otherwise class = b

A tree with a replicated subtree

Replicated subtree problem: tree contains identical subtrees

If x=1 and y=1 then class = a

If z=1 and w=1 then class = a

Otherwise class = b



Executing a rule set

- Two ways of executing a rule set:
 - Ordered set of rules ("decision list")
 - Order is important for interpretation
 - Unordered set of rules
 - Rules may overlap and lead to different conclusions for the same instance

Neural Network

• A neural network, when used for classification, is typically a collection of neuron-like processing units with weighted connections between the units.



Classification Techniques

• Examples of classification techniques:

- Decision Trees
- Classification Rules
- Neural Network
- Naïve Bayesian classification
- Support vector machines
- k-nearest neighbor classification

Classification vs. Association Analysis

- Difference association analysis to classification learning:
 - Can predict any attribute's value, not just the class
 - More than one attribute's value at a time
 - There are more association rules than classification rules

• Numeric prediction:

- predicting a numeric quantity
- Numeric prediction is a variant of classification learning in which the outcome is a numeric value rather than a category.
- Learning is supervised
 - Process is being provided with target value
- Measure success on test data

 A version of the weather data in which what is to be predicted is the time (in minutes) to play

Outlook	Temperature	Humidity	Windy	Play time (min.)
sunny	85	85	false	5
sunny	80	90	true	0
overcast	83	86	false	55
rainy	70	96	false	40
rainy	68	80	false	65
rainy	65	70	true	45
overcast	64	65	true	60
sunny	72	95	false	0
sunny	69	70	false	70
rainy	75	80	false	45
sunny	75	70	true	50
overcast	72	90	true	55
overcast	81	75	false	75
rainy	71	91	true	10

• To find the important attributes and how they relate to the numeric outcome is more important than predicting value for new instances.

• Representing numeric prediction:

- Linear regression equation: an equation to predicts a numeric quantity
- Regression tree: a decision tree where each leaf predicts a numeric quantity
 - Predicted value is average value of training instances that reach the leaf
- Model tree: a regression tree with linear regression models at the leaf nodes

Linear regression equation

• Linear regression equation for CPU data

PRP =
 -56.1
 +0.049 MYCT
 +0.015 MMIN
 +0.006 MMAX
 +0.630 CACH
 -0.270 CHMIN
 +1.46 CHMAX

Regression tree

• Regression tree for the CPU data



Regression tree

- We calculate the average of the absolute values of the errors between the predicted and the actual CPU performance measures,
- It turns out to be significantly less for the tree than for the regression equation.

Model tree

Model tree for the CPU data



Cluster Analysis

Cluster Analysis

Clustering

- grouping similar instances into clusters
- Clustering is unsupervised
 - The class of an example is not known
- Example:
 - a version of the iris data in which the type of iris is omitted
 - Then it is likely that the 150 instances fall into natural clusters corresponding to the three iris types.

Iris data as a clustering problem

	Sepal length (cm)	Sepal width (cm)	Petal length (cm)	Petal width (cm)
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
 51	7.0	3.2	4.7	1.4
52	6.4	3.2	4.5	1.5
53	6.9	3.1	4.9	1.5
54	5.5	2.3	4.0	1.3
55	6.5	2.8	4.6	1.5
101	6.2	2.2	6.0	25
101	0.3 E 0	5.5 27	5.0	2.5
102	J.O 7 1	2.7	5.1	1.9
103	7.1	3.0	5.9	2.1
104	6.3	2.9	5.6	1.8
105	6.5	3.0	5.8	2.2
• • •				

Cluster Analysis

• A 2-D plot of customer data with respect to customer locations in a city, showing three data clusters. Each cluster "center" is marked with a "+".



Clustering

 Clustering may be followed by a second step of classification learning in which rules are learned that give an intelligible description of how new instances should be placed into the clusters.

Representing clusters

- The output takes the form of a diagram that shows how the instances fall into clusters.
- Different cases:
 - Simple 2D representation: involves associating a cluster number with each instance
 - Venn diagram: allow one instance to belong to more than one cluster
 - Probabilistic assignment: associate instances with clusters probabilistically
 - Dendrogram: produces a hierarchical structure of clusters (dendron is the Greek word for tree)

Representing clusters





Representing clusters

Probabilistic assignment

	1	2	3
а	0.4	0.1	0.5
b	0.1	0.8	0.1
С	0.3	0.3	0.4
d	0.1	0.1	0.8
е	0.4	0.2	0.4
f	0.1	0.4	0.5
g	0.7	0.2	0.1
ĥ	0.5	0.4	0.1

Dendrogram



Clustering

- Examples of clustering algorithm:
 - k-Means algorithm
 - k-medoids algorithm
 - AGNES (AGglomerative NESting)
 - DIANA (Divisive ANAlysis)
 - BIRCH (Balanced Iterative Reducing and Clustering Using Hierarchies)

Interesting Patterns

Interesting Patterns

 Data mining may generate thousands of patterns: Not all of them are interesting

• What makes a pattern interesting?

- 1. Easily understood by humans,
- 2. Valid on new or test data with some degree of certainty
- 3. Potentially useful
- 4. Novel
- 5. Validates some hypothesis that a user seeks to confirm

Interesting Patterns

 Objective vs. subjective interestingness measures

- Objective: based on statistics and structures of patterns, e.g., support, confidence, etc.
- Subjective: based on user's belief in the data, e.g., unexpectedness, novelty, actionability, etc.

References

References

- Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, Elsevier Inc., 2005. (Chapter 2 & 3)
- J. Han, M. Kamber, **Data Mining: Concepts and Techniques**, Elsevier Inc. 2006. (Chapter 1)

The end