ASSOCIATION RULE LEARNING

Rule-based machine learning - ×

https://en.wikipedia.org/wiki/Rule-based_machine_learning

WIKIPEDIA
The Free Encyclopedia

Search Wikipedia | Search

Create account Log in

Rule-based machine learning

3 languages

Contents [hide]

Article Talk

Read Edit View history Tools

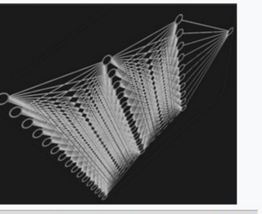
From Wikipedia, the free encyclopedia

Rule-based machine learning (RBML) is a term in computer science intended to encompass any machine learning method that identifies, learns, or evolves 'rules' to store, manipulate or apply.[1][2][3] The defining characteristic of a rule-based machine learner is the identification and utilization of a set of relational rules that collectively represent the knowledge captured by the system. This is in contrast to other machine learners that commonly identify a singular model that can be universally applied to any instance in order to make a prediction.[*clarification needed*][*citation needed*]

Rule-based machine learning approaches include learning classifier systems,[4] association rule learning,[5] artificial immune systems,[6] and any other method that relies on a set of rules, each covering contextual knowledge.

While rule-based machine learning is conceptually a type of rule-based system, it is distinct from traditional rule-based systems, which are often hand-crafted, and other rule-based decision makers. This is because rule-based machine learning applies some form of learning algorithm to automatically identify useful rules, rather than a human needing to apply prior domain knowledge to manually construct rules and curate a rule set.

Part of a series on
Machine learning
and data mining



Paradigms	[show]
Problems	[show]
Supervised learning (**classification** • **regression**)	[show]
Clustering	[show]
Dimensionality reduction	[show]
Structured prediction	[show]
Anomaly detection	[show]
Artificial neural network	[show]
Reinforcement learning	[show]
Learning with humans	[show]
Model diagnostics	[show]

Rules [edit]

Rules typically take the form of an ***'{IF:THEN} expression'***, (e.g. *{IF 'condition' THEN 'result'}*, or as a more specific example, *{IF 'red' AND 'octagon' THEN 'stop-sign}*). An individual rule is not in itself a model, since the rule is only applicable when its condition is satisfied. Therefore rule-based machine learning methods typically comprise a set of rules, or knowledge base, that collectively make up the prediction model.

See also [edit]

- Learning classifier system

RULES

```
if (/* Condition / Antecedent */)
{
    // Something happens / Consequent
}
```

Association of Rules is describing how or why two or more objects / items are related to one another.

Association rule learning - Wiki ×

https://en.wikipedia.org/wiki/Association_rule_learning

WIKIPEDIA
The Free Encyclopedia

Search Wikipedia | Search

Create account Log in

Association rule learning

18 languages

Article Talk Read Edit View history Tools

Contents [hide]

From Wikipedia, the free encyclopedia

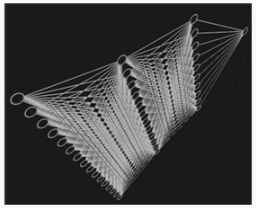
Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness.[1] In any given transaction with a variety of items, association rules are meant to discover the rules that determine how or why certain items are connected.

Based on the concept of strong rules, Rakesh Agrawal, Tomasz Imieliński and Arun Swami[2] introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule $\{\mathrm{onions}, \mathrm{potatoes}\} \Rightarrow \{\mathrm{burger}\}$ found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, they are likely to also buy hamburger meat. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements.

In addition to the above example from market basket analysis, association rules are employed today in many application areas including Web usage mining, intrusion detection, continuous production, and bioinformatics. In contrast with sequence mining, association rule learning typically does not consider the order of items either within a transaction or across transactions.

The association rule algorithm itself consists of various parameters that can make it difficult for those without some expertise in data mining to execute, with many rules that are arduous to understand.[3]

Part of a series on
Machine learning and data mining



Paradigms	[show]
Problems	[show]
Supervised learning (classification • regression)	[show]
Clustering	[show]
Dimensionality reduction	[show]
Structured prediction	[show]
Anomaly detection	[show]
Artificial neural network	[show]
Reinforcement learning	[show]
Learning with humans	[show]
Model diagnostics	[show]

Definition [edit]

Following the original definition by Agrawal, Imieliński, Swami[2] the problem of association rule mining is defined as:

Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of n binary attributes called *items*.

Let $D = \{t_1, t_2, \ldots, t_m\}$ be a set of transactions called the *database*.

Each *transaction* in D has a unique transaction ID and contains a subset of the items in I.

Example: Market Basket Analysis

Transactions

```
t0: milk, bread, eggs
t1: milk, juice
t2: juice, butter
t3: milk, bread, eggs
t4: coffee, eggs
t5: coffee
t6: coffee, juice
t7: milk, cookies, bread, eggs
t8: cookies, butter
t9: milk, bread
```

Association Rules

```
milk -> eggs, bread
bread, eggs -> milk
```

Approaches for Transaction Database Storage

Simple Storage

Transaction ID	Items
T1	i1, i2, i5
T2	i3, i1, i5

Horizontal Storage

TID	i1	i2	i3	i4	i5
T1	1	1	0	0	1
T2	1	0	1	0	1

Vertical Storage

Items	T1	T2
i1	1	1
i2	1	0
i3	0	1
i4	0	0
i5	1	1

OUR ATTEMPT

TID	i1	i2	i3	i4	i5	binary
T1	1	1	0	0	1	11001 = 25
T2	1	0	1	0	1	10101 = 21

APRIORI

1 of 32 Page Width

Fast Algorithms for Mining Association Rules

Rakesh Agrawal Ramakrishnan Srikant*

IBM Almaden Research Center
650 Harry Road, San Jose, CA 95120

Abstract

We consider the problem of discovering association rules between items in a large database of sales transactions. We present two new algorithms for solving this problem that are fundamentally different from the known algorithms. Experiments with synthetic as well as real-life data show that these algorithms outperform the known algorithms by factors ranging from three for small problems to more than an order of magnitude for large problems. We also show how the best features of the two proposed algorithms can be combined into a hybrid algorithm, called AprioriHybrid. Scale-up experiments show that AprioriHybrid scales linearly with the number of transactions. AprioriHybrid also has excellent scale-up properties with respect to the transaction size and the number of items in the database.

Reference to this paper.

INTERESTINGNESS MEASURES

Support, Confidence and Lift

Support

- Frequency / Probability of an itemset.
- `Support(X → Y) = P(XY)`
- Number of times X and Y appears together DIVIDE Total Number of Transactions

Confidence

- Conditional probability that Y will follow when X has already been occurred.
- $P(Y \mid X) = P(X \cap Y) / P(X)$
- $\text{Confidence}(X \rightarrow Y) = \text{Support}(XY) / \text{Support}(X)$

Lift

- Confidence of the rule does not depend on the frequency of Y
- candle → coke and candle → matchbox
- `Lift(X → Y) = P(Y|X) / P(Y)`
- `Lift(X → Y) = Confidence of (X → Y) / P(Y)`

Solved Example

Antecedent	Consequent
A	0
A	0
A	1
A	0
B	1
B	0
B	1
C	0
D	0
C	0
C	1
E	0

Rule $A \rightarrow 0$

$$\text{Support}(A \rightarrow 0) = \boxed{\frac{3}{12}}$$

$$\text{confidence}(A \rightarrow 0) = \frac{\text{Support}(A,0)}{\text{Support}(A)} = \boxed{\frac{3}{4}}$$

$$\text{Lift}(A \rightarrow 0) = \frac{\text{confidence}(A,0)}{\text{Support}(0)} = \boxed{\frac{\frac{3}{4}}{\frac{8}{12}}}$$

Let's have a look at the implementation in visual studio...