Mining Relationships Among Records

Sumeet Gupta

Outline

- Basic Concepts
- Performing Association Rule Mining

Featured Stores
- Hindi Bookstore
- Tamil Bookstore
- 100 Books to Read in a Lifetime
- 100 Must-reads for Children
- Must reads in Crime, Thriller & Mystery
- CAT

Around the Books Store
- Bestsellers
- Pre-orders & New Releases
- Textbooks
- Upcoming Exams
- Young Adults Books Store
- Man Booker Prize Longlist 2014

Amazon Kindle
- Kindle
- Kindle eBooks
- Free Kindle Reading Apps
- 31,000 Free eBooks

Show results for

New Arrivals
- Last 30 days (1,04,347)
- Last 90 days (3,56,354)
- Next 90 days (56,102)

Books
- Action & Adventure (1,08,977)
- Arts, Film & Photography (7,24,383)
- Biographies, Diaries & True Accounts (4,25,664)
- Business & Economics (12,15,482)
- Children's & Young Adult (12,74,538)
- Comics & Mangas (80,040)
- Computing, Internet & Digital Media (2,65,236)

Books

New Releases & Pre-orders | Textbooks | Hindi Bookstore | Tamil Bookstore | 100 Books to Read in a Lifetime | Upcoming Exams | India's Largest Bookstore with over 10 Million Books

This Children's Day
Discover the Magical World of Books

FLAT 50% OFF | on best-selling Children's Books

› Shop now


Children's Day Discounts
Shop now


Amar Chitra Katha
Up to 50% off


Kannada Store
Explore more

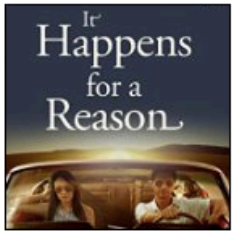

CN Remix Collection
Amazon Exclusive


Short Story Collections
See all

More to Explore

Great Deals on Books

Great Deals in Books
- Find great deals and offers on books. Happy shopping! *See all*



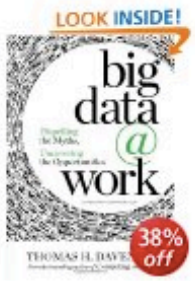
Pre-Orders & New Releases
- Explore the latest releases and pre-order books*See all*

Your Dreams are Mine Now
- Best-selling author, Ravindar Singh, is out with his latest book, Your Dreams are Mine Now *Pre-order now*

The Way Things Were
- Aatish Taseer's The Way Things Were is a magisterial novel about the pressures of history upon the present moment. *Pre-order now*

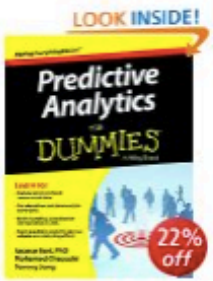
100 Books to Read in a Lifetime
- The Amazon.com editors have put together a list of the 100 must-read books in a lifetime, which includes all time classics, new age romance, and must read series. *See all*

Recommendations for You in Books



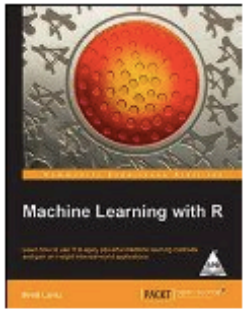
Big Data at Work: Dispelling the...
Thomas H. Davenport
Hardcover
★★★★★ (1)
~~₹1,250.00~~ ₹776.00
Fix this recommendation

Big Data for Dummies
Judith Hurwitz, Alan Nugent, Dr...
Paperback
★★★★½ (4)
~~₹399.00~~ ₹339.00
Fix this recommendation



Predictive Analytics for Dummies
Anasse Bari, Mohamed Chaouchi, Tommy...
Paperback
~~₹549.00~~ ₹428.00
Fix this recommendation

Complete Business Statistics
Amir Aczel, Jayavel Sounderpandian, P...
Paperback
★★★★☆ (1)
~~₹725.00~~ ₹567.00
Fix this recommendation



Machine Learning With R: Learn How to...
Brett Lantz
Paperback
★★★☆☆ (2)
₹775.00
Fix this recommendation

› See more recommendations

Recommendations for You in Electronics

VEEGEE Anti-Finger Print Anti Glare...
★★★½☆ (41)
~~₹600.00~~ ₹350.00
Fix this recommendation

VEEGEE Anti Fingerprint Anti Glare...
★★★½☆ (10)
~~₹500.00~~ ₹350.00
Fix this recommendation

Screen Protector Scratch Guard For...
SCREENWARD
★★★☆☆ (15)
~~₹600.00~~ ₹325.00
Fix this recommendation

KINDLE PAPERWHITE Leather Flip Case...
CUBIX
★★★½☆ (103)
~~₹1,800.00~~ ₹900.00
Fix this recommendation

HOKO Brown Slim Leather Flip Case...
★★★★½ (10)
~~₹1,800.00~~ ₹1,000.00
Fix this recommendation

› See more recommendations

Basic Concepts: Example

Bound Away
Last Train Home

Share your own customer images

List Price: $16.98
Price: **$16.98** and eligible for **FREE Super Saver Shipping** on orders over $25. See details.

Availability: Usually ships within 24 hours

Want it delivered Tomorrow? Order it in the next 4 hours and 9 minutes, and choose **One-Day S** checkout. See details.

41 used & new from $6.99

See more product details

Based on customer purchases, this is the #82 Early Adopter Product in Alternative Rock.

801x612

Buy this title for only $.01 when you get a new Amazon Visa® Card

Apply now and if you're approved instantly, **save $30** off your first purchase, earn **3% rewards**, get a **0% APR**,* and pay **no**

Amazon Visa discount: $30.00
Applied to this item: - $16.97
Discount remaining: $13.03
Find out how
(Don't show again)

Customers who bought this title also bought:

- Time and Water ~ Last Train Home (Why?)
- Cold Roses ~ Ryan Adams & the Cardinals (Why?)
- Tambourine ~ Tift Merritt (Why?)
- Last Train Home ~ Last Train Home (Why?)
- True North ~ Last Train Home (Why?)
- Universal United House of Prayer ~ Buddy Miller (Why?)
- Wicked Twisted Road [ENHANCED] ~ Reckless Kelly (Why?)
- Hacienda Brothers ~ Hacienda Brothers (Why?)

Basic Concepts

- Proposed by Agrawal et al in 1993.
- It is an important data mining model studied extensively by the database and data mining community.
- Assume all data are categorical.
- No good algorithm for numeric data.
- Initially used for Market Basket Analysis to find how items purchased by customers are related.

$$\text{Bread} \rightarrow \text{Milk} \qquad [\text{sup} = 5\%, \text{conf} = 100\%]$$

Basic Concepts

- What are Association Rules?
  - Study of "what goes with what"
    - "Customers who bought X also bought Y"
    - What symptoms go with what diagnosis
  - Transaction-based or event-based
  - Also called "market basket analysis" and "affinity analysis"
  - Originated with study of customer transactions databases to determine associations among items purchased

Basic Concepts: Association Rules

"IF" part = **antecedent** "THEN" part = **consequent**

"Item set" = the items (e.g., products) comprising the antecedent or consequent

- Antecedent and consequent are *disjoint* (i.e., have no items in common)
  - An association rule is an implication of the form:
    $X \rightarrow Y$, where $X, Y \subset I$, *and* $X \cap Y = \varnothing$
  - An itemset is a set of items.
    - E.g., X = {milk, bread, cereal} is an itemset.
  - A *k*-itemset is an itemset with *k* items.
    - E.g., {milk, bread, cereal} is a 3-itemset

Basic Concepts: Association Rules

Transaction	Faceplate Colors Purchased			
1	red	white	green	
2	white	orange		
3	white	blue		
4	red	white	orange	
5	red	blue		
6	white	blue		
7	white	orange		
8	red	white	blue	green
9	red	white	blue	
10	yellow			

Basic Concepts: Association Rules

For example: Transaction 1 supports several rules, such as

- "If red, then white" ("If a red faceplate is purchased, then so is a white one")
- "If white, then red"
- "If red and white, then green"
- + several more

Basic Concepts: Association Rules

Transaction	Red	White	Blue	Orange	Green	Yellow
1	1	1	0	0	1	0
2	0	1	0	1	0	0
3	0	1	1	0	0	0
4	1	1	0	1	0	0
5	1	0	1	0	0	0
6	0	1	1	0	0	0
7	1	0	1	0	0	0
8	1	1	1	0	1	0
9	1	1	1	0	0	0
10	0	0	0	0	0	1

An example

t1:	Butter, Bread, Milk
t2:	Butter, Cheese
t3:	Cheese, Boots
t4:	Butter, Bread, Cheese
t5:	Butter, Bread, Clothes, Cheese, Milk
t6:	Bread, Clothes, Milk
t7:	Bread, Milk, Clothes

- Transaction data
- Assume:
  minsup = 30%
  minconf = 80%

- An example frequent *itemset*:
  {Bread, Clothes, Milk} [sup = 3/7]

- Association rules from the itemset:
  Clothes → Milk, Bread [sup = 3/7, conf = 3/3]
  …
  Clothes, Bread → Milk, [sup = 3/7, conf = 3/3]

Basic Concepts: Objective

- Ideally, we want to create all possible combinations of items
- Problem: computation time grows exponentially as # items increases
- Solution: consider only "frequent item sets"
- Criterion for frequent: support

=> Apriori Algorithm

Basic Concepts: Rule Strengths

- Support: The rule holds with support *sup* in *T* (the transaction data set) if sup% of transactions contain $X \cup Y$.
  - $sup = \Pr(X \cup Y)$.
- Confidence: The rule holds in *T* with confidence *conf* if *conf*% of transactions that contain *X* also contain *Y*.
  - $conf = \Pr(Y \mid X)$
- **Lift** = *confidence*/(*benchmark confidence*)
  - *Benchmark confidence* = transactions with consequent as % of all transactions
  - Lift > 1 indicates a rule that is useful in finding consequent items sets (i.e., more useful than just selecting transactions randomly)

Basic Concepts: Rule Strengths

- Support count: The support count of an itemset X, denoted by $X.count$, in a data set T is the number of transactions in T that contain X. Assume T has n transactions.
- Then,

$$support = \frac{(X \quad Y).count}{n}$$

$$confidence = \frac{(X \quad Y).count}{X.count}$$

Apriori Algorithm

Goal: Find all rules that satisfy the user-specified *minimum support* (minsup) and *minimum confidence* (minconf).

For *k* products…

1. User sets a minimum support criterion
2. Next, generate list of one-item sets that meet the support criterion
3. Use the list of one-item sets to generate list of two-item sets that meet the support criterion
4. Use list of two-item sets to generate list of three-item sets
5. Continue up through *k*-item sets

Apriori Algorithm
Interpretation

- *Lift ratio* shows how effective the rule is in finding consequents (useful if finding particular consequents is important)
- *Confidence* shows the rate at which consequents will be found (useful in learning costs of promotion)
- *Support* measures overall impact

Apriori Algorithm
Caution: The Role of Chance

Random data can generate apparently interesting association rules

The more rules you produce, the greater this danger

Rules based on large numbers of records are less subject to this danger

Summary

- Association rules (or *affinity analysis,* or *market basket analysis*) produce rules on associations between items from a database of transactions
- Widely used in **recommender systems**
- Most popular method is **Apriori algorithm**
- To reduce computation, we consider only "frequent" item sets (=support)
- Performance is measured by *confidence* and *lift*
- Can produce a profusion of rules; review is required to identify useful rules and to reduce redundancy

Thank You