



E-BUSINESS

PROF. MAMATA JENAMANI

DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING
IIT KHARAGPUR

NPTEL

Week 11: Lecture 1

COLLABORATIVE FILTERING BASED RECOMMENDER SYSTEM





We are going to learn

Collaborative filtering



E-Business





Two approaches

- User-User based
 - Identify like-minded users
 - Absolutely no offline processing
 - Likely to be slow
 - The basic collaborative filtering algorithm
- Item-Item based
 - Identify buying patterns
 - Offline processing of major computations
 - Amazons recommender system belongs to this category





User-User based Collaborative filtering The Phases

- Dimension reduction
 - Transform the original user preference matrix into a lower dimensional space to address the sparsity and scalability problem
- Neighborhood formation
 - For an active user, compute the similarities between all other users and the active user to form a proximity-based neighborhood with a number of like minded users.
- Recommendation generation
 - Generate recommendations based on the preferences of the set of nearest neighbors of the active user





Dimension reduction

- Dealing with sparsity and scalability problem

	Andre	Star	Wars	Batman	Rambo	Hiver	Whispers
Lyle	У	У					
Ellen	У	У				У	
Fred		У		У			
Dean		У		У	У		
Jason						У	У

	Actio	n		Forei	gn	Classic
	Batman	Rambo	Andre	Hiver	Whispers	Star Wars
Lyle			у		_	у
Ellen			у	\mathbf{y}		у
Jason				\mathbf{y}	У	
Fred	У					У
Dean	\mathbf{y}	\mathbf{y}				У





Methods for dimension reduction

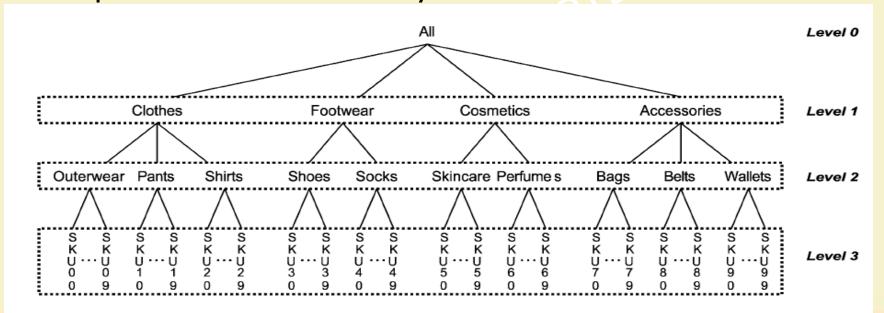
- Semi-manual Methods
 - Use product features
 - Cluster products first, then cluster users
 - Works only if we have descriptive features
- Automatic Methods
 - Adjusted Product Taxonomy
 - Latent Semantic Indexing





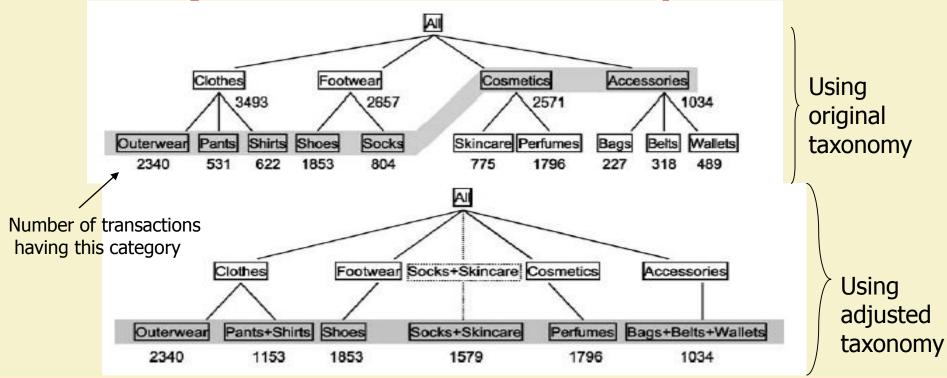
Adjusted Product Taxonomy

- Input: product taxonomy
- Output: modified taxonomy with even distribution





Adjusted Product Taxonomy (2)







Latent semantic indexing

- Reduce the original matrix n×m preference matrix using latent semantic indexing technique a to a lower dimensional n×d matrix with d metaitems. Where d<m
- Uses singular value decomposition method to obtain a rank-d approximation of the original matrix.



Neighborhood Formation

- For an active user u_a find a list of I like minded users
 - Interested in similar items (meta-items)
- Measuring similarity
 - Pearson correlation coefficient
 - Constraint Pearson correlation coefficient
 - Spearman rank correlation
 - Cosine Similarity
 - Mean-square difference
- Neighborhood Formation
 - Weight threshold
 - Center based best-k neighbors
 - Aggregate based best –k neighbors





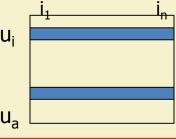
Measuring similarity

Pearson correlation coefficient

$$sim(u_a, u_i) = \frac{\sum\limits_{\mathbf{j} \in \text{Commonly Rated Items}} (p_{aj} - \overline{p_a})(p_{ij} - \overline{p_i})}{\sqrt{\sum\limits_{\mathbf{j} \in \text{Commonly Rated Items}} (p_{aj} - \overline{p_a})^2 \sum\limits_{\mathbf{j} \in \text{Commonly Rated Items}} (p_{ij} - \overline{p_i})^2}}$$

where p_{ij} : preference of i^{th} user for j^{th} item

 \overline{p}_i : mean prefernce of i^{th} user for Commonly Rated Items u_i



Neighborhood Formation

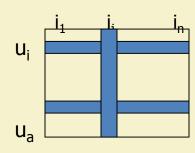
- Weight threshold method
 - Set an absolute threshold
 - Selects all the neighbors whose similarity coefficient is greater that this threshold



Method for Recommendation generation

- Weighted Average
 - Preference scores on item j is a weighted average score of the preference scores with correlation as the weight
- Deviation from the mean

$$p_{aj} = \overline{p_a} + \frac{\sum_{i} sim(u_a, u_i)(p_{ij} - \overline{p_i})}{\sum_{i} sim(u_a, u_i)}$$



Offline Vs. Online processing

- Offline phase:
 - Do nothing...just store transactions
- Online phase:
 - Identify highly similar users to the active one
 - Predict





Item-Item based Collaborative filtering

- Search for similarities among items
- All computations can be done offline
- Item-Item similarity is more stable than user-user similarity
 - No need for frequent updates
- First Order Models
 - Correlation Analysis
 - Linear Regression
- Higher Order Models
 - Belief Network
 - Association Rule Mining

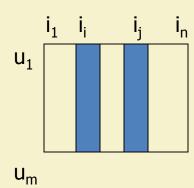




Search for similarities among items - Correlation-based Method

- Same as in user-user similarity but on item vectors
- Pearson correlation coefficient
 - Look for users who rated both items

$$sim(i, j) = \frac{\sum_{u \in \text{Users Rated Both Items}} (p_{uj} - \overline{p_j})(p_{ui} - \overline{p_i})}{\sqrt{\sum_{u \in \text{Users Rated Both Items}} (p_{uj} - \overline{p_j})^2 \sum_{u \in \text{Users Rated Both Items}} (p_{ui} - \overline{p_i})^2}}$$



Predict rating - Correlation-based Method Offline phase:

- Calculate n(n-1) similarity measures
- For each item

 Determine its k-most similar items

Online phase:

■ Predict rating for a given user-item pair as a weighted sum over similar items that he rated

						n <u>i∈similar items</u>				
2		3		٠.		4		P_{aj} –	\sum_{i}	sim(i, j)
				j					i∈similar item.	(/ J /

 $sim(i, j)p_{ai}$





Collaborative Filtering in Amazon -A case

- Amazon.com uses recommendations as a targeted marketing tool in many email campaigns and on most of its Web sites' pages, including the high traffic Amazon.com homepage.
- Amazon.com extensively uses recommendation algorithms to personalize its Web site to each customer's interests.





A challenging environment for recommendation algorithms

- A large retailer like amazon has huge amounts of data, tens of millions of customers and millions of distinct catalog items.
- Many applications require the results set to be returned in realtime, in no more than half a second, while still producing high-quality recommendations.
- New customers typically have extremely limited information, based on only a few purchases or product ratings.
- Older customers can have a glut of information, based on thousands of purchases and ratings.
- Customer data is volatile: Each interaction provides valuable customer data, and the algorithm must respond immediately to new information.





Need to develop a new algorithm

- Because existing recommendation algorithms cannot scale to Amazon.com's tens of millions of customers and products, they develop their own algorithm.
- Their algorithm, item-to-item collaborative filtering, scales to massive data sets and produces high-quality recommendations in real time.





How It Works

Offline Component

- To determine the most-similar match for a given item, the algorithm builds a similar-items table by finding items that customers tend to purchase together.
- It could build a product-to-product matrix by iterating through all item pairs and computing a similarity metric for each pair. However, many product pairs have no common customers, and thus the approach is inefficient in terms of processing time and memory usage.
- Online Component
 - To generate recommendation based on the similarity table produced offline





The Algorithm for generating similar item table

For each item in product catalog, /1

For each customer C who purchased 11

For each item 12 purchased by customer C

Record that a customer purchased /1 and /2

For each item 12

Compute the similarity between /1 and /2





Measuring Similarity

• The similarity can be measured between two products, A and B, in various ways; Amazon's recommendation system uses a common method that measures the cosine of the angle between the two vectors.

$$similarity(\vec{A}, \vec{B}) = cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| * \|\vec{B}\|}$$





Complexity of the algorithm

- Offline computation of the similar-items table is extremely time intensive
- Sampling customers who purchase best-selling titles reduces runtime even further, with little reduction in quality.



Recommendation generation

- Given a similar-items table, the algorithm finds items similar to each of the user's purchases and ratings, aggregates those items, and then recommends the most popular or correlated items.
- This computation is very quick, depending only on the number of items the user purchased or rated.





Collaborative filtering assignment

The *i*th row in the following matrix represents a single transaction by the buyer b_i . The non-zero entries in a row represent the items bought together during the transactions and the corresponding value represents the preference scores assigned to the items by the buyer. If an active buyer 'a' has put i_4 in his shopping cart, recommend one more item to him. Use item-item collaborative filtering algorithm for recommendation generation.

	Items							
		i_1	i_2	i_3	i_4	i_5	i_6	
	b_1		5		6	4		
	b_2			5		8		
	b_3		9	7	8		6	
Duyers	b_4	2	4	6	4			
4	b_5			8	3	5		
	а				*			





NPTEL

Week 11: Lecture 2

ASSOCIATION BASED RECOMMENDER SYSTEM





We are going to learn

Association based Recommender system







Introduction to frequent pattern analysis

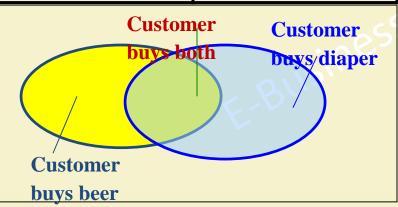
- Frequent pattern: a pattern (a set of items, subsequences, substructures, etc.) that
 occurs frequently in a data set
- Frequent pattern analysis is the basis of association rule mining
- Motivation: Finding inherent regularities in data
 - What products were often purchased together?— Beer and diapers?!
 - What are the subsequent purchases after buying a PC?
 - What kinds of DNA are sensitive to this new drug?
 - Can we automatically classify web documents?
- Applications
 - Basket data analysis, cross-marketing, catalog design, sale campaign analysis, Web log (click stream) analysis, and DNA sequence analysis.





Basic Concepts: Frequent Patterns and Association Rules

Transaction-id	Items bought
10	A, B, D
20	A, C, D
30	A, D, E
40	B, E, F
50	B, C, D, E, F



Itemset $X = \{x_1, ..., x_k\}$

Find all the rules $X \rightarrow Y$ with minimum support and confidence

- support, s, probability that a transaction contains $X \cup Y$
- confidence, c, conditional probability that a transaction having X also contains Y

Let $sup_{min} = 50\%$, $conf_{min} = 50\%$ Freq. Pat.: {A:3, B:3, D:4, E:3, AD:3} Association rules:

 $A \to D$ (60%, 100%), $D \to A$ (60%, 75%)





Interestingness measures

- Association rule mining searches for interesting relationships among items in a given data set.
- Two measures of interestingness: Support and Confidence
- Find all the rules X → Y with minimum support and confidence
 - support, S, probability that a transaction contains $\mathsf{X} \cup \mathsf{Y}$
 - S = (# of tuples containing both X and Y)/(total number of tuples)
 - Support Count = # of tuples containing both X and Y)
 - confidence, C, conditional probability that a transaction having X also contains Y
 - C = (# of tuples containing both X and Y)/(# of tuples containing X alone)
 = (Support count of tuples containing X ∧Y)/(Support count of tuples containing A)



Algorithms for association rule mining

- Three major approaches
 - Apriori algorithm
 - Frequent pattern growth
 - Vertical data format approach



The apriori Algorithm

- Apriori Principle
 - Suppose an item set is not frequent (i.e. does not have the minimum support). If an item A is added to this set then the resulting set cannot occur more frequently.
 - It is an anti-monotone property
 - If a set cannot pass a test then all its supersets will also fail the test.
 - Two steps of the algorithm
 - Join
 - Prune





The algorithm

- scan DB once to get frequent 1-itemset C₁
- $C_1 = Prune(C_1)$
- $L_1 \leftarrow C_1$
- Continue join step till no frequent or candidate set can be generated
- Join
 - − C_k ← A set of k-item sets generated by joining L_{k-1} with itself
 - $-C_{\nu}^{\sim}=Prune(C_{\nu})$
 - $-L_{k}\leftarrow C_{k}$
- Prunë(C_k)

 - Delete the tuples in C_k that do not satisfy the apriori property
 If any (k-1)-subset of a candidate is not in L_{k-1}, then the k-item set cannot be frequent
 - Scan D to get the frequency count of each set in C_k. Delete the sets that does not satisfy the minimum support count.

Assignment

- Derive the frequent pattern from the given transaction database.
- Generate association rules

Tid	Items
10 P	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E



Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

Solution

$Sup_{min} = 2 (50\%)$

1st scan

Itemset	sup
{A}	2
{B}	3
{C}	3
{D}	1
{E}	3

T	Itemset	sup
L_{I}	{A}	2
	{B}	3
N	{C}	3
	{E}	3

Itemset sup

{A, C}	2
{B, C}	2
{B, E}	3
{C, E}	2
-	

Itemset	sup
$\{A, B, C\}$	1
$\{A, B, C, E\}$	1
$\{A, C, E\}$	1
{B, C, E }	2

\boldsymbol{C}	Itemset	sup
C_2	$\{A, B\}$	1
	{A, C}	2
	{A, E}	1
80	{B, C}	2
	{B, E}	3
	{C, E}	2

Itemset $\{A, B\}$ 2nd scan $\{A,C\}$ $\{A, E\}$ $\{B,C\}$ $\{B, E\}$ $\{C, E\}$ **Itemset** sup

 $\{B, C, E\}$





3rd scan

Solution

- Association Rules and Confidence
 - $B \rightarrow C \{2/3\}$
 - $C \rightarrow B \{2/3\}$
 - B→E {3/3}
 - $E \rightarrow B \{3/3\}$
 - $B \rightarrow \{C, E\} \{2/3\}$
 - {C, E}→B {2/2}
 - $C \rightarrow \{B, E\} \{2/3\}$
 - $\{B, E\} \rightarrow C \{2/3\}$

- $E \rightarrow \{B, C\} \{2/3\}$
- $\{B, C\} \rightarrow E\{2/2\}$
- Assuming we go for the rules with 100% confidence only 4 rules qualify



Association rule based recommendation generation

- Generate association rules from the transaction database
- To generate Top-N recommendation
 - Find the association rule supported by the active user (rules whose LHS appears in the active user's transaction)
 - Let I_p be the set of unique items suggested by the RHS of the rules
 - Sort I_p based on confidence score with respect to the association rules.
 Confidence is more if an item appears in more rules.
 - Choose the top N of these items
- Prediction
 - An item can be recommended if it appears in the RHS of the association rules supported by the active user.
- Top M users
 - **–** \hat{i}





NPTEL

Week 11: Lecture 2

DEMOGRAPHICS BASED RECOMMENDER SYSTEM AND WEBSITE PERSONALIZATION





We are going to learn

Association based Recommender system







The approach

- Recommends items to a user based on the preferences of the users whose demographics are similar to those of the user
- Unlike other approaches, where the recommendations are made at the item level here the recommendations are made at the category level to help
 - Providing more generalized information
 - Addressing sparsity problem
- Typical Application: Target advertising in electronic storefront





The steps

- Data transformation
 - Generate a set of training examples each of whose input attributes are the demographics of a user and the decision outcomes are the category preference of the user
- Category preference model
 - Automatically induce the preference model for each category based on the training examples pertaining to the category
 - ANN, Decision tree or any other form of induction learning technique
- Recommendation generation
 - Given the demographic data of an active user, generate recommendations by performing reasoning on the category preference models induced previously





Data transformation

- Transformation of the preference data collected in the item level to the category level
 - Counting-based (frequency threshold) method
 - Expected value method
 - Statistics based method
- Preference data is modeled as discrete values numerically scaled on the user preferences
 - Mostly binary





Counting based methods

 Considers the frequency of favorite preferences of a user on the items in a particular category

$$cp_{aj} = \begin{cases} 1 \text{ if } \sum_{i \in C_j} p_{aj} \ge w \\ 0 \text{ otherwise} \end{cases}$$



Recommendation generation

- Prediction
 - Reasoning on the category preference model
- Top-N items
 - Reasoning over all the category preference models for a single user
 - Estimating prediction accuracy for these predictions and choosing the top-N most accurate ones
- Top-M users
 - Reasoning over a single category over all users
 - Estimating prediction accuracy for these predictions and choosing the top-M most accurate ones





Web site personalization

- Providing each user with individually tailored Web pages to decrease information overload
- Types of personalization
 - personalizing Content
 - personalizing Structure
 - personalizing Layout, presentation, media format etc.
- A kind of recommender system?





Advantages

- Increasing site usability
- Converting users to buyers
- Retaining current customers
- Re-engaging customers
- Penetrating new markets





Two general approaches

- Buyer driven
 - Buyer decides on the rules of personalization
- Seller driven
 - Seller decides on the rules
 - Used for cross selling, up-selling, target advertising etc.





Personalization process

- Data collection
 - User data, usage data and environmental data
 - Reactive approach (explicit) and non-reactive approach (implicit)
- Preprocessing
- User profiling
 - Data mining algorithms: Clustering, classification, association rule mining, sequential pattern discovery
- Personalized output





NPTEL

E-Business





NPTEL

Week 11: Lecture3

DYNAMIC PRICING





We are going to learn

- Concepts of a market and scope for dynamic pricing
- E-Commerce as a driver for dynamic pricing
- Types of dynamic pricing





A market

- A market is a mechanism through which the buyers and sellers interact to determine the prices and exchange of goods and services.
- Prices coordinates the decisions of producers and consumers
- Higher prices tend to reduces consumer purchases and encourages production
- Lower prices encourages consumption and discourages production
- Prices are the balance wheel of the market mechanism



Market equilibrium

- Market equilibrium comes at the price at which the quantity supplied is the quantity demanded.
- The market demand (demanded by all individuals) however changes over time.
- So also the supply.
- Hence there is a scope for the price to change dynamically.



Understanding the scope for dynamic pricing







Some important observations

- It appears that prices should change dynamically depending on the market condition
 - In fact dynamic pricing has a history which is as old as the human civilization
 - Fixed pricing has a history which is only 100 years old.
- Dynamic pricing ensures perfect competition
 - No consumer or the producer is large enough to control the market.
 - Efficient allocation of resources
- If dynamic pricing is so natural why did people go for fixed pricing?





Advantages of fixed pricing

- Convenient
 - Easy to model
 - Designed to recover the cost of production (break-even)
 - Difficulty in estimating demand
- Decrease price uncertainty in the market
 - Loyal customers
- An instrument to control the market
- ...





Fixed pricing method

- Markup pricing
 - Unit price = unit cost / (1 markup on sales)
- Target return pricing
 - Unit cost + [(desired return*invested capital)/unit sales]
- Perceived value pricing
 - Service, warranty, reliability etc.
- Value pricing
 - Quality Vs. Price
- Going rate pricing
 - Competitor's price





Was the traditional pricing really fixed?

- Volume discounts?
- Bargaining and Negotiations?
- Product mix pricing?
- Promotional pricing?
- •





E-Commerce as a driver of dynamic pricing

- What buyers can do
 - Gets instant price comparisons
 - Instant search for the substitutes that fits the budget
- What sellers can do
 - Monitor customer behavior and instant tailoring of customized price
 - Change price on the fly after sensing demand
- Both can do
 - Instant negotiation on price
 - Auctions
 - Exchanges
- Internet has created a conducive environment for perfect competition hence dynamic pricing is becoming a reality



E-Commerce as a driver of dynamic pricing

- Transaction cost for implementing the dynamic pricing have been reduced by
 - Eliminating the need for people to be physically present in time and space
 - Reducing the search cost
 - Reducing the menu cost of informing the changed price
- Increased number of customers, competitors, and increased amount of information has lead to price uncertainty and demand volatility.
- Companies are finding that using a single fixed price in these volatile market is inefficient and ineffective





Defining Dynamic Pricing

- Dynamic pricing is defined as the buying and selling of goods and services in free markets where the prices fluctuate in response to changing supply and demand.
- Also called flexible pricing/ customized pricing
- Includes two aspects
 - Price dispersion
 - price differentiation





Price dispersion

- Spatial
 - Several seller offers a given item in different prices
- Temporal
 - A seller varies the price of a given item over the time
 - Ex: Seasonal discounts





Price Differentiation

Based on one product

- First degree differentiation
 - Perfect differentiation
 - Same product, different price for different people
 - Extracts maximum consumer surplus from the market
 - Ex: Auction
- Second degree differentiation
 - Non-linear pricing
 - Different quantity for different unit price but the rule is same for each individual
 - Ex: Volume discounts, Utility prices
- Third degree differentiation
 - Group pricing
 - Same unit price for any quantity but unit price is different for different groups of people
 - Ex. Telecom pricing for business and households





Price Differentiation

Based on one product which can be customized

- Addition or deletion of attributes
- Decreased substitutability
- Customized product
- Ex.: Dell Computers with customized features
- Ex: Airline industry Product differentiation based on refund policy

Dynamic pricing success stories

- Airline Industry
 - Yield management
- Priceline.com
 - Negotiation with major airlines to fill up the vacant seat with the marginal revenue
- Online auctions
- ...





Dynamic pricing failure stories

- DVDs from Amazon.com
 - Lost customer loyalty
- Buy.com
 - Price competion
 - Profit is low or even sometimes negative



Conditions under which dynamic pricing will be successful

- Customer must be heterogeneous in their willingness to pay
- Market must be segmentable
- Reselling at a higher price should be prohibited
- The cost of segmenting and price differentiation must not exceed the revenue due to price customization
- Customer should feel fairness in dynamic pricing
- Dynamic pricing must be based on sophisticated mathematical models.





Models for dynamic pricing

- Inventory based model
 - Models based on inventory type, inventory levels and customer service levels
- Data driven models
 - Models based on statistical techniques/machine learning that uses the data available on customer preferences and buying patterns
- Auctions
 - Models where prices vary based on the market condition
- Simulation models





NPTEL

Week 11: Lecture4

INTRODUCTION TO AUCTION





We are going to learn

- Framework for classifying auction
- Applications

E-Business



Auction

- An oldest form of market
 - A history of at least 500 BC
 - From Babylonian auction to eBay
- Theoretical studies started in 1970 with
 - Organization of Petroleum Exporting Countries (OPEC) increased the price
 - US Dept of Interior decided to action the drilling rights in the costal area
 - Economists were hired by the organizations to design bidding strategies
- Federal communication commission
 - Radio spectrum auction
 - Since 1994, the FCC has conducted 87 spectrum auctions, which raised over \$60 billion for the U.S.
- New Zealand Spectrum auction Vickrey Auction
 - Winning bid NZ\$100, 000 second highest bid NZ\$ 6 !! (1990)





A framework for classifying the auctions

- Resources
- Market structure
- Preference structure
- Bid structure
- Bidding rule
- Matching supply to demand
- Information feedback
- Nature of the good

http://www.eecs.harvard.edu/econcs/pubs/ehandbook.pdf





Classification based on the resources

- Identify the set of resources over which the negotiation is to be conducted
- Single item single unit
- Single item multiple unit
 - Multi-unit auction
- Multiple Items
 - Combinatorial auction
 - Homogeneous items or heterogeneous items
 - Sequential or simultaneous
- Items with multiple attributes
 - Pricing out mechanism for non-priced attributes through some utility function
 - Multi attribute auction





Classification based on Market Structure

- An auction is a mechanism for negotiation between buyers and seller
- One seller multiple buyer
 - Forward auction
- One buyer multiple seller
 - Reverse auction
- Multiple buyers multiple sellers
 - Double auction
 - Trading securities and financial instruments





Classification based on Bidding rules

- Ascending bid auction
 - English auction
 - Reserve price
 - Bid Increment
- Descending bid auction
 - Dutch auction
- Sealed-bid auction
 - First-price
 - Second-price (Vickrey auction)





Classification based on Preference structure

- Preference defines an agent's utility for different outcomes
- In case of multi unit auctions
 - Agent may show a decrease in marginal utility for additional units of the product
- In case of multi-attribute auction
 - Agent 's preference structure for different attributes is to be designed in terms of scoring rules used to signal information





Classification based on Bid structure

- Structure of a bid defines the flexibility with which an agent can express his resource requirement
- Ex: In single-item single-unit auction the buyer needs to specify the price
- Ex: In single-item multi-unit case the buyer needs to specify both quantity and price
- Ex: In case of multi-item case a bid may be specified as allor-nothing over a basket of items.



Classification based on Payment rule

- First price
- Second price
- All pay







Classification based on Matching supply with demand

- Market clearing or winner determination problem
- Single sourcing
 - A sorting problem
- Multiple sourcing
 - A combinatorial problem
- The problems range from simple sorting problems to NP-Hard Optimization problems

Classification based on Information feedback

- Auction protocol based on direct or indirect mechanism
- Direct mechanism
 - No feedback
 - Price signal
 - Ex. First-price sealed bid auction
- Indirect Mechanism
 - Feedback on the state of the auction
 - Price signal
 - Provisional allocation
 - English auction





Issues involved in online auction

- Strategic Issues
 - Economic Issues (Auction design issues)
 - Business Issues (Business rules)
 - **—** ..
- Implementation Issues
 - Modeling the decision making process
 - Computational Issues
 - Security





End of Week 11 E-Business



