Recommendation System

AGENDA

- The what
- Why
- Where
- How
- Types
- Deep dive into each type and live examples
- Data & Few supporting concepts

What is a recommendation system?

- A Recommendation System refers to a system that is capable of predicting the future preference of a set of items for a user, and recommend the top items.

- In a very general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries).

Why recommendation system?

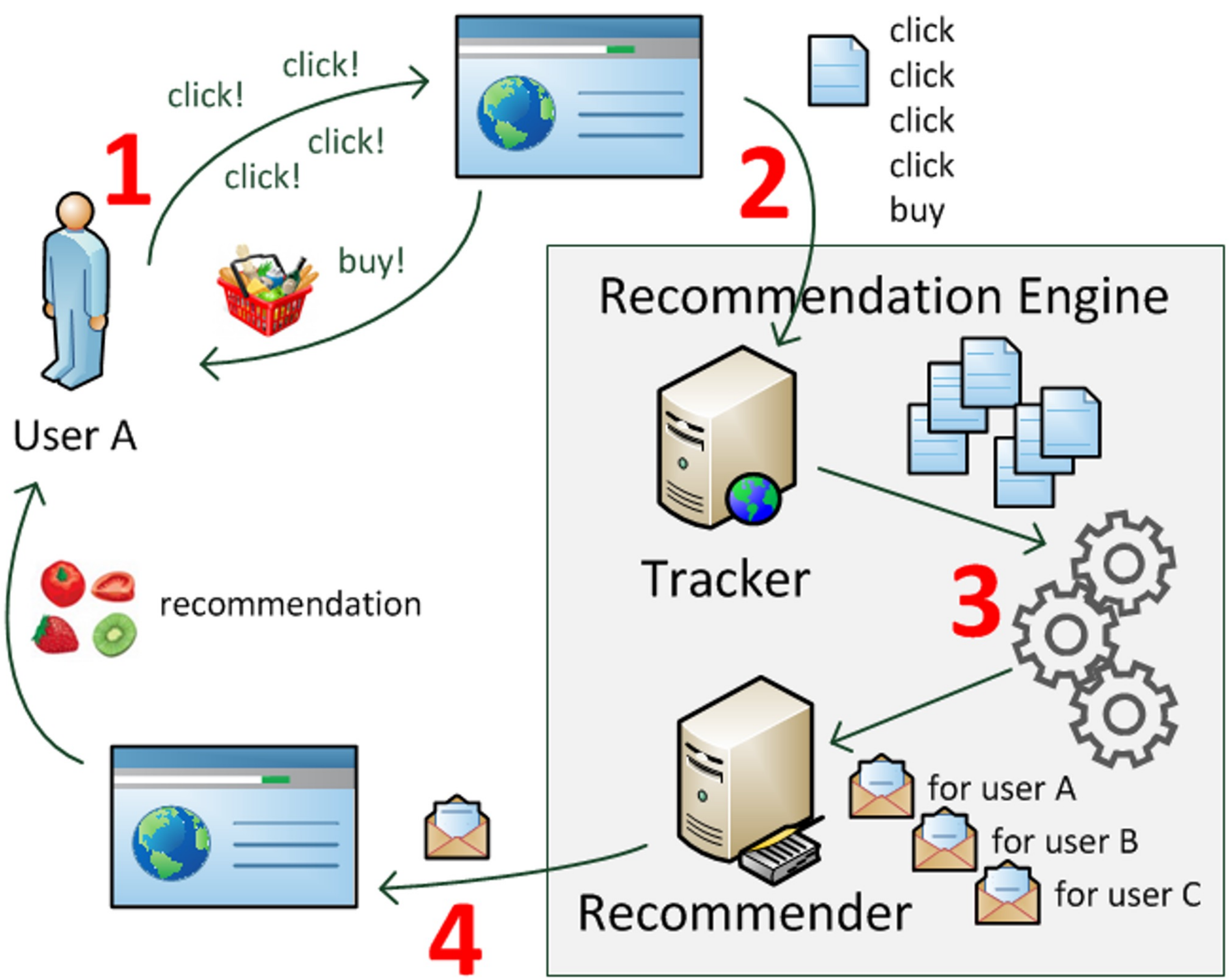
- One of the primary reasons for the necessity for a recommender system in current culture is that people have too many options due to the widespread usage of the Internet.
- People used to shop at actual stores, where the selection of things was limited.

Where is a recommendation system used?

- E-commerce
- Retail
- Banking
- Media
- Telecom
- Utilities

But what business problem are we solving?

How does a Recommendation System work?



Types of Recommendation systems

- Rule based system - Popularity, Trending by region, Buy again, New launches, Top offers, etc.
- Market based analysis – Combo offers
- Content based – similar items
- Collaborative filtering - User based and item based
- Hybrid – Ensemble - content and collaborative filtering
- Unsupervised clustering based - User behavior
- Classification based - Buying propensity model
- Deep learning based
- Graph based
- Image based (especially in fashion domain)

Rule Based

Rule Based

- Popularity
- Buy again
- Trending by region and category
- Top offers
- New launches

Popularity Based

- It's a form of recommendation system that works on the basis of popularity or anything that's currently popular.
- These algorithms look for products or movies that are currently trending or popular among consumers and then immediately recommend them.
- For example, if a product is frequently purchased by the majority of people, the system will learn that it is the most popular, and it will recommend that product to every new user who has just joined up, increasing the likelihood that the new user will purchase it as well.
- Example
  - Google News: News filtered by trending and most popular news.
  - YouTube: Trending videos.

Popularity Based

Merits

- It does not suffer from cold start problems which means on day 1 of the business also it can recommend products on various different filters.
- There is no need for the user's historical data.

Demerits

- Not personalized
- The system would recommend the same sort of products/movies which are solely based upon popularity to every other user.

Glimpse and Example of Rule Based

Market Basket Analysis

Market Basket Analysis

- Market Basket Analysis is one of the key techniques used by large retailers to uncover associations between items.
- It works by looking for combinations of items that occur together frequently in transactions.
- In the retail and restaurant businesses, **market basket analysis (MBA)** is a set of statistical affinity calculations that help managers better understand – and ultimately serve – their customers by highlighting purchasing patterns.
- In simplest terms, MBA shows what combinations of products most frequently occur together in orders.
- These relationships can be used to increase profitability through cross-selling, recommendations, promotions, or even the placement of items on a menu or in a store.

Market Basket Analysis

- The strategy is founded on the idea that customers who purchase one thing (or a combination of items) are more inclined to purchase another (or group of items).
- For example, if someone orders a sandwich and cookies on an application like Swiggy or Zomato, they are more likely to order a drink than if they do not order a sandwich.
- If the link is found to be stronger than the one between the sandwich and the drink without the cookies, it becomes more valuable.
- MBA can be used to recommend a purchase based on the lack of a typical match.
- For example, when a consumer orders only a tiny sandwich, they may be more inclined to purchase a dessert or a second sandwich than someone who purchased a huge lunch.
- Recommendation Systems that have been trained to recognize these scenarios might offer the extra things to their consumers, possibly at a discount to make the alternative more appealing.

Market Basket Analysis

- A collection of items purchased by a customer is an **itemset**.
- The probability that a customer will purchase two specific items together is referred to as the **confidence** of the rule.
- Market Basket Analysis is done using the Apriori Algorithm.
- The **Apriori algorithm** is a commonly-applied technique in computational statistics that identifies itemsets that occur with a support greater than a pre-defined value (frequency) and calculates the confidence of all possible rules based on those itemsets.

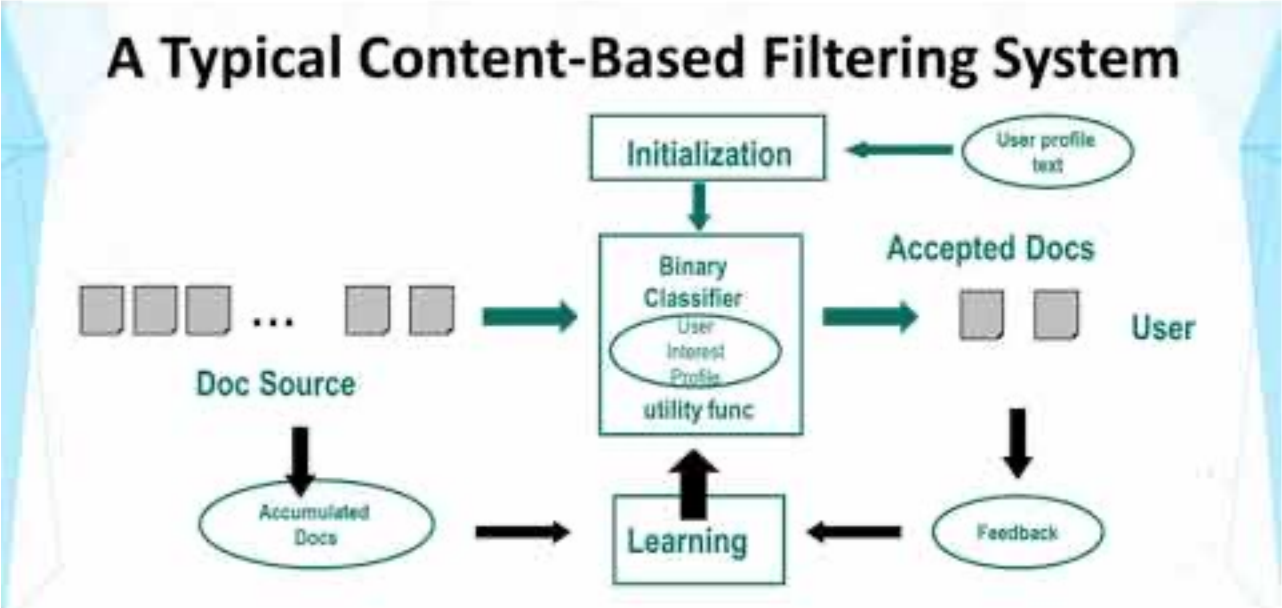
Glimpse and Example of Market Basket Analysis

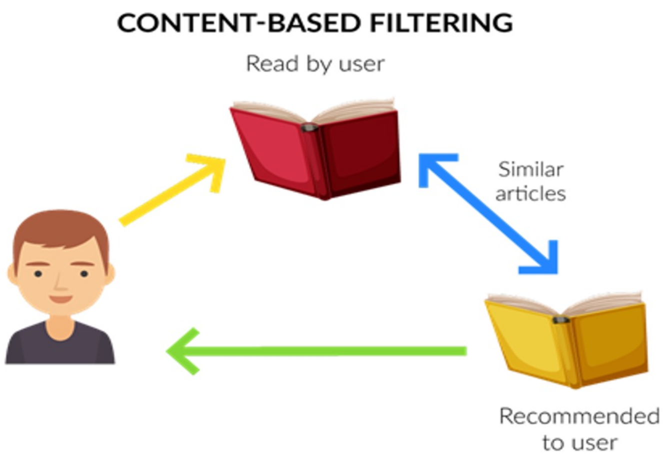
Content Based

Content Based

- Content-based filtering algorithms are based on an item's description and a user's preference profile.
- Keywords are used to define products in a content-based recommendation system, and a user profile is developed to indicate the type of item this user enjoys.
- To put it another way, these algorithms attempt to propose things that are comparable to those that a user previously enjoyed (or is examining in the present).
- Specifically, numerous candidate goods are compared to items that the user has already rated, and the best-matching items are recommended.
- This method has its origins in the fields of information retrieval and information filtering.

Content Based





Content Based

Merits

- There is no requirement for much of the user behavior data.
- We just need item data that enable us to start giving recommendations to users.
- A content-based recommender engine does not depend on the user’s data, so even if a new user comes in, we can recommend the user as long as we have the user data to build his profile. It does not suffer from a cold start.

Demerits

- Items data should be in good volume.
- Features should be available to compute the similarity.

Glimpse and Example of Content Based

Collaborative Filtering

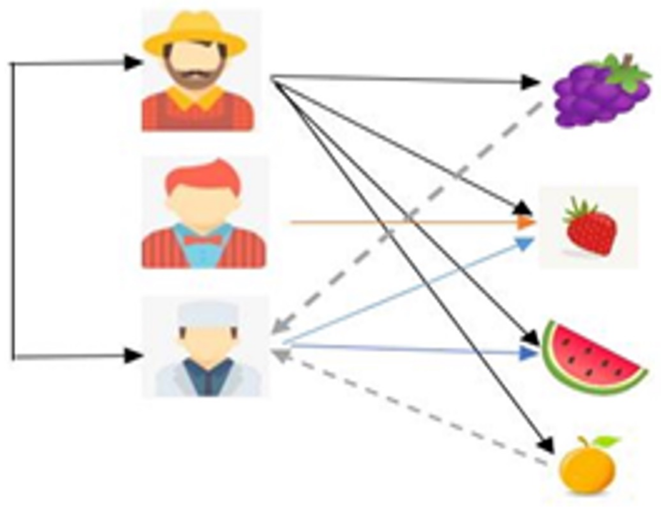
Collaborative Filtering

- Collaborative filtering approaches rely on gathering and analyzing a huge amount of data on users' actions, interests, and preferences, as well as predicting what they will like based on their resemblance to other users.
- The collaborative filtering strategy has the advantage of not relying on machine analyzable content, which allows it to accurately recommend complicated objects like movies without requiring a "knowledge" of the item.
- In recommender systems, many algorithms have been employed to measure user or item similarity.
- For example, the k-nearest neighbor (k-NN) method and the Pearson Correlation.

Collaborative Filtering

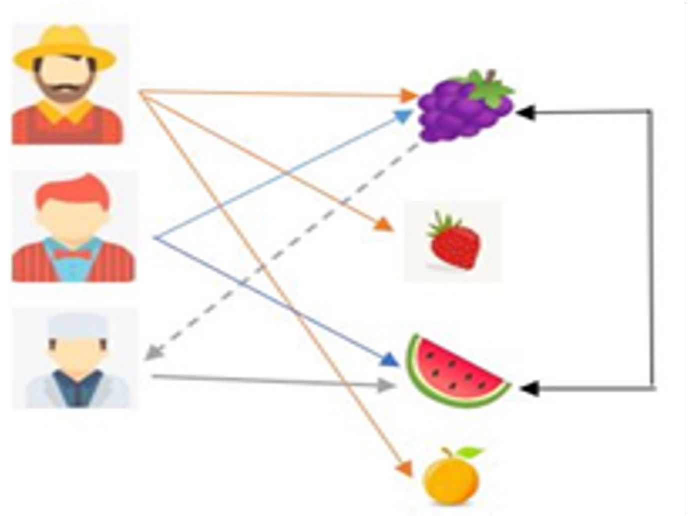
User-User Collaborative Filtering

- The system finds out the users who have the same sort of taste of purchasing products and similarity between users is computed based upon the purchase behavior.



Item-Item Collaborative Filtering

- The system computes the similarity between different items based on the items and not the users for the prediction.



Collaborative Filtering

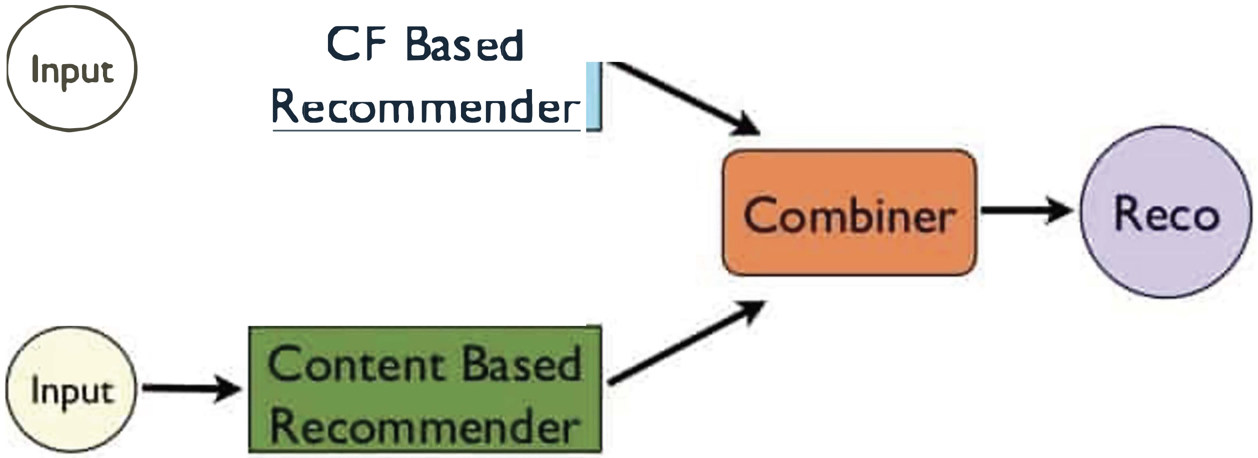
Limitations

- Enough users required to find a match. To overcome such cold start problems, often hybrid approaches are made use of between CF and Content-based matching.
- Even if there are many users and many items that are to be recommended often, problems can arise of user and rating matrix to be sparse and will become challenging to find out about the users who have rated the same item.
- Sparsity problems.

Glimpse and Example of Collaborative Filtering

Hybrid Recommendation System

Hybrid Recommendations



Hybrid Recommendation System

- Recent study has shown that in some circumstances, a hybrid strategy that combines collaborative filtering and content-based filtering is more effective.
- Hybrid techniques can be applied in a variety of ways, including separately producing content-based and collaborative-based predictions and then merging them, adding content-based capabilities to a collaborative-based approach (and vice versa), or unifying the approaches into one model.
- Several empirical studies compare the efficacy of hybrid methods with pure collaborative and content-based methods, demonstrating that hybrid methods can deliver more accurate suggestions than pure approaches.
- These techniques can also be utilized to solve some of the most typical issues in recommendation systems, such as cold start and sparsity.

Hybrid Recommendation System

A good example of a hybrid system is Netflix. They produce suggestions by comparing comparable users' viewing and searching behaviors (collaborative filtering), as well as by suggesting films that share qualities with films that a user has rated highly (content-based filtering).

Glimpse and Example of Hybrid Recommendation System

Clustering Based

Clustering Based

- Group customers based on their behavior
- Based on the cluster properties, target them with offers and recommendations

Mostly into electronics and groceries

Mostly into fashion

Mostly into luxury items

Mostly into Beauty and healthcare

Glimpse and Example of Clustering Based

Classification Model Based

Classification Based

- The model that uses features of both products as well as users to predict whether a user will like a product or not.
- An application of classification based recommendation system is the buyer propensity model.
- **Propensity modeling attempts to predict the likelihood that visitors, leads, and customers will perform certain actions.**
- It's a statistical approach that accounts for all the independent and confounding variables that affect said behavior.
- So, for example, propensity modeling can help a marketing team predict the likelihood that a lead will convert to a customer.
- Or that a customer will churn. Or even that an email recipient will unsubscribe.
- The propensity score, then, is the *probability* that the visitor, lead, or customer will perform a certain action.

Classification Based

Classification Based

Limitations

- It is a rigorous task to collect a high volume of information about different users and also products.
- Also, if the collection is done then also it can be difficult to classify.
- Flexibility issue.

Glimpse and Example of Classification Based

Deep Learning Based

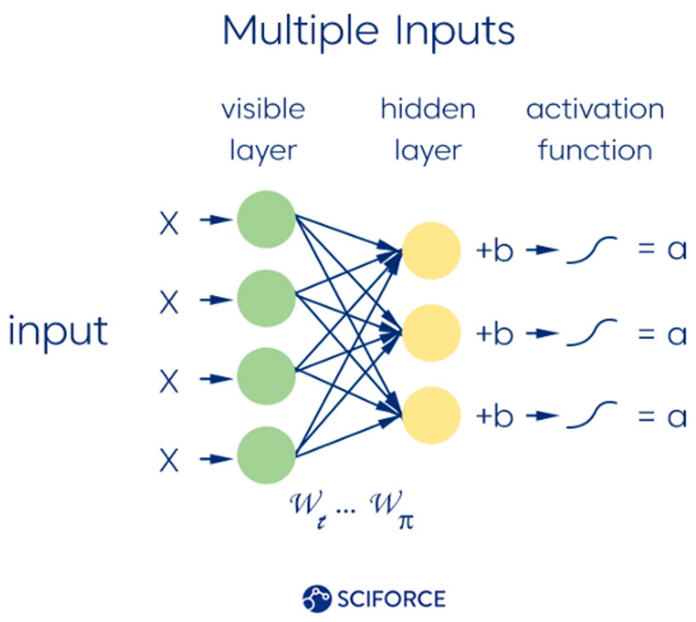
Deep Learning Based

- Deep learning (DL) is a branch of machine learning that learns from numerous levels of data representations and abstractions.
- Some tech companies are already utilizing deep learning (DL) systems based on various neural networks (NNs) to improve customer experience.
- Deep neural networks (DNNs) are used by YouTube, eBay, Yahoo, and Twitter, while convolutional neural networks are used by Spotify (CNNs).
- Meanwhile, DNNs and CNNs are merely a few types of networks that have been used, and the list of deep learning algorithms can go on and on.
- Why do we require a variety of them?
- The solution has something to do with the business domain, a specific activity, or a recommender scenario.

Deep Learning Based

Restricted Boltzmann machine based recommender systems

- The Restricted Boltzmann Machine (RBM) is one of the oldest RSs, dating back to 2007, yet it is still in use.
- RBM, in combination with collaborative filtering, was awarded the Netflix Prize in 2009 for better recommendations on the streaming site.
- Furthermore, RBM-based algorithms are still scalable to big data sets and capable of delivering high-quality item recommendations for each individual user.

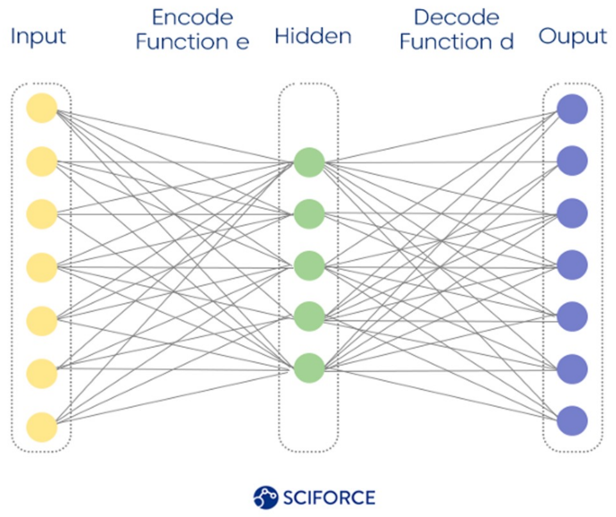


Popular Techniques for Deep Learning Based Recommendation System

Deep Learning Based

Autoencoder based recommender systems

- An autoencoder is a neural network that reconstructs its input data at the output layer.
- It has a hidden layer on the inside that defines the code that is used to represent the input.
- The autoencoder is divided into two components.
- They consist of an encoder that converts data into code and a decoder that reverses the process to recreate the input.
- The flexibility of the autoencoder in terms of data dimensionality reduction, data reconstruction, and feature extraction is what makes it so appealing.



Popular Techniques for Deep Learning Based Recommendation System

Deep Learning Based

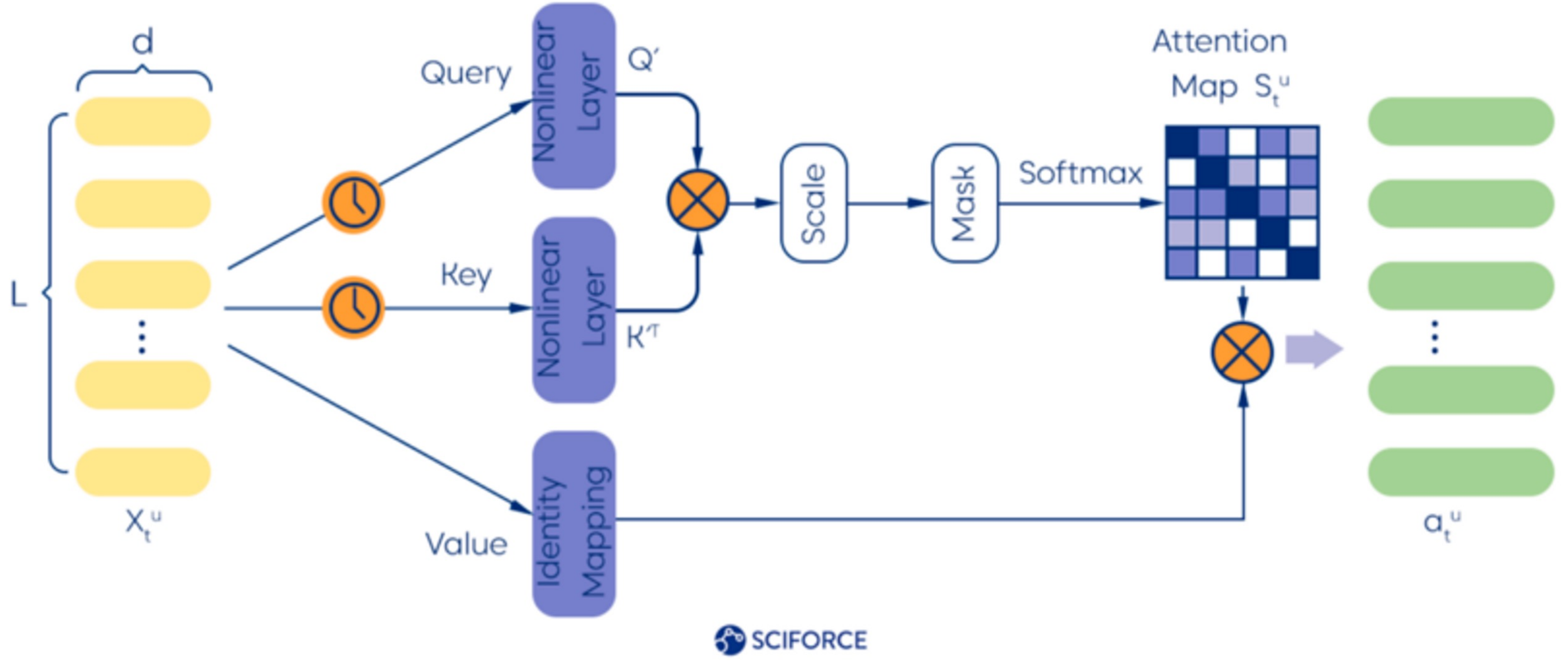
Neural attention based recommender systems

- The attention mechanism is derived from the disciplines of computer vision and natural language processing.
- Simply explained, it anticipates the next item using a vector of importance weights.
- Correlation with other elements lies at the heart of the attention mechanism (e.g., a pixel in the image or the next word in a sentence).
- The source of inspiration for this technique is, in essence, human visual attention.
- To make its next recommendation, the algorithm might "focus" on a certain element.
- Applying an attention mechanism to the recommender system can aid in filtering out irrelevant content and selecting the most representative things.
- At the same time, it has a high level of interpretability. DNNs or CNNs can also be used to merge neural attention models.

Popular Techniques for Deep Learning Based Recommendation System

Deep Learning Based

5. Neural attention based recommender systems



Glimpse and Example of Deep Learning Based

Knowledge Graph Based

Knowledge Graph Based

- Knowledge graph connects users and items through different relationships to obtain an explainable candidate list for target user, and the path between target user and recommended item is used as an explanation basis.
- Google's Knowledge Graph is a semantic network for natural language processing.
- It can be represented as a graph with nodes and edges, with the node representing an entity and the edge representing a relationship.
- The entity in the actual world could be a person, a firm, or a concept, and the relation is simply the relationship between them.
- One approach by applying Knowledge Graphs into Recommendation Systems is using the KGE(Knowledge Graph Embedding).
- By replacing the FastText embedding, or using KGE as additional feature vectors, we could improve the performance of recommendation.

Knowledge Graph Based

ARCHITECTURE

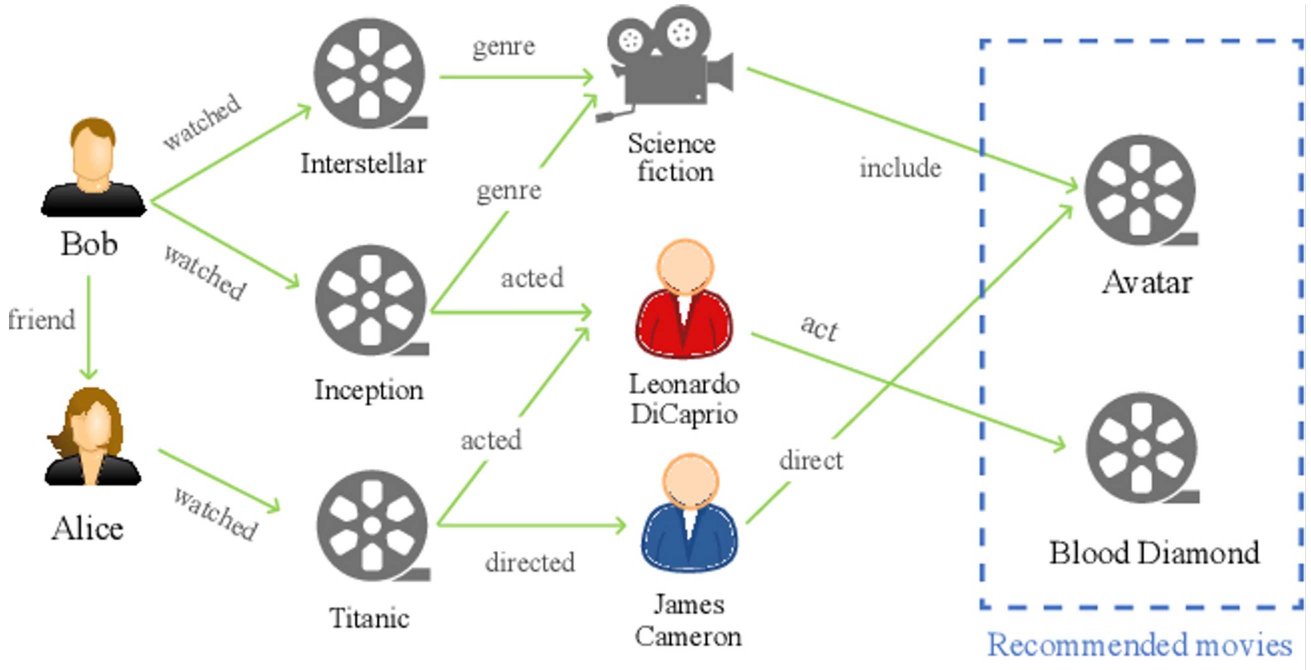
Looking inside of KG, there are 3 basic layer and they are:

- Storage layer: determine the graph data storage
- Data layer: determine the triple data structure and the entity & relation set
- Calculation layer: the optional algorithms for recommendation based on KG

ROADMAP

Stage	Task	Description
#1	Triple Generation	- define the relation & entity set - generate triples for all data - store all triples
#2-1	KGE	- extract subgraphs of KG, store in Neo4j if necessary - try embedding algorithms for KG - compare performance of different KGE models
#2-2	Knowledge Inference	- try and compare inference algorithms

Knowledge Graph Based



Glimpse and Example of Knowledge Graph Based

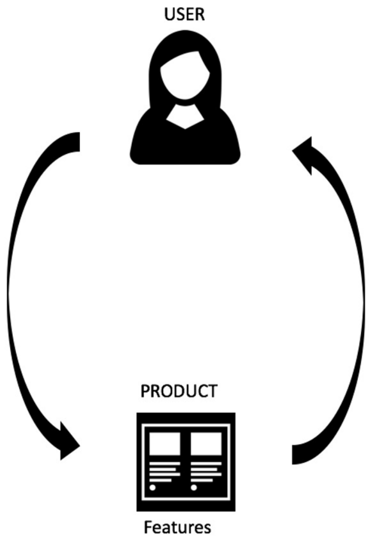
Image Based

Image Based

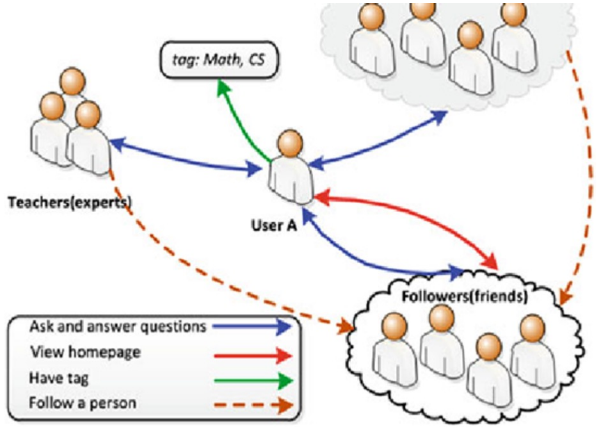
- Product Recommended based on images of the product browsed by customer
- *(Deep Learning – CNN)*
- *Can be alternative to content based. Instead of product description, you use product image.*

Glimpse and Example of Image Based

Data & Few supporting concepts



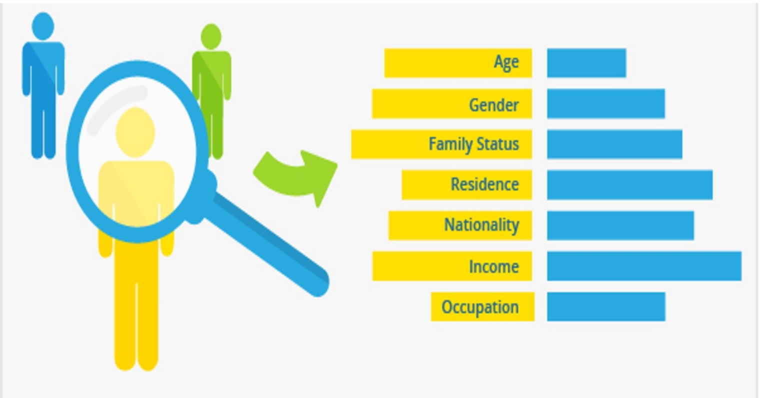
Understanding Relationships



Relationships give recommender systems a lot of information and help them understand their clients. There are three categories of occurrences:

1. **User-Product Relationship**
   - When some users have an affinity or desire for specific items that they require, this is known as the user-product connection.
   - A cricket player, for example, may have a taste for cricket-related things, therefore the e-commerce website will create a player->cricket user-product relationship.
2. **Product-Product Relationship**
   - When items are similar in nature, whether by look or description, they form product-product associations.
   - Books or music in the same genre, dishes from the same cuisine, or news stories on a certain event are all examples.
3. **User-User Relationship**
   - When two or more customers have similar tastes in a product or service, they form user-user relationships.
   - Mutual friends, same backgrounds, comparable ages, and so on are examples.

Data and Recommendation Systems



Recommender systems use the following types of data in addition to relationships:

1. **User Behavior Data**
   - Users' behavior data is useful information regarding a user's involvement with a product.
   - Ratings, clicks, and purchase history can all be used to get this information.
2. **User Demographic Data**
   - User demographic information is linked to personal details such as age, education, income, and location.
3. **Product Attribute Data**
   - Product attribute data refers to information on the product itself, such as the genre of a book, the actors of a movie, or the cuisine of a meal.

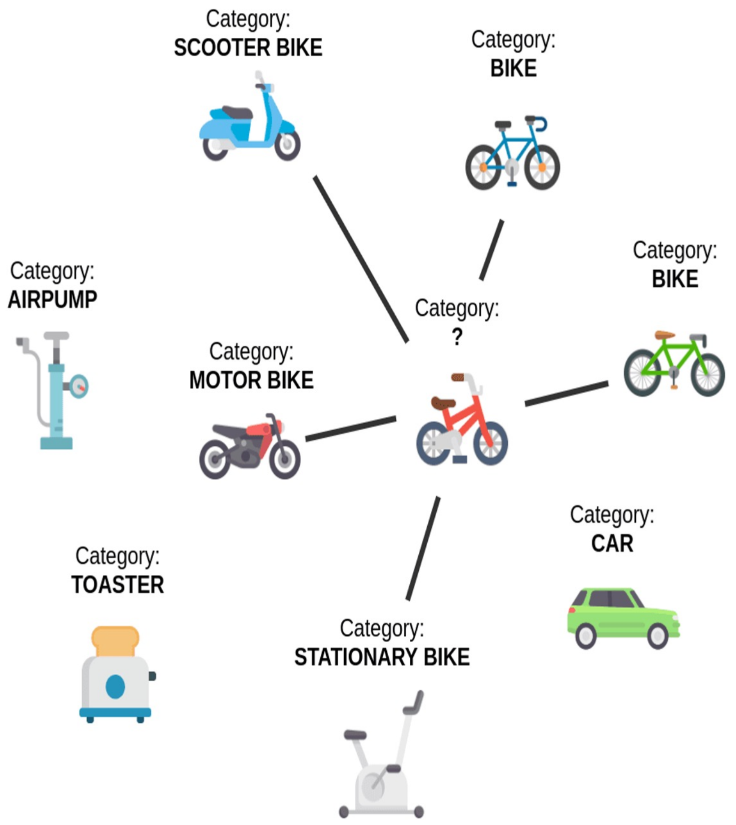
How do we provide data for Recommendation Systems?

Data can be delivered in a number of different ways. The explicit and implicit rating methods are two of the most essential.

1. **Explicit Ratings**
   - The user expresses his or her opinion through explicit ratings.
   - Explicit Ratings make assumptions about the user's preferences.
   - Star ratings, reviews, feedback, likes, and following are just a few examples.
   - Exact ratings can be difficult to obtain because people do not always rate things.

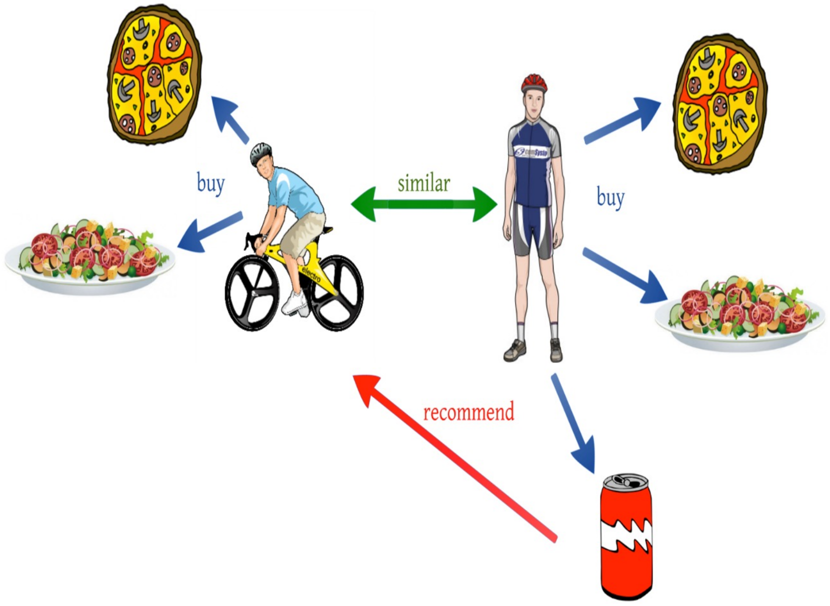
2. **Implicit Ratings**
   - When people interact with an item, they contribute implicit ratings.
   - They infer a user's activity and are simple to obtain because consumers click subconsciously.
   - Clicks, views, and purchases are all examples.

How do we provide data for Recommendation Systems?



3. **Product Similarity (Item-Item Filtering)**
   - The most useful approach for proposing products is product similarity which depends on how much the consumer would like the product.
   - Similar items can be shown to users who are browsing or looking for a specific product.
   - Users frequently expect to locate the things they desire fast and abandon their search if they are unable to do so.
   - When a user clicks on a product, we can show them another comparable product, or if the user purchases the goods, we can send them adverts or coupons for a related product.
   - When we don't know much about the user but do know what products they're looking at, product similarity comes in handy.

How do we provide data for Recommendation Systems?



4. **User Similarity (User-User Filtering)**
   - The difference between two users' similarity is checked using user similarity.
   - We can presume two users have comparable interests if they have similar product selections.
   - It's like a friend is promoting a product to you.
   - However, one drawback of user similarity is that it requires all of the user's data to propose items.
   - The problem is known as a cold start problem since it necessitates previous data from consumers to begin the suggestion process.
   - Because it has a small number of users, a freshly created e-commerce website, for example, suffers from the cold start problem.

Similarity Measures

- The distance metric is used to determine similarity.
- The points closest to each other are the most similar, while the points farthest apart are the least relevant.
- Similarity is a subjective concept that varies greatly depending on the domain and application.
- Two films, for example, are comparable in terms of genre, length, or actors.
- When measuring distance between unrelated dimensions/features, caution is advised.
- The relative values of each element must be normalized, else the distance calculation will be dominated by one feature.

Similarity Measures

Some similarity measures are as follows:

1. **Minkowski Distance**: When the dimension of a data point is numeric, the general form is called the **Minkowski distance.** It is a generic distance metric where Manhattan(r=1) or Euclidean(r=2) distance measures are generalizations of it.
   - **Manhattan Distance:** The distance between two points measured along axes at right angles. It is also called rectilinear distance, L1-distance/L1-norm, Minkowski's L1- distance, city block distance and taxi cab distance.
   - **Euclidean Distance**: The square root of the sum of squares of the difference between the coordinates and is given by Pythagorean theorem.
     Euclidean Distance: The square root of the sum of squares of the difference between the coordinates and is given by Pythagorean theorem.

Similarity Measures

2. Cosine Similarity:

- Measures the cosine of the angle between two vectors.
- It is a judgment of orientation rather than magnitude between two vectors with respect to the origin.
- The cosine of 0 degrees is 1 which means the data points are similar and cosine of 90 degrees is 0 which means data points are dissimilar.
- Cosine similarity is subjective to the domain and application and is not an actual distance metric.
- For example data points [1,2] and [100,200], are shown as similar with cosine similarity, whereas the Euclidean distance measure shows them as being far away from each other (i.e., they are dissimilar).

Similarity Measures

3. **Pearson Coefficient**:
   - It is a measure of correlation between two random variables and ranges between [-1, 1].
   - If the value is 1, it is a positive correlation, and if -1 then there is a negative correlation among variables.
4. **Jaccard Similarity**:
   - In the other similarity metrics, we discussed some ways to find the similarity between objects, where the objects are points or vectors.
   - We use Jaccard similarity to find similarities between finite sets. It is defined as the cardinality of the intersection of sets divided by the cardinality of the union of the sample sets.

Similarity Measures

5. **Hamming Distance:**
   - All the similarities we discussed were distance measures for continuous variables.
   - In the case of categorical variables, Hamming distance must be used.
   - If the value (x) and the value (y) are the same, the distance D will be equal to 0, otherwise D=1.
   - If we have data that is binary (i.e., classification), one would go for Hamming distance.
   - The lower value means high similarity and higher value means less similarity between variables.
   - For example, the Hamming distance between 1101111 and 1001001 is 3, while the Hamming distance between ‘batman’ and ‘antman’ is 2