Chapter 1

Recommender System Concepts

The rapid growth of web information provides excellent opportunities for developing various e-services applications in many domains. However, it has also caused information overload problems, whereby users are unable to efficiently locate relevant information to meet their needs when using the current Internet search tools. To help users retrieve the most relevant information from a massive amount of online information, thereby providing personalized services, recommender systems are proposed. A recommender system aims to recommend the most suitable items (products or services) to a particular user (individual or business) by predicting the user's interest in an item based on user historical records. Recommender systems provide users with personalized online information, product and service recommendations, assisting them to make decisions. Since the mid-1990s, various recommender system frameworks, methods and tools have been proposed and applied in e-commerce, e-business, e-government and other areas.

This chapter will present the basic concepts of recommender systems. The background on information overload and personalization are first introduced in Section 1.1. Then the recommender system definition, characteristics and types are discussed in Section 1.2. A general recommender system framework is given in Section 1.3. Section 1.4 presents the design and development of recommender systems, including industrial applications and the benefits of developing recommender systems. Section 1.5 summarizes this chapter.

1.1 Information Overload and Personalization

In current web systems, a huge amount of information is provided to users when they are making decisions about any kind of product or e-service. Usually, users are unwilling to explore the vast amount of information offered by companies in relation to their products and services and, choose a product or service in which they are truly interested. Thus, in this fiercely competitive marketplace it is crucial for a company to help its customers deal with information overload in decision-making so as to retain their loyalty [Schafer *et al.* (1999)]. Providing personalized e-services to customers is an appropriate approach to handle the information overload problem and improve user's experience. In this way, services to different users are customized, making it easy for customers to find what they need.

Providing customized products and services to customers, while they are basically being passive, is known as personalization, and is becoming a key factor for a company to satisfy their customers. This is where recommender systems come in. Using recommender systems to provide personalized e-services has become the most commonly used technique to solve the information overload problem that has been brought about by Web 2.0 [Adomavicius and Tuzhilin (2005)]. The recommendation process predicts a user's potential interest (user's preference) in items that they haven't previously bought, according to the user's historical records, thus creating a personalized list from which users can choose.

Recommender systems were first applied in e-commerce to solve the information overload problem, and were quickly expanded to the personalization of e-government, e-business, e-learning and e-tourism [Lu *et al.* (2015b)]. Nowadays, recommender systems are an indispensable part of Internet websites such as Amazon.com, YouTube, Netflix, Yahoo, Facebook, Last.fm and Meetup.

1.2 Definition, Characteristics and Types of Recommender Systems

In this Section we provide the definition of a recommender system. Then we list its characteristics. Finally we introduce the four different types of recommender systems.

1.2.1 *Definition*

A recommender system can be defined as a set of programs that attempt to recommend the most suitable items to particular users by predicting a user's interest in an item based on the information about the items, the users and the interactions between items and users [Bobadilla *et al.* (2013)]. The users could be individuals or businesses, such as book buyers, job seekers or small businesses who want to find a potential partner. The items could be either products or services, such as a book, a movie or a package of mobile services, as shown in Fig. 1.1.

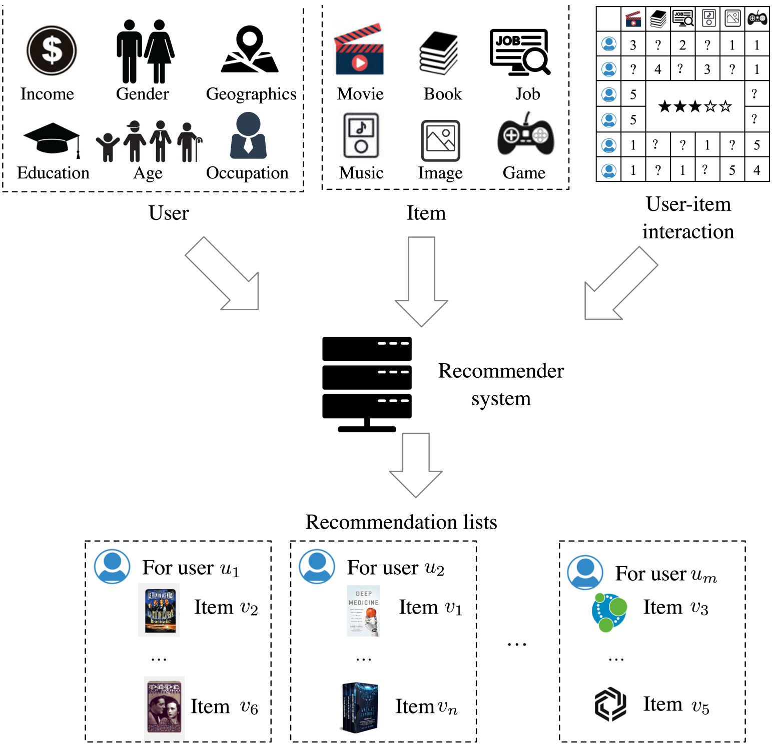

Fig. 1.1: A graphical illustration of a recommender system.

In brief, recommender systems are designed to estimate the utility of an item and predict whether it is worth recommending. The core part of recommender systems is a function to define the utility of a specific item to a user [Adomavicius and Tuzhilin (2005)]:

$$f : \mathcal{U} \times \mathcal{V} \rightarrow \mathcal{D} \tag{1.1}$$

This is a function to define the utility of a specific item $v \in \mathcal{V}$ to a user $u \in \mathcal{U}$. $\mathcal{D}$ is the final recommendation list containing a set of items in a ranked order. This list is ranked according to the utility of all the items the user has not consumed.

The key purpose of most recommendation methods is to find an item for the user so as to maximize the utility function, formulated as follows:

$$\forall u \in \mathcal{U}, \ \underset{v \in \mathcal{V}}{\operatorname{argmax}} f(u, v) \tag{1.2}$$

A final recommendation list, containing a set of items in a ranked order, will then be provided. Essentially, the utility of an item is presented as ratings of a user. A rating given by user u to an item v is $r_{u,v}$. With the number of users M and the number of items N, the rating matrix is $\boldsymbol{R} \in \mathbb{R}^{M \times N}$. The recommendation task is to either fill in the rating matrix or give a ranking score for each user-item pair. Thus the items that the user has not seen before can be ranked and the Top-K items will be recommended. Predicting the utility of items to a particular user varies in different recommendation methods. Also, other information, for example item attribute information, user-generated information and user social information, have been used in different methods.

1.2.2 *Characteristics*

According to the definition of recommender systems and their various applications, we list the main characteristics of a recommender system below:

(1) It mainly deals with information overload problem;
(2) It supports both individual and group users in decision making;
(3) It provides users with personalized services;
(4) It usually generates recommendation without additionally requiring information from users;
(5) It supports a variety of business use cases and scenarios such as data uncertain and data sparsity;
(6) It can access a wide variety of data sources across multiple domains simultaneously;

With these characteristics, we can develop different types of recommender systems, use different estimation and prediction models, and apply them in different domains for various users to help them make decisions.

1.2.3 *Types*

Recommendation methods are the key to implementing a recommender system. In principle, the methods can be classified into four types: collaborative filtering (CF)-based, content-based, knowledge-based and hybrid recommendation method. Each recommendation method has both advantages and limitations. For example, CF-based methods require the rating of user data and so it is hard to deal with new users or users who do not have enough rating data (known as cold start or data sparsity problem). However, advanced developments in recommender systems to overcome disadvantages and provide better user experiences are growing each year. Here we give a brief overview of the four recommendation method types. More detail can be found in Chapter 2.

(1) CF-based recommendation methods

CF-based recommendation methods help users to make choices based on the opinions of other users who share similar interests. The CF-based methods can be divided into memory-based CF and model-based CF. Memory-based CF methods contain user-based and item-based CF [Sarwar *et al.* (2001)]. In the user-based CF, a user will receive recommendations of items liked by similar users. In the item-based CF, a user will receive recommendations of items that are similar to those they have loved in the past. The similarity between users or items can be calculated by different measures.

With the development of machine learning techniques, model-based CF methods are built on the basis of optimizing an objective function between the model prediction and the true label. The model-based CF is more superior, as it benefits from various artificial intelligence (AI) techniques. To overcome the disadvantages of prediction in one way, multi-criteria CF has also been developed.

(2) Content-based recommendation methods

Content-based recommendation methods focus on recommending items that are similar to items previously preferred by a specific user [Ricci *et al.* (2010)]. The basic principles of content-based recommender systems are: (1) To analyze the description of the items preferred by a particular user so as to determine the principal common attributes (preferences) that can be used to distinguish these items. These preferences are stored in a user profile. (2) To compare each item's attributes with the user profile so that only items that have a high degree of similarity with the user profile will be recommended.

Content-based recommender systems generate recommendations heuristically using traditional information retrieval methods such as the cosine similarity measure. Also, they generate recommendations using statistical learning and machine learning methods, which are capable of learning users' interests from the historical data of users. One disadvantage of content-based recommendation methods is overspecialization, since any recommendation is provided based on the user's historical consumption records. Details of content-based recommendation methods will be discussed in Chapter 2.

(3) Knowledge-based recommendation methods

Knowledge-based recommendation methods offer items to users based on external knowledge about the users, items and/or their relationships. Usually, knowledge-based recommendation methods retain a functional knowledge base that describes how a particular item meets a specific user's need. This is performed based on inferences about the relationship between a user's need and a possible recommendation. Case-based reasoning (CBR) is a common expression of knowledge-based recommendation methods in which items are represented as cases and recommendations are generated by retrieving the most similar cases to the user's query or profile [Brusilovski *et al.* (2007)]. Ontology, as a formal knowledge representation method, represents the domain concepts and the relationships between those concepts. It also has been used to express domain knowledge in recommender systems. The most obvious disadvantage of the knowledge-based recommendation method is the human intervention required to acquire the knowledge and build the knowledge base.

(4) Hybrid recommendation methods

To overcome the disadvantages of using only one type of recommendation method, various hybrid recommendation methods, which combine two or more recommendation methods, have been developed [Burke (2002)]. The most common practice is the combination of CF-based and content-based methods, including user-based CF with content-based and item-based CF with content-based. Knowledge-based recommendation methods have also been widely combined with CF-based and content-based recommendation methods.

1.3 Framework of Recommender Systems

The framework of recommender systems contains three main components: data sources (input), recommendation engine (core technique) and recommendation generation (output). A recommender system can use data from various sources. This means the information about a user, an item and the interaction between them can be taken care of by recommender systems as shown in Fig. 1.2. Then, the recommendation engine processes and generates results to fit diverse requirements. The output of the recommender system varies in accordance with the application scenarios.

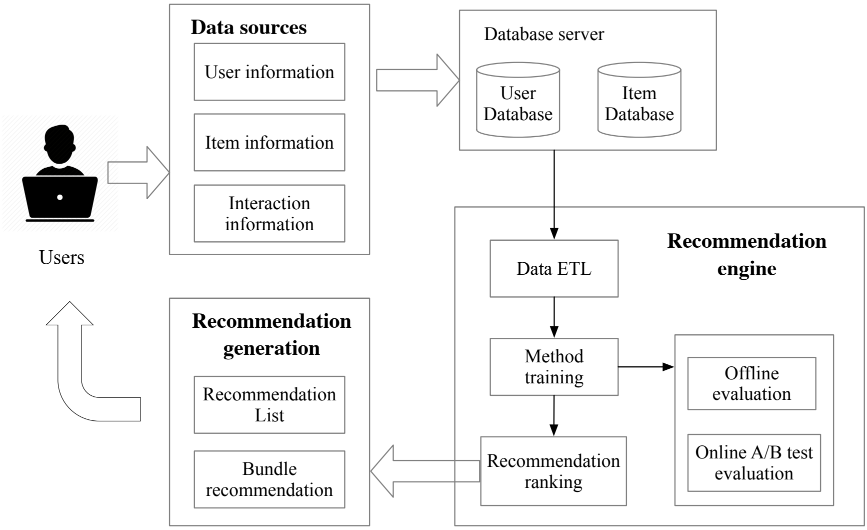

Fig. 1.2: A general recommender system framework.

1.3.1 *Data sources*

Data about users, items and their interactions come from different sources and can be in various forms. We will first give a detailed description of the data the recommender system can make use of.

Users have different preferences and various purposes. In order to provide users with personalized recommendation, the recommender system needs to profile each user using the user's related information. For example, when a user is selecting a movie on a website, the available information of that user includes: demographic information such as age, gender, nationality, the language he/she speaks, search history, clicked movies, watched movies, user social relations or user generated tags or comments. All of the above can be used to profile the user's preferences. The process of data selection and the way the data is used vary based on the recommendation method. Also, users can be individuals or something more complex such as a group or a business. For example, a user can be a channel group for online videos or a company where each member may have different purposes.

Items refer to anything that needs to be recommended to users. The attributes of an item are not just text descriptions, but also comprise images or knowledge graph between a set of items. The attributes of an item can greatly affect the recommendation methods. For example, groceries can be recommended to users repetitively and music or news can be consumed in large quantity. On the other hand, luxuries, such as a mobile phone or a house, may be consumed by a user with careful considerations. Additionally, some complicated items where knowledge and rules accounted for a large proportion in recommendation such as jobs, hotels or insurance policies.

The interaction between users and items contains all the information the user has left behind after browsing the web or using an App. This information can be in different forms such as ratings, tags or free-text comments. In a broad sense, the user feedbacks can be divided into explicit feedbacks and implicit feedbacks:

(1) Explicit feedbacks refer to the rating stars given by the user as $\star$ to $\star\star\star\star\star$ as the ordinal rating $\{1,2,3,4,5\}$. In such case, the rating contains both positive and negative feedback from the users.
(2) Implicit feedbacks refer to the binary values given by user behaviors such as whether the user has viewed, clicked or bought an item or not. The implicit feedbacks only represent positive feedbacks - users' negative feedbacks are not collected.

1.3.2 *Recommendation engine*

The recommendation engine is the core part of the recommender system. First, data from various sources need to be processed by data extract, transform and load (ETL). Then structured data, such as user features and item features, are stored in the database. After data are processed, the engine is to train the recommendation methods. The recommendation engine contains recommendation methods that take those input data to classify, measure and match users or items; predict a score indicating preferences; and generate recommendation lists.

We now describe the two kinds of evaluations for the recommendation engine: (1) Offline evaluation: Data engineering techniques and recommendation methods are validated and evaluated with offline data where they are selected for online serving according to their performance on offline evaluation criteria. (2) Online A/B test evaluation: Except for offline evaluation, online A/B test is also needed for a new recommendation method before it is deployed. Two groups of users (corresponding to A and B) are randomly selected and are provided with recommendations generated by two different methods. The feedbacks of users from group A and B are collected and the significance of the difference is statistically analyzed as a "two-sample hypothesis test" to evaluate which recommendation method is more effective.

1.3.3 *Recommendation generation*

Recommendation results are usually generated as a ranked list of items and are presented to users to help them make decisions. Moreover, some recommender systems can bundle related items and give a package for the users to choose from [Zhu *et al.* (2014)]. The bundle recommendation can not only improve the sales of items, but also integrate price strategy where users are offered diverse choices and discount promotions. All in all, recommender systems can be used in various scenarios and situations to offer customers better experiences and companies commercial value through their ability to filter information.

1.4 Design and Development of Recommender Systems

Almost every website is powered by a recommender system. It has become an indispensable part of e-commerce in industry. In this section, we introduce the design and development of recommender systems in industry. Firstly, we give a general introduction about the industrial application of recommender systems. Then the benefits of developing recommender systems are listed.

1.4.1 *Applications*

Recommender systems are adopted in many industrial companies with applications in different areas. The business target is decided by the uniqueness of each application area. As a result, the constructed recommender systems vary according to the different users, items, contexts and business problems. In this Section, we give a general system design of the recommender system used in industry, ignoring the techniques applied. Also, we review several well-known applications in industry, as examples of how recommendation methods help to build the system and how industry benefits by building recommender systems.

The first application of the recommender system in a large company was at Amazon [Linden *et al.* (2003)]. The item-based CF method is employed in Amazon as part of its online shopping system, and deployment of the recommender system has helped Amazon increase its profits around 30% (analyzed by McKinsey & Company in 2017). The success in Amazon has attracted other e-commerce companies to attach importance to the application of recommender systems, including eBay [Schafer *et al.* (1999)] and Tmall [Wang *et al.* (2018a)]. Embedding and graph embedding techniques are applied to process the huge amount of items and complex relationships between them. The recommender system is now an indispensable part of online shopping websites.

E-entertainment followed the step of e-commerce with the well-known Netflix Prize [Bennett and Lanning (2007)]. Video recommendation taking various data sources from structured meta data such as actors, titles to unstructured data such as images and videos are applied in YouTube [Davidson *et al.* (2010)]. Similarly, the music genome project is launched where features are extracted containing information about the music artists, band, albums, genre, style and emotion are extracted. These features are integrated with the CF recommendation method which generates music play lists for users. Also, social platforms, such as Twitter and Facebook, connect users with artists and their friends. Spotify deployed Echo Nest for powering the recommendation in its music streaming platform in 2014 [Eriksson *et al.* (2019)]. With Apple Music entering the competitive market, music platforms all recommend a play list for users once they launch the App or website. These e-entertainment items can be repetitively consumed by users, particularly for music, and recommender systems are developed to fit this requirement.

Recommender systems are also directly connected and widely used with ad promotions. The click through rate is the metric that is directly linked with the profits from advertising revenues. Facebook uses logistic regression and boosted decision trees to improve the click through rate in ad recommendation [He *et al.* (2014)]. LinkedIn provides personalized feed to users to improve user participation on the activities [Agarwal *et al.* (2015)]. The click through rate is also a crucial indicator for evaluation on news recommendation, which is used by Google and Yahoo [Das *et al.* (2007)].

Except for recommender systems on single domain, some companies also develop cross-domain recommender systems. The cross-domain recommender systems can leverage a source domain containing rich information to alleviate data sparsity problem. They can also help to promote a combination of items from several domains and increase the diversity of consumption. Microsoft has developed cross-domain recommender systems by using Apps and News together with Movie/TV on Xbox to generate recommendation to users [Elkahky *et al.* (2015)].

Current development in the industry shows that recommender systems have become one of the most important business intelligence techniques. Because of the opportunities they have provided in the area of promotions, recommender systems are connected with the profit of companies. At the same time, recommender systems play a crucial role in improving user experience.

1.4.2 *Benefits*

The advantages of recommender systems for both customers and companies are summarized below.

(1) Personalized services.
One of the key benefits recommender systems provide is a personalized service for users. Since users are limited to their own knowledge and experience, the personalized service of a recommender system can help users quickly find what they need while discovering new items which they did not know about previously. This reduces the amount of time the user needs to spend searching and exploring.

(2) Improved user profile.
If recommender systems have the ability to give users personalized recommendation, users need to be profiled accurately. The recommender systems can assimilate the various pieces of information about a user, including demographic information, browsing history, user preference interactions and business knowledge, to create user profiles. These user profiles can then be utilized for business analysis.

(3) Retain the loyalty of users.
Recommender systems provide a guide for users when they are browsing the website. They can make it easier for customers to find what they need and limit the time spent completing an order. In such case, users' needs are satisfied and their experience of using the website or service are improved. This can encourage users to stick with the service based on it being user-friendly and helpful. The user experience is crucial to companies in keeping user loyalty and capturing market share.

(4) Increase the revenue.
Recommender systems provide a guide for users when they are browsing the website. They can make it easier for customers to find what they need and limit the time spent completing an order. Users can be swamped by the excessive number of product choices, particularly when there are long-tail products. Users are more likely to discover diverse range of items that they are more likely to be interested in.

(5) Understanding market trends.
A secondary but integral benefit of the recommender system is to provide reports on sales direction and market trends. These reports and analysis can help a company make decisions about the direction to drive the market.

1.5 Summary

The development of recommender systems aims to provide personalized items (products and services) to users (business or individual). Recommender systems, as one of the hottest areas in artificial intelligence, have brought great profits to companies who are prompting frontier development with more sophisticated techniques to support consumers. This chapter presents the definitions, types of recommendation methods and the general framework of the recommender system. It also presents the design and development of recommender systems in practice.