Delivering Relevant Product Recommendations in Finance*

Mykhailo Chepara

Slovak Technical University in Bratislava

Faculty of Information and Information Technologies

`xchepara@stuba.sk`

October 1 2024

Abstract

To begin with, why is that topic? One of my friends is working in a financing company, therefore I'm genuinely curious how does a bank recommend products to it's customers. As well as how do some systems recommend stock recommendations to investors. It's a useful tool for bank to increase it's revenue. For an investor a significant instrument that can offer good offers on a trade market. Can it be enhanced? And if yes then how?

Now to the points that I'm going to untangle in this article:

- What is a recommendation system?
- Distribution of the systems based on their category
- Usage cases in real life
- Architecture and concept of work
    - Architecture and ways of utilizing data
    - Algorithms applied and models used
    - Data fetch and processing
- Conditions to be met before deployment of the system
    - Learning conditions
    - Duration of learning
    - And other conditions
- Issues with the system
    - Lack of data
    - New item introduction
    - Inability to recommend anything relevant to a user
    - And other issues
- Enhancement of the system
    - Augmentation of one system with another
    - Systems that can be merged with recommendation system in finance
    - Other improvements

[15]

Contents

1   Recommendation system

1.1   Definition of a recommendation system and its objective

Nowadays, people buy a lot of things on the internet, they watch a lot of films and do a lot of other stuff. Since there's such a rich number of choices, it's hard to find the ones people would choose. For that reason, a recommendation system was introduced. A recommendation system, also called, recommender system, is a program whose main point is to recommend users a product/service which he/she most probably will buy. Ultimately, it tries to predict the next probable choice of a customer and then recommends it. It could be the next film watched, the next purchase bought or the next route on the road recommended. The data is provided from the customer's overall experience. So it's a system, adapting to each customer's needs and preferences. [13]

1.2   Recommendation system in Finance

The system we're going to talk about is a system which recommends offers on stock or any other liquidity markets. Since a lot of people buy stocks to earn money, and a lot of them would be happy to know where prices on the stock market will go, a necessity in finance recommendation systems has appeared. It's hard to predict stock prices: it depends on factors such as inflation, supply, demand, company's state, economic reports, news, investor's sentiment etc. Adding to that probability of factors influencing each other, prices can strongly fluctuate. Offers are chosen according to the certainty of a successful sale. Thus, it's going to be easier to buy and sell them. [9] [3]

1.3   Distribution of the systems based on their category

The systems are divided into few main categories. Some of them are content-based systems, collaborative systems and hybrid systems. Here is a graph of most famous ones

Here are the most widely used recommendation systems today.

Collaborative systems is one of the most used types of recommendation systems in the world. Collaborative system recommends something based on preferences of another user which has common traits with the current one. Advantages are filtering complex information, recommending completely new information, utilizing feedback of similar users. Disadvantages are rarity of scores, multiple content, scalability.

Content-based system is the most used type of system in academic search engines. The plain idea of it is to record which object the user has chosen from the list of recommended products and then analyze it, based on the most relevant object next. For example, academic, fresh attributes. The point is to recommend an object which has the biggest similarity in terms of description to the description provided by the user. Advantages are cold start, ability to adapt to user's preferences, no problem with new objects recommendation. Disadvantages: in case of a user's account deletion, his preferences are gone, it might recommend the same product multiple times, and lack of history.
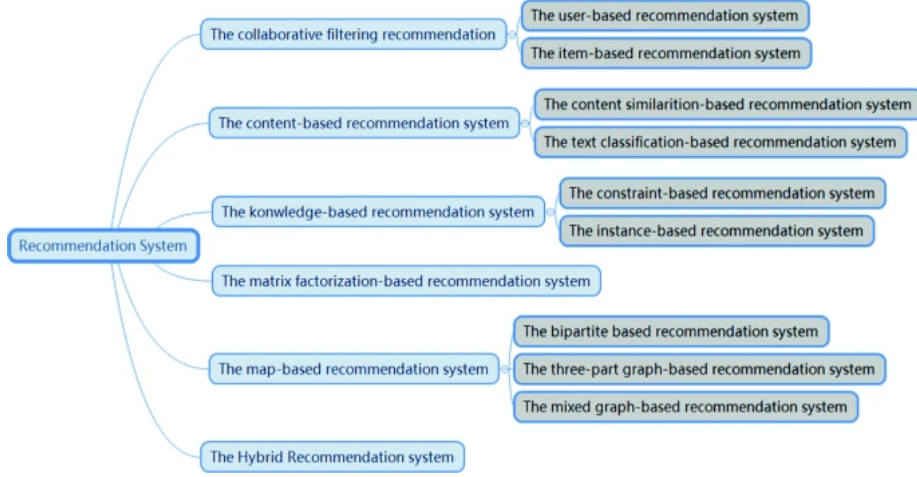
Figure 1: Distribution of systems based on their category
[17]

Knowledge-Based Recommendation system is a type of recommendation system which uses auxiliary information about products and users to give out appropriate results. This kind of recommendation system has an advantage against regular systems are content-based cold start. Advantages: history of ratings isn't required, recommendations aren't dependent on user's preferences, so user's data isn't needed. Disadvantages: knowledge acquisition bottleneck, lack of needed information.

Hybrid-Based Recommendation system consists of multiple system types which cancel out each other's flaws. It has 7 main strategies for filtering. Here are a few of them.

Switching: this method basically selects a recommender system based on its accuracy. Feature combination: ultimately, one system is complemented by features of another. Feature Augmentation: each time a new item is recommended, new feature is offered based on which next recommendation is made, so each new recommendation is made using previous features. Meta-level: recommendation system uses an input output of another system.

1.4   Usage cases in real life

Here are the main variations of recommendation systems which are most commonly used.

Spotify, Apple Music, YouTube music and other platforms use recommendation systems to match their user's preferences and show appropriate results for search requests. In that particular field of human art, collaborative filtering and content-based filtering is used in most cases. [7]

Google Play Books, Open Library, and Goodreads are apps that manifest the power of recommendation system for recommending books, articles and other handwritten works to their customers. Matrix factorization - collaborative filtering algorithm which is used for this scope as a default decision for recommending desired results. [11]
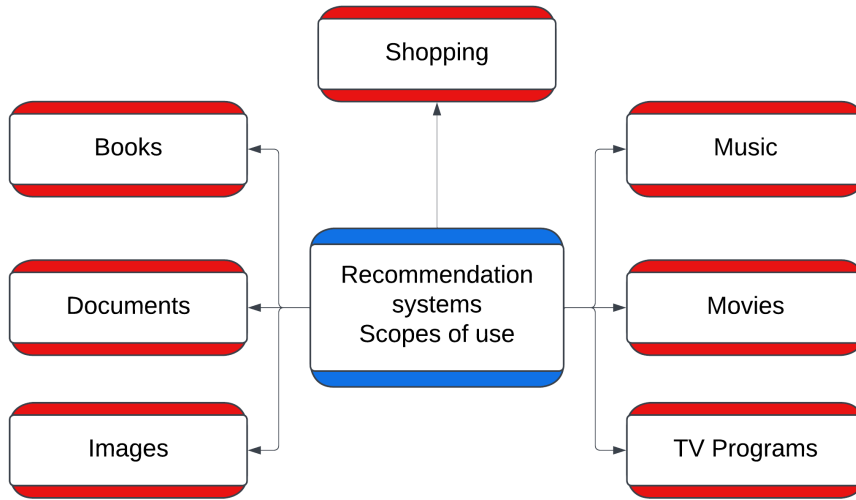
Figure 2: Most commonly used system types

Microsoft Word, Excel, PowerPoint, all of them use file recommendation in order to keep track of important work by recommending files which were frequently edited, commented or people were mentioned in them. [10]

Netflix, Amazon Prime Video, and Hulu use, systems recommendation systems to recommend movies, TV shows and animes to their clients, which increases profits significantly. For example, Netflix reported saving $1 billion dollars by engaging customers with recommended movies. Techniques of collaborative filtering, content-based filtering, and hybrid filtering are used to generate precise apposite recommendations to a customer. [6]

Amazon, Target and L'Oréal are companies which honed their recommendation systems to the highest possible precision to gain the most revenue. Amazon even made its own personalization system called Amazon Personalize systems applies machine learning algorithms to sort out necessary data and use it for systems. In general, it uses a collaborative filtering method. [16] [14]

2 Architecture and concept of work

Recommendation system in finance can recommend products as well as it can recommend offers which can lead to revenue of individuals. Here's an overall architecture of system capable of recommending fitting stock recommendations

2.1 Data fetch

In a nutshell, data is collected as a first step. News data is sent into a NLP(Natural Language Processing) preprocessor which decodes sentences and assesses whether news is good or bad. Later conclusion of news processing is used in evaluation of offers. Subsequently, stock market data is analyzed and the conclusion about market volatility is made. Volatility is considered a measurement of the speed

Figure 3: Architecture of finance recommendation system

at which prices and their ranges change. And then all the data left is separated into training and test data. After which training is made and then, the system recommends the most favorable offers on the market, also including investor's preferences.

Data is collected using API (Application Programming Interface) from trustworthy news sources such as Yahoo Finance, American Stock Market and Alpha Vintage, CNBC, Stock Market and many others. [2]

Steps for an API to parse data from news websites and stock market data providers. Firstly, it authenticates, uses tokens and API keys to access a website's data, so nobody else will get to it. Then it configures and sends what type of data will be retrieved. Based on the response from the server, it either sends back another request or waits in case of the server being down. Then it receives requested data and saves it into an already existing database.

2.2   Data processing

Data variables retrieved from stock market providers are: Date, Open, Close, Volume, High, Low, Adjusted close. Percentage Change (Returns) is a crucial variable which predicts how much the price has grown or declined. It measures the volatility of the market. So, to summarize, text data types and number data types are used in calculations for a recommendation. Text obtained from news is preprocessed. Preprocessing consists of a few stages:

- Text cleaning, which includes stop word removal, lowercasing words, removing characters, removing numbers, removing unnecessary words.

- Normalization is stemming and lemming words. It essentially means to get the root of each word and normalizing words according to their dictionary forms.

- Tokenization. Ultimately, it's dividing text into meaningful sections such

as paragraphs, sentences or words. In our case, it's words. So each part is called a token.

- Feature extraction is achieved by vectorization. Vectorization means that each token is converted into a corresponding numerical value.

- Data transformation. This stage is optional. PCA (Principal, component analysis), t-SNE and UMAP (Manifold Learning), ICP (Independent component analysis), LDA (Linear Discriminant analysis) and other techniques can be used to decrease dimension of feature set matrix. [12]

- Splitting data for analysis. Data is split into training and testing data. [5] [8]

For example, in the article about stock recommendation systems [9] at the data preprocessing stage there were used methods and algorithms: Stemmer Algorithm, WordNet Lemmatizer, and Stop word removal. After, what data was vectorized using the TF-IDF(Term Frequency Inverse Document Frequency) vectorizer. [1] [4] TF-IDF calculates the tangency value for each word in all the documents in our case datasets. Term frequency(TF) is calculated using the formula:
$$\alpha = \frac{\beta}{\gamma}$$
Where α represents term(word) frequency. β is number of occurrences of word in the document and γ is total number of words in the document.
Inverse document frequency(IDF) is calculated using the following formula
$$\sigma = \log(\frac{\rho}{\epsilon} + 1)$$
Where σ represents inverse document frequency value. ρ is total number of documents and ϵ is number of documents in which our word occurs.
And so the TF-IDF is represented by this formula
$$TF - IDF = \alpha * \sigma$$

Main analysis of the data can be conducted using Random Forest Algorithm, LSTM model and Logistic Regression. Nevertheless, it could be done differently using other ANNs(Artificial Neural Networks) and RNNs(Recursive NEural Networks) complemented by Naive Bayes classifier or SVM(Support Vector Machines).

2.3 Algorithms and models used

2.3.1 Models used to predict prices on the market

Now for the data analysis of the stock market dataset. There are lots of models for analyzing and predicting data, but in our case, to foretell prices in such an unpredictable market, sophisticated models are required, which must forecast the volatility of the market. Furthermore, they must know how to work with time series.

Here are some examples of models that can be used for prediction.
Statistical models:
ARIMA(AutoRegressive Integrated Moving Average) models. They work well with a series of data points. By finding, dependency between them, they find

the volatility rate of the market.
GARCH(Generalized Autoregressive Conditional Heteroskedasticity) are also
used for forecasting market volatility as well as the degree of risk.
Deep learning Models:
ANN(Artificial Neural Networks) can learn complex relationships and are made
of several layers of interconnected points(nodes). But are not suitable for un-
predictable market.
RNN(Recurrent Neural Networks) are designed to work with successive data.
But they're suffering form vanishing gradient problem
LSTM(Long Short-Term Memory) model is a type of RNN which has resolved
the issue of vanishing gradient. It fits the best among models to predict prices.
There are, of course, other model types, but we'll stop on the LSTM model,
since it's the most efficient and suited for our purpose.

LSTM model is made of three main parts named gates. Input gate, forget
gate and output gate. All the gates control the flow of information. Informa-
tion is splitted into cells and hidden states. Cell state represents the long-term
memory. The hidden state represents short-term memory. Gates are used to
forget, save and update information throughout the period of learning.
Forget Gate:
$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$$
W_f and b_f are weights and bias of the forget state. f_t is the output of the
forget state. If it's close to 0 it's forgotten if close to 1 it's remembered. h_{t-1} is
previous short-term memory value. x_t is current input value(record)
Input Gate:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh W_C * [h_{t-1}, x_t] + b_C$$

i_t determines whether to remember or not the value of the new input. C_t is
value for the new input.
Cell state update:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

C_t is new long-term memory f_t is forget gate, C_{t-1} is previous long-term memory
and i_t, C_t are values from input gate.
Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t)$$

h_t is output of an entire LSTM unit and is called new short-term memory.
Whereas o_t decides whether to keep that new memory or discard of it. The
hidden layer output is the value predicted by our system.

3 Conditions to be met before deployment of the system

3.1 Learning conditions

3.2 Duration of learning

4 Issues with the system

4.1 Lack of data

4.2 New item introduction

4.3 Other problems

5 Enhancement of the system

5.1 Augmentation of one system with another

5.2 Systems competent of complementing stock recommendation system

References

[1] Tiago Duque. Building a stemmer. `https://medium.com/analytics-vidhya/building-a-stemmer-492e9a128e84`, February 2022. Last accessed November 1 2024.

[2] FeedSpot. Top 20 stock market news websites in 2024. `https://news.feedspot.com/stock_market_news_websites/`, October 2024. last accessed November 1 2024.

[3] DAVID R. HARPER. Forces that move stock prices. `https://www.investopedia.com/articles/basics/04/100804.asp`, October 2024. Last accessed October 31 2024.

[4] Eda Kavlakoglu IBM, Contributors:Jacob Murel Ph.D. What are stemming and lemmatization? `https://www.ibm.com/topics/stemming-lemmatization`, December 2023. Last accessed November 1 2024.

[5] Daniel Jurafsky. Speech and language processing. `https://www.dcs.bbk.ac.uk/~Dell/teaching/nlp/book/slp/slp_ch01.pdf`, 2000.

[6] Arkadiusz Krysik. Movie recommendation system: How it works and how to introduce it in your business. `https://stratoflow.com/movie-recommendation-system/`, June 2024. Last accessed October 31 2024.

[7] Arkadiusz Krysik. Music recommendation system: How do streaming platforms leverage ai in 2024? `https://stratoflow.com/music-recommendation-system-guide/`, August 2024. Last accessed October 31.

[8] Lillian Lee. Foundations of statistical natural language processing. *Comput Linguist*, 26(2):277–279, 2000.

[9] D Lokesh, V Vani, N Karthik, et al. Stock recommendation system for better investment plan. In *2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*, pages 1–6. IEEE, 2024.

[10] Microsoft. Recommended files in office. `https://support.microsoft.com/en-us/office/recommended-files-in-office-232c8322-1060-4453-bd69-a4770efe731e#:~:text=Microsoft%20Office%20displays%20a%20list,by%20people%20you%20interact%20with`. Last accessed October 31 2024.

[11] Moneka. Book recommendation system. `https://medium.com/@mmoneka11/book-recommendation-system-75963c7d4144`, August 2019. Last accessed October 31 2024.

[12] Stephen Oladele. Top 12 dimensionality reduction techniques for machine learning. `https://encord.com/blog/dimentionality-reduction-techniques-machine-learning/`, March 2024. Last accessed November 1 2024.

[13] Fayyaz Z.; Ebrahimian M.; Nawara D.; Ibrahim A.; Kashef R. Recommendation systems: Algorithms, challenges, metrics, and business opportunities. appl. sci. 2020, 10, 7748. `https://doi.org/10.3390/app10217748`, November 2020. Last accessed 24 Octopber 2024.

[14] Milos Simic Safa Bouguezzi. How does the amazon recommendation system work? `https://www.baeldung.com/cs/amazon-recommendation-system#:~:text=Collaborative%20filtering%20techniques%20are%20the,Their%20browsing%20behavior%20and%20ratings`, March 2024. Last accessed October 31 2024.

[15] Anand Subramaniam. Meeting customer expectations with ai-powered recommendation systems in banking: Use cases & insights. `https://kanini.com/blog/banking-recommendation-system-use-cases/`, May 2024. Last accessed 24 Octopber 2024.

[16] Trapica. Top 10 ecommerce recommendation systems. `https://medium.com/trapica/top-10-ecommerce-recommendation-systems-ce0a7a2bf4d1`, February 2020. Last accessed October 31 2024.

[17] Ning Wang, Hui Zhao, Xue Zhu, and Nan Li. The review of recommendation system. In *Geo-informatics in Sustainable Ecosystem and Society: 6th International Conference, GSES 2018, Handan, China, September 25–26, 2018, Revised Selected Papers 6*, pages 332–342. Springer, 2019.