Long-Tail Data-Driven Recommendations - Innovative Solutions for Financial Recommender Systems

Mariia V. Sigova   Igor K. Klioutchnikov   Oleg I. Klioutchnikov

Sobchak International Banking Institute

Saint-Petersburg, Russia;

2021

Long-tailed financial models are being introduced to overcome the problems with the realism of classical financial models, expand the supply of financial services, and better address the needs of netizens.

Classical financial time series models are based on homoscedasticity, variance uniformity, and normality. However, they cannot explain asymmetries, tail queries, and their volatility.

Non-standard variances in the financial sector began to be widely analyzed after Robert Engle's proposal to use Autoregressive Conditional Heteroscedasticity (ARCH) - 1980-90.

However, this method is poorly applicable for analyzing the niche demand of network users and preparing recommendations when using generative models.

Long-tailed financial models overcome the disadvantages of ARCH.

The Problem We Solved

The current situation

Explain

Solution

Discussion

Switching to using tails:
- to analyze consumer вemand in the financial sector;
- offering non-standard financial services.

Information processed by Generative Adversarial Networks (GANs)

Tails help improve the reliability of recommendations and prepare recommendations for non-standard users

Prospects for tails for the development of a new generation of GANs.

1. Individualization of financial solutions and expanding supply and demand
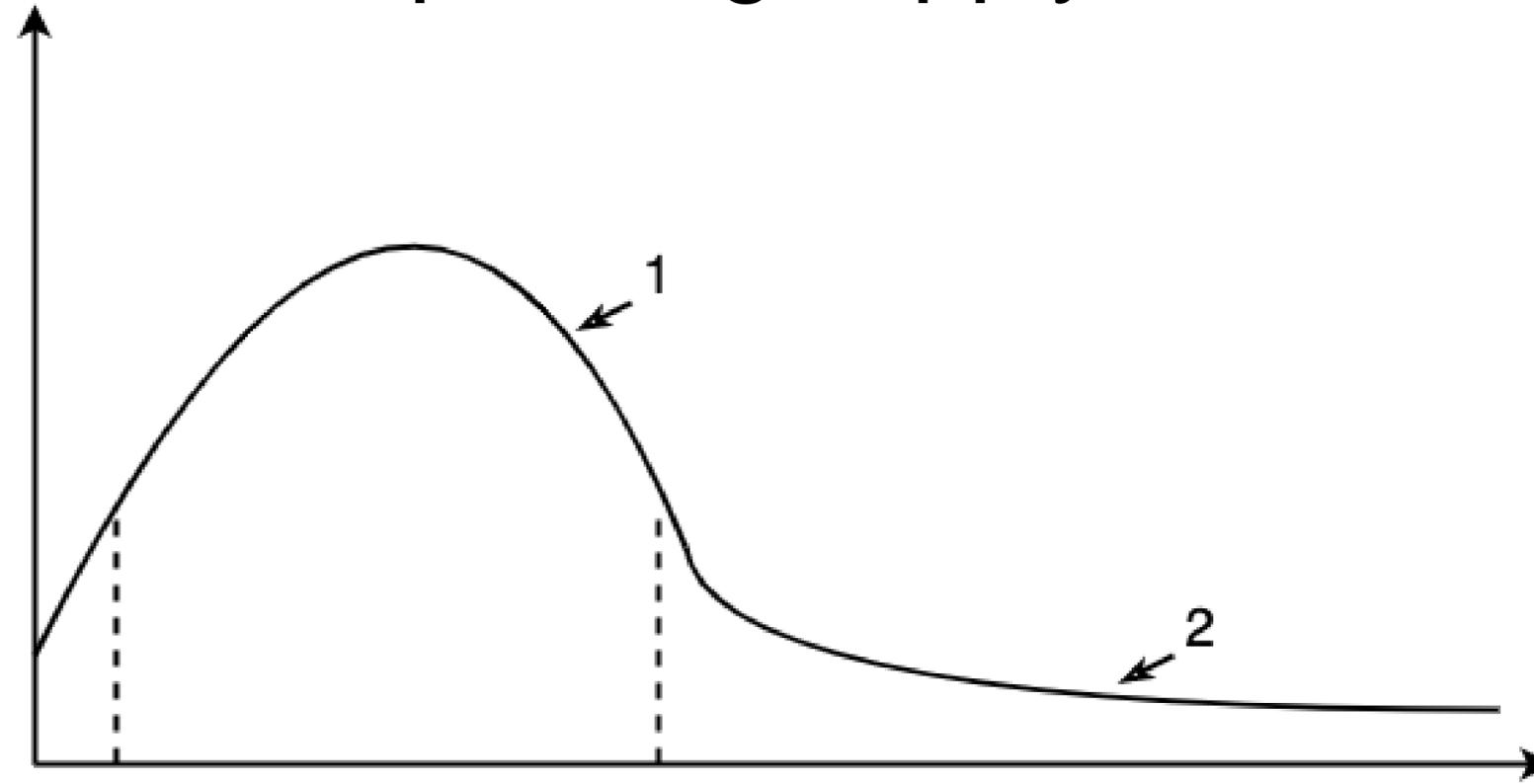
The head and long-tail graph is a complete model of the demand for financial services. The general concept assumes that the patterns of development of efficient markets (Fama) - (1) and behavioral finance (Kahneman & Tversky) - (2) are taken into account.

Network users are likely to move from standard financial services to personalized and niche services.

Long-tails in the distribution of demand for financial services often have some behavioral roots (emotional buying and spending, optimism or pessimism of netizens about their financial prospects) and therefore are studied in behavioral finance.

Benoit Mandelbrot - "the father of long tails"

Chris Anderson popularized the term and gave it new life (business strategy)

- Statisticians began to study long-tailed frequency distributions after World War II.

- In the 1950s-60s, this term began to be used in the financial and insurance business, and at the turn of the 20th and 21st centuries in stock trading.

- At the very end of the last decade, the concept of "long tails" began to be implemented in financial recommendation systems.

- In financial time series, not only in the head but also in the tails, observations run counter to simple random walk models and are more consistent with the phenomenon of clustering volatility (a wavy return to certain niche interests).

Long-tail distributions in finance

Two options for its application in generative system

- In the financial business, the term "long tail" is applied to the distribution of rank-size or rank-frequency distribution (primarily the popularity of financial services. The term is used to describe the strategy for accounting for various non-standard and niche financial requests of network users (long tail), which distinguishes such requests from popular and standard ("head").

- In the article, the term "long-tailed distribution" has a narrow technical meaning. Two options for its application in generative systems are considered:

- for a "critical" assessment of recommendations based on head distributions (through discriminatory systems),

- to satisfy non-standard and niche financial "appetites" of network users, as well as in microfinance, crowdfunding, and viral marketing of financial services (primarily in insurance).

Long-tail distributions in finance

Long-tail retail business strategies

- For retail, Chris Anderson defined the term in 2004. With its help, the range of online trade was dramatically expanded and non-standard consumers were attracted.

- In the insurance industry, long-tail strategies have dramatically expanded retail insurance.

- In the management of investment portfolios, long-tail strategies made it possible to move towards accounting for non-standard risks.

- Long-tail strategies in retail finance are now spreading rapidly. Social networks are becoming the leading arena for their distribution.

- Hypothesis 1:

The accuracy of the (R) recommendations is enhanced by using long-tailed conventional distributions (LTD) data.

Hypothesis testing

Generative model (GM).

Two options:

- LTD → DM → GM → R;

LTD → GAN → R.

- Hypothesis 2:

- Social networks are connected to the financial intermediation system;

- Recommender systems are the core of the system of interaction between banks and social networks.

Three main forms of the hypothesis are put forward:

The main goal of the hypothesis is to introduce tail data into the process of preparing recommendations that were not previously included in the information circle of recommendation systems.

"Weak shape" - tails allow you to correct the recommendations sharpened on the head;

"Medium Shape" - tails provide additional representativeness to recommendations based on head analysis;

"Strong form" - the tails have increased representativeness and can be independently used as a basis for developing recommendations.

- By leveraging the special properties of long-tail data, the parameters of specific consumer demand are more fully defined. This approach can be viewed as a "paradoxical enrichment" since otherwise, the long-tail data could invalidate the normal distribution result.

- The data of network users are included in recommendation systems as follows (Fig. 2):

1) A basic set for decision making, which determines the perception of external signals and the reaction to them.

2) There are certain techniques that allow you to focus the attention of the netizen and redirect his interest. These methods are used by an operator - a recommendation system.

3) There are two options for accessing the operator: first - the user does not have direct access to the operator's work - P (D, F); second, the user can use the feedback method to influence him (in this case, the conditional distributions are blurred - P (D, | F).

- The second principle corresponds to the work of generative opposing systems.

The data of network users are included in recommendation systems as follows (Fig. 2):

1) A basic set for decision making, which determines the perception of external signals and the reaction to them.

2) There are certain techniques that allow you to focus the attention of the netizen and redirect his interest. These methods are used by an operator - a recommendation system.

3) There are two options for accessing the operator: first - the user does not have direct access to the operator's work - P (D, F); second, the user can use the feedback method to influence him (in this case, the conditional distributions are blurred - P (D | F).

The second principle corresponds to the work of generative opposing systems.

2. The traditional method of developing recommendations

The traditional decision-making process is developing recommendations, including selecting data from the user for whom the recommendation is intended, comparing with other users, preparing a solution ("primitive" model - GM), validating it (DM), and final validation with the development of the recommendation.
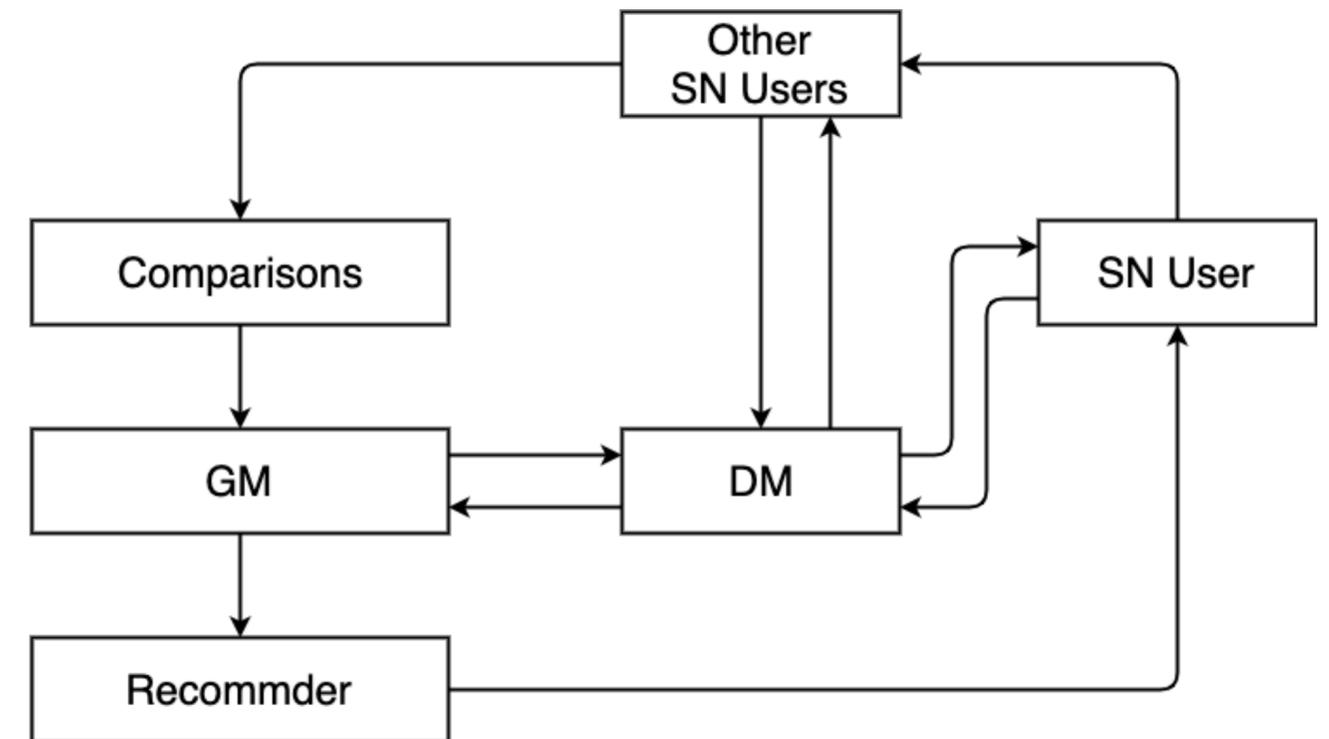
3. Using the behavior of other netizens as a basis for modeling personalized recommendations

Solutions are developed based on the system's perception of user search queries, identifying other web users with similar profiles, and comparing the queries of different web users. The idea is that, by comparing with other network users, this makes it possible to increase the accuracy of recommendations and more fully predict the behavior of the network user for whom a personal recommendation is made. This basic engine produces primitive predictions (recommendations), the accuracy of which is at the heart of the feedback learning engine.
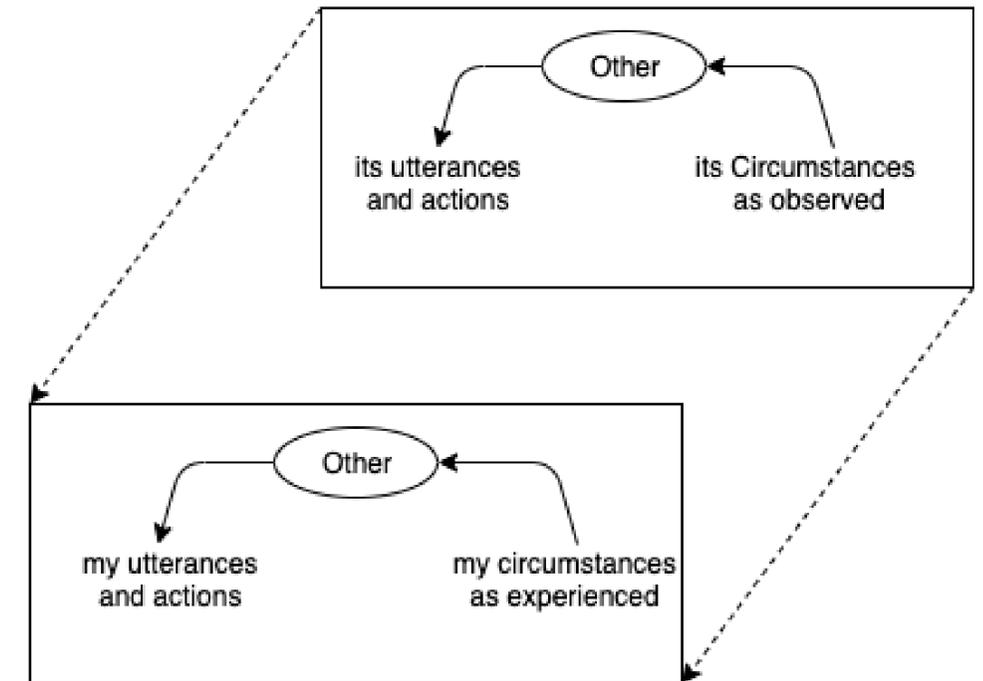
How are better recommendation reliability and better customer satisfaction achieved?

- The traditional approach takes into account the user's actions against the background of developing recommendations for other users (for example, according to the principle of a neighbor). This approach improves prediction accuracy using machine learning with feedback, during which long-tail data is connected to the validation of recommendations.

- Recursive processing of results improves the accuracy of recommendations. Data processing is carried out by a generative adversarial network (GAN), which consists of two adversarial networks - a generative filter (GM), which prepares a decision, and a discriminatory filter (DM), which "checks" and corrects the decision by connecting a new set of data obtained in tail sections (LT) using machine learning (ML) technologies.

4. Long-tail data is involved in adjusting the inferences of the generative model
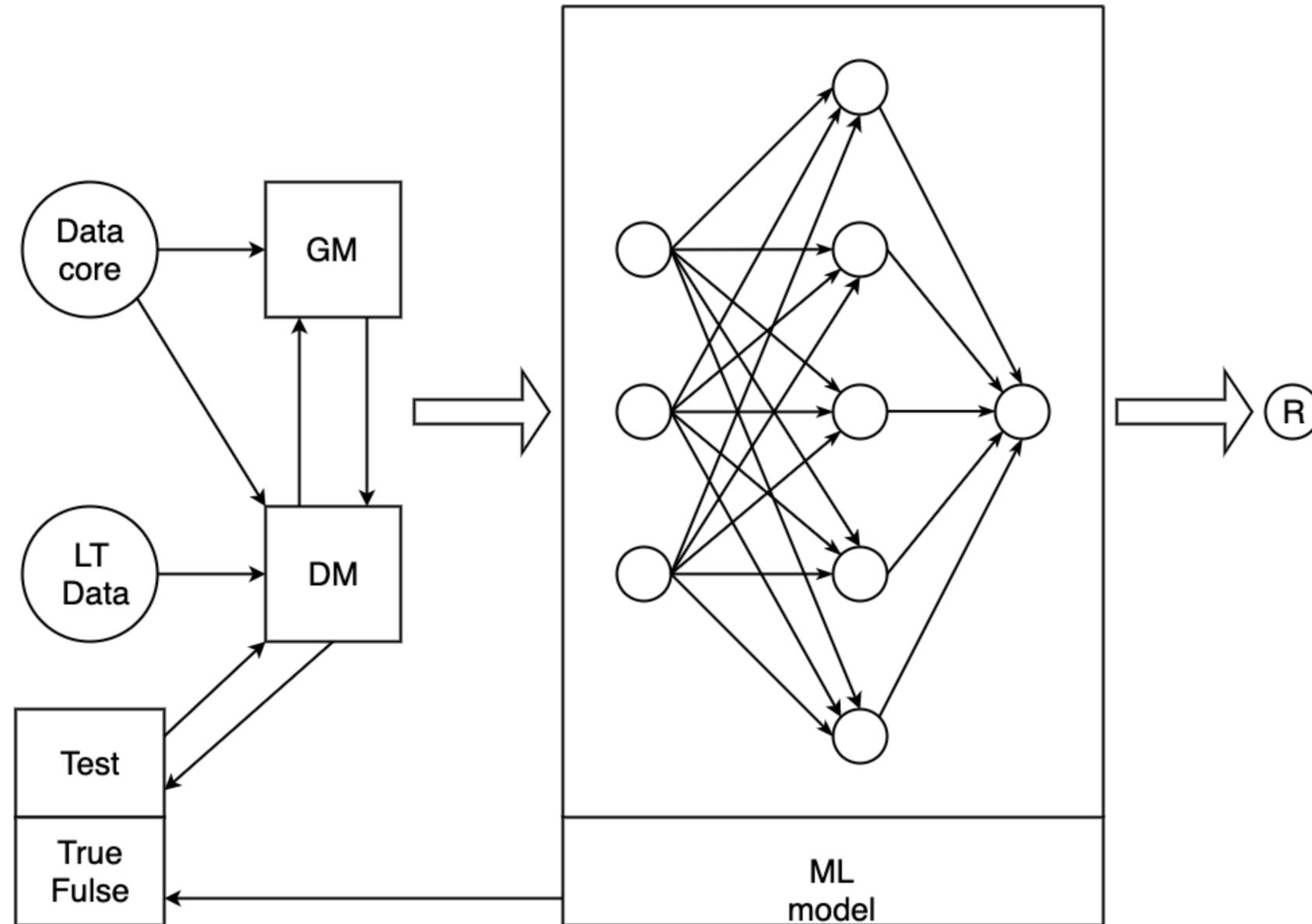
Limitations of using long-tail data in recommendations

and prospects

- long-tail data can have the opposite effect of reducing the accuracy of recommendations and showing negative economies of scale (bias towards the peripheral interests of netizens can skew the results);

- the cost and time to scale the data may not match the improvement in the accuracy of the recommendations;

- reducing the problem of long-tail data to a state of complete no data, as the data can be extremely garbled, completely unstructured, very blurry and with a lot of noise.

However, long-tail data in recommender systems could revolutionize more than just quantitative finance but also financial services, and transform work processes, business strategies, relationships, and society.

Discussion suggestions and possible follow-up research:

Long-tail data is an innovative solution for financial recommendation systems. It aims to expand and deepen the application of the recommendations in financial intermediation.

- The data of financial networks on social networks has a long memory.

Long memory can be found in various areas where long-tail data is a standard feature of commonly used stochastic models. A high level of probability of undulating recurrence of non-standard and niche financial requests of network users. By means of undulating clustering, the accountability of such queries is achieved.

- The move to widespread use of long-tail data is seen as part of the move to big data, modern business intelligence, and online financial intermediation. - The importance of working with data in strong noise interactions is growing. Automatic intelligent systems for analyzing text messages and browser searches of the user with filtering and highlighting peripheral signals in the general noise in order to identify and/or clarify key problems among users;

- Impact of edge query analytics on recommendation performance and the ratio of edge queries to main queries. How and why social media analytics create super-additive value through synergies based on the complementarity of data;

- The impact of tail-based advice on the behavioral intentions of netizens;

- A promising area of research is the definition of "attractor / s" in long tails.

Contacts

Autonomus Nonprofit Organization of Higher Education «International Banking Institute»

191023, Russian Federation, Saint Petersburg, 60 Nevsky ave

+7(812) 571-65-55

ibispb@ibispb.ru

Thank you for your attention