Non-IID Recommendation Theories and Systems

Longbing Cao
University of Technology Sydney, Australia
Philip S Yu
University of Illinois at Chicago

***Abstract*—A good recommender system must build an in-depth understanding of recommended users and items. This involves the non-IIDness, namely heterogeneity and couplings, of user preferences, behaviors and their couplings with item properties. The couplings consist of subjective and objective as well as explicit and implicit relationships both within and between users and items and between users and items. This article outlines the need and concepts of next-generation recommendation theories and techniques to address non-IIDness in recommendation.**

***Keywords*—recommender systems, behavior informatics, non-IIDness learning**

I. Introduction

Recommendation behaviors, services and systems are increasingly important for online business and economy, such as online shopping websites, online broadcasting systems, IPTV, and social media. So far, respective recommendation algorithms and systems have considered (1) the rating information, (2) user comments on items, (3) user friendliness, (4) group preferences, and (5) user cross-domain preference transfer. Some researchers also consider user behaviors [1], [5] of viewing or commenting on items. Most of existing work focuses on high-level aspects such as user, item, social relationship, and comment etc.

However, very few existing algorithms and systems incorporate (1) low-level properties of users and/or items, (2) heterogeneity between users and items, and (3) couplings [2] within and between users and items and between users and items. Often, existing theories treat users and their properties similar to items and item properties and essentially assume users and items are IID (independent and identically distributed). For instance, Matrix Factorization learns latent user and item matrix in the same way.

II. Non-IIDness in Recommendation

An effective recommender system for real-life recommendation services such as those for Google and Amazon needs to consider the heterogeneity and couplings, namely non-IIDness [3], embedded within and between recommended users and items.

A. Heterogeneity

Heterogeneity refers to different aspects, including

Different types of users and items, referring to various types, categories, domains, preferences, and behaviors of users and items, and the difference between user and item at the above aspects.

Different types of user properties and item properties, referring to the diversity of attribute types of users and items, and the difference between user properties and item properties. Here user properties refer to any aspects related to or describe a user, including user demographics, group, preference, comment, social relation, and behavior. Item properties refer to anything related to or describe an item, including item domain, category, and other basic attributes.

Different distributions of users and items, meaning users (and items) are not identically distributed, and the distributions between users and items are different.

B. Couplings

There are different types of intrinsic coupling relationships [2], including *User-user couplings*, referring to couplings within and between users, namely (1) intra-user property couplings between values of a user property, such as couplings between user preferences, groups, domains, behaviors, and between social relations; (2) inter-user property couplings between user properties, such as between user groups and their social relations, and between user preferences and their comments; and (3) user couplings between users and between user groups.

Item-item couplings, similar to user-user couplings, including (1) intra-item property couplings, (2) inter-item property couplings, and (3) item couplings between items, and between item categories.

User-item couplings, referring to the couplings within and between user-item connections, including (1) explicit user-item couplings, such as a user's rating on an item, and a user's comment on an item, and (2) implicit user-item couplings, such as the influence of or connections between a user's properties on his/her rated item properties. User-item couplings may take place on different levels, say within the rating table, between rating tables of various user groups, between a user property and an item property, or between user property matrix and item property matrix.

III. Theoretical Framework

To ensure accurate recommendations, comprehensive heterogeneity and couplings within and between user/item attributes, users and items, and between users and items need to be considered. Here a new theoretical framework is introduced, which draws a comprehensive picture of non-IID recommendation theories and systems. The objectives of this non-IID recommendation include: (1) Capturing both heterogeneity

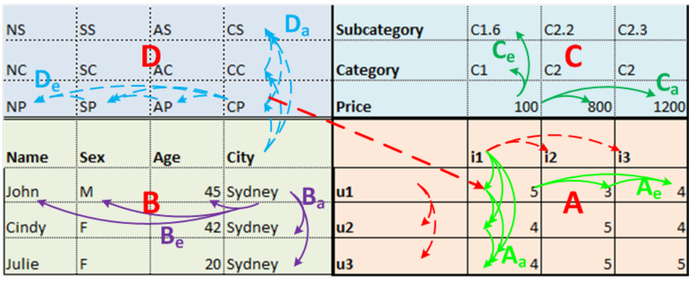

Fig. 1. Non-IIDness in recommendation.

and couplings, namely non-IIDness in recommendation, (2) Catering for both explicit non-IIDness such as user-user couplings and between ratings and implicit non-IIDness such as user-item couplings, and (3) Seizing both subjective non-IIDness such as in ratings and objective non-IIDness such as in user/item and between user and item properties.

Accordingly, as illustrated in Figure 1, a general framework for discussing and understanding non-IIDness in recommendation identifies multi-sources (spaces) of heterogeneity and couplings.

A. Subjective user-item non-IIDness

Subjective user-item non-IIDness RS^A in recorded in the user-item interaction space: Table A, capturing the explicit and subjective interactions and preferences of a user for an item, which is the area mainly explored by the current recommender system community. There are user-user rating and item-item rating non-IIDness in Table A. The rating non-IIDness $A(\cdot)$ thus indicate the relationship between the ratings given by a user to an item, which can be categorized as intra-rating non-IIDness between users $A_a(\cdot)$ and inter-rating non-IIDness between items $A_e(\cdot)$, and the aggregated rating non-IIDness $A(A_a(\cdot), A_e(\cdot))$.

By incorporating the non-IIDness between items and users in Table A, the overall explicit non-IID user-item similarity $\mathbb{RS}^A$ can be measured by:

$$\mathbb{RS}^A = A(A_a(\cdot), A_e(\cdot)) = A(A_{j_1,j_2}, A_{i_1,i_2}) \quad (1)$$

where A_{j_1,j_2} refers to rating non-IIDness across j_1 and j_2 items (namely intra-rating non-IIDness in Table A), and A_{i_1,i_2} refers to rating non-IIDness across i_1 and i_2 users (the inter-rating non-IIDness in Table A).

Alternatively, if the non-IIDness in Table A can be or is actually overlooked, then

$$\mathbb{RS}^A = A(A_{i,j}) \quad (2)$$

where $A_{i,j}$ represents the preference rating matrix, in which A_{i_1,j_1} is the rating of user u_{i_1} on item i_{j_1}, and $A(\cdot)$ represents the aggregation function.

B. User-user non-IIDness

User-user non-IIDness $\mathbb{RS}^B$ is in the user information space: Table B, capturing user demographic, group, domain and even behavioral and social relation information, etc. user properties. The user information Table B consists of intra-user non-IIDness $B_a(\cdot)$, inter-user non-IIDness $B_e(\cdot)$ and the aggregative non-IIDness $B(B_a(\cdot), B_e(\cdot))$. Hence, users are non-IID.

$$\mathbb{RS}^B = B(B_a(\cdot), B_e(\cdot)) \quad (3)$$

C. Item-item non-IIDness

Item-item non-IIDness $\mathbb{RS}^C$ is indicated in the item information space: Table C, storing item properties such as price, category and subcategory for each item, and the relationships and connections between items through item properties and item-item connections. The item information Table C consists of intra-item non-IIDness $C_a(\cdot)$, inter-item non-IIDness $C_e(\cdot)$, and the aggregative non-IIDness $C(C_a(\cdot), C_e(\cdot))$. Accordingly, items are non-IID.

$$\mathbb{RS}^C = C(C_a(\cdot), C_e(\cdot)) \quad (4)$$

D. Objective user-item non-IIDness

Objective user-item non-IIDness $\mathbb{RS}^D$ is in the user-item interaction space: Table D, representing the implicit and objective interactions between user properties and item properties. Any aggregation function for a specific cell such as $_{i_1 j_1}$ in Table D is the product of matrix C_a for specific item property q_{j_1} and matrix B_a for specific user property p_{i_1}. We call this a *coupled user-item cell*. The non-IIDness in each cell in Table D, namely $D_{i_1 j_1}$, may be a matrix to learn.

Implicit non-IIDness in a user-item coupled cell $D_{i_1 j_1}$ is $\mathbb{RS}^{D_{i_1 j_1}}$, which consists of two parts: the non-IIDness of a user i_1's specific property on all items with item property number j_1, namely $D_a(D_{i_1 j_{1*}})$ $(1 \leq j_{1*} \leq J)$; and the non-IIDness of an item j_1's specific property on all users with the user property number i_1, namely $D_e(D_{i_{1*} j_1})$ $(1 \leq i_{1*} \leq I)$. As a result, the implicit non-IIDness in a user-item coupled cell is $\mathbb{RS}^{D_{i_1 j_1}}$:

$$\mathbb{RS}^{D_{i_1 j_1}} = RS^{D_{i_1 j_1}}(D_a(D_{i_1 j_{1*}}), D_e(D_{i_{1*} j_1})) \quad (5)$$

The overall implicit user-item non-IIDness hidden in Table D has the similarity $\mathbb{RS}^D$, which is the aggregation of all user-item cell non-IIDness in Table D:

$$\mathbb{RS}^D = RS^D(\mathbb{RS}^{D_{i_1 j_1}}) \quad (6)$$

where $(i_1 \neq i_2) \vee (j_1 \neq j_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq J$. Alternatively, we may say the overall implicit user-item non-IIDness consists of two parts: a user i_1's property-based item non-IIDness $D_a(\cdot)$ on all item properties, and an item j_1's property-based user non-IIDness $D_e(\cdot)$ on all user properties, which are further coupled in terms of RS^D.

$$\mathbb{RS}^D = RS^D(D_a(\cdot), D_e(\cdot)) \quad (7)$$

where $D_a(\cdot)$ represents the non-IIDness between different item properties on a particular user property, and $D_e(\cdot)$ indicates the non-IIDness between user properties in terms of a specific item property.

E. Complete non-IIDness

The aggregated user-item non-IIDness is the combination through an aggregation function $RS^{A+D}(\cdot)$ of explicit user-item non-IIDness $\mathbb{RS}^A$ and implicit user-item non-IIDness $\mathbb{RS}^D$:

$$\mathbb{RS}^{A+D} = RS^{A+D}(\mathbb{RS}^A, \mathbb{RS}^D) = \sum_{i_1,i_2=1}^{I} \sum_{j_1,j_2=1}^{J} RS^{A+D}(A(A_{i_1j_1}, RS^D(RS^D_{i_1j_1}))) \odot (A_{i_1j_1}, RS^D_{i_1j_1}) \quad (8)$$

where $\sum_{j_1,j_2=1}^{J} RS^{A+D}(A(A_{i_1j_1}, RS^D(RS^D_{i_1j_1})))\odot$ means the subsequent non-IIDness of RS^{A+D} are $A_{i_1j_1}$ coupled with $A(A_{i_1j_1},\ RS^D(RS^D_{i_1j_1}))$, and $RS^D_{i_1j_1}$ coupled with $RS^D(RS^D_{i_1j_1})$, and so on, with non-determinism.

The complete non-IIDness in a recommender system is embodied through four sources: coupled users in Table B, coupled items in Table C, explicit user-item interactions in Table A, and implicit user-item interactions in Table D. Hence, the complete non-IIDness $\mathbb{RS}$ in a recommender system aggregates the actual preference $\mathbb{RS}^A$ of a user on an item in Table A, an accumulative indication $\mathbb{RS}^B$ of the influence between users determined by user properties and its impact on prediction, the relation between items $\mathbb{RS}^C$ shown by item properties, and the underlying hidden yet comprehensive interactions $\mathbb{RS}^D$ between the user properties and item attributes driving the rating.

$$\mathbb{RS} = RS(\mathbb{RS}^A, \mathbb{RS}^B, \mathbb{RS}^C, \mathbb{RS}^D) \quad (9)$$

Note that $\mathbb{RS}^D$ is not a simple matrix, because it carries much information from Tables B and C and their interactions on many different properties, at different layers and on various forms. It is in fact very complicated to obtain $\mathbb{RS}^D$, so future studies must incorporate the implicit interactions between properties in Tables B and C.

IV. Case Studies

In practice, as there are various non-IIDness as discussed above, the existing IID approaches such as collaborative filtering models do not work well. A new paradigm shifts the focus of recommendation design from understanding explicit user-item preference to considering both explicit and implicit user-item interactions across multiple sources, with different properties, layers and forms. This will trigger research into the next generation of recommendation algorithms and systems.

Coupled item recommendation A coupled item-based collaborative filtering algorithm is introduced in [2] by explicitly considering both intra- and inter-coupling between item attributes, and aggregating them in terms of the coupled behavior analysis [4].

Coupled matrix factorization The basic matrix factorization approach for recommendation builds on an assumption that users are IID. In [2], an MF algorithm incorporating couplings is discussed, which considers user-user and item-item couplings.

V. Conclusions

A comprehensive framework is proposed to review the non-IID complexities built in recommendation problems, which considers explicit and implicit, subjective and objective, local and global heterogeneity and couplings. Limited work has been done in this direction, which hopefully will inspire deep understanding and analysis of recommendation behaviors, problems and theories.

References

[1] Longbing Cao. In-depth Behavior Understanding and Use: the Behavior Informatics Approach, Information Science, 180(17); 3067-3085, 2010.

[2] Longbing Cao. Coupling Learning of Complex Interactions, Journal of Information Processing and Management, 51(2): 167-186 (2015).

[3] Longbing Cao. Non-IIDness Learning in Behavioral and Social Data. Comput. J. 57(9): 1358-1370, 2014.

[4] Longbing Cao, Yuming Ou, Philip S. Yu. Coupled Behavior Analysis with Applications. IEEE Trans. Knowl. Data Eng. 24(8): 1378-1392, 2012.

[5] Longbing Cao, Philip S. Yu. Behavior Computing - Modeling, Analysis, Mining and Decision. Springer, 2012.