September 2022

Using Cross-User Understanding to Develop Better User Embeddings

Nithya Kota

Maneesh Dewan

Ganesh Mallya

Jatin Chhugani

Using Cross-User Understanding to Develop Better User Embeddings

ABSTRACT

Digital payment apps, e.g., available via smartphones, can make multiple payment instruments available to a user; however, a user may not link all their cards to a payment app. Some payment apps include analytics features based on data from cards linked to the app. However, such data can give an incomplete financial picture of the user. This disclosure describes techniques to develop a more complete representation of a user that has provided only a partial view of their finances. Two embedding vector spaces are created, one trained over users with incomplete financial profiles and another trained over users with complete financial profiles. A map is created between the two vector spaces. The map is used to extend the representation of a user with a partial financial profile to that of a user with a complete financial profile.

KEYWORDS

- Payment app
- Financial profile
- User profile
- Financial information
- Spending pattern
- Credit card spend
- Cross-user embedding

BACKGROUND

Users exhibit preferences for certain payment instruments such as specific credit or debit cards under certain contexts. For example, the preferences can be driven by card benefits, e.g., certain cards that allocate more points for fuel spend, restaurant spends, airline spends, etc., user habits, or other criteria. Digital payment apps, e.g., available via smartphones, can make multiple payment instruments available to a user; however, a user may not link all their cards to a payment app. Some payment apps include analytics features, implemented with user permission. Such analytics are performed by the payment app based on data from cards linked to the app. However, such data can give an incomplete financial picture of the user. For example, if the linked card is primarily used for purchasing fuel while an unlinked card is used for restaurant payments, data analytics performed on the card linked to the payment app cannot provide the user with information about their dining spends.

DESCRIPTION

This disclosure describes techniques to develop a more complete representation, e.g., embedding, of a user that has provided a payment app only a partial view of their finances. The techniques are implemented with specific user permission. Users are provided with options to disable or limit such features

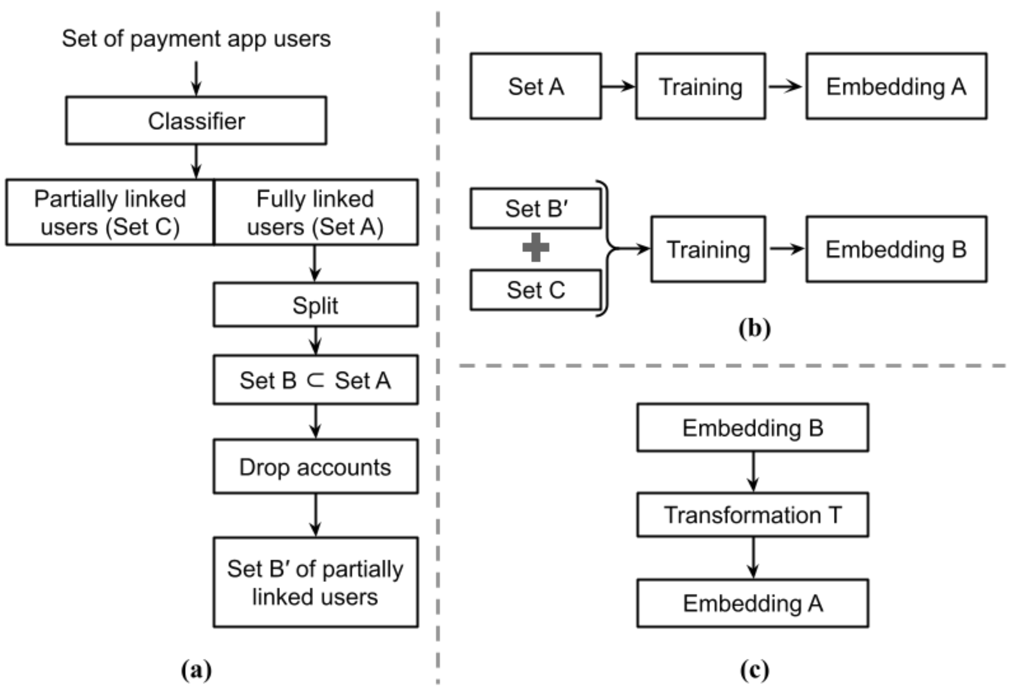

Fig. 1: Using cross-user understandings to develop a more complete user embedding: (a) Creating a set of partially linked users from a set of fully linked users; (b) Creating embeddings of fully linked and partially linked users; (c) Creating a transformation between embeddings of fully linked and partially linked users

Fig. 1 illustrates using cross-user understanding to develop a more complete user embedding. Data from users that have provided consent for use in cross-user understanding is used; data from other users is not included. Further, the data is processed to remove user-identifiable information.

In Fig. 1(a), a classifier divides the space of payment app users into users that have provided complete information (set A) and those who have provided partial information (set C). The classifier can be trained to separate users into the two subsets using user features such as spending, number of cards, demographics, etc. Subset B of the set A of fully linked users is formed by randomly selecting users from set A. A set B′ is formed from set B by randomly deleting some linked account data, e.g., transaction histories from one or more cards, for users in set B. Thus, set B′ comprises users with originally complete financial information but whose information is made incomplete by design.

In Fig. 1(b), a machine learning (ML) model trains over set A (fully linked users, which includes set B) to generate embeddings A. Example ML models that can be used to develop embeddings include frameworks such as auto-personalize, etc. Users in set A are represented by embeddings A. The machine learning model trains over the union of sets B′ and C which is a combined group of partially linked users, both original and artificially generated, to generate embeddings B. Partially linked users are represented by embeddings B.

In Fig. 1(c), a transformation T which can be a function or a model, is trained which maps the embedding B of a partially linked user (e.g., a user in set B′ or C) to the embedding A of a fully linked user. This can be done based on users in Set B, who have both embedding B and embedding A. When presented with the embedding of a user from set A, the transformation T acts as an identity map or nearly so.

Fig. 2: Transforming the embeddings of a partially linked user (a user in set C) to an embedding of a fully linked user

As illustrated in Fig. 2, the transformation T derived earlier is used to transform the embeddings of a partially linked user (a user in set C) to an embedding of a fully linked user. In this manner, a richer, more complete embedding can be obtained for users that have provided only partial financial information. The richer embedding (or representation) of the user can be used as an input feature in downstream applications such as offer recommendations, budgeting, subscription identification, category identification, etc.

The described techniques of improved user representation based on partial user information are applicable to products within financial services as well as other products that may utilize user information to make recommendations, e.g., search engines, recommendation engines, virtual assistants, automatic personalization modules, etc. The techniques enable obtaining a more complete representation or embedding of a user given an embedding obtained from a partial view of the user.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable the collection of user information (e.g., information about a user's credit or debit cards and other payment instruments, a user's payment apps, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level) so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques to develop a more complete representation of a user that has provided only a partial view of their finances. Two embedding vector spaces are created, one trained over users with incomplete financial profiles and another trained over users with complete financial profiles. A map is created between the two vector spaces. The map is used to extend the representation of a user with a partial financial profile to that of a user with a complete financial profile.