December 28, 2018

Attribute Prediction Based On Online Connections

Anonymous

Attribute Prediction Based On Online Connections

ABSTRACT

Techniques to predict attributes for a user or user group are described. User profiles and connections to other users are analyzed to determine user groups such as a household. Clusters are formed based on matching attributes of similar user groups. Available databases are queried using the user group identifiers to obtain a group-wide attribute value, e.g., household income, for a subset of user groups of a cluster. The obtained values are used to predict missing attribute values that were not found in the database, after adjusting for variations between user groups, such as a number of users in the group. User/group profiles are updated with the predicted attribute values. The predicted attribute values are utilized to customize content delivery to users.

KEYWORDS

user profile; incomplete profile; user group; social group; clustering; attribute prediction; attribute aggregation; content delivery; customization; social network; online connections

BACKGROUND

Online service providers such as video-hosting websites, online social networks, news portals, etc. deliver content to users. Some content providers customize content delivery based on available user information to provide a personalized user experience. Such customization can include selecting particular content for delivery or excluding content from delivery to particular users, organizing content in a particular order tailored for the user, etc. Different service providers provide different types of content such as audio/video, text articles, advertisements, etc.

Content customization is dependent on availability of user data. While some service providers maintain user profiles based on user data, such profiles are often incomplete. If the

available information about a user is incomplete, such providers may send users irrelevant content or exclude content that the users are interested in, providing a poor quality of experience.

DESCRIPTION

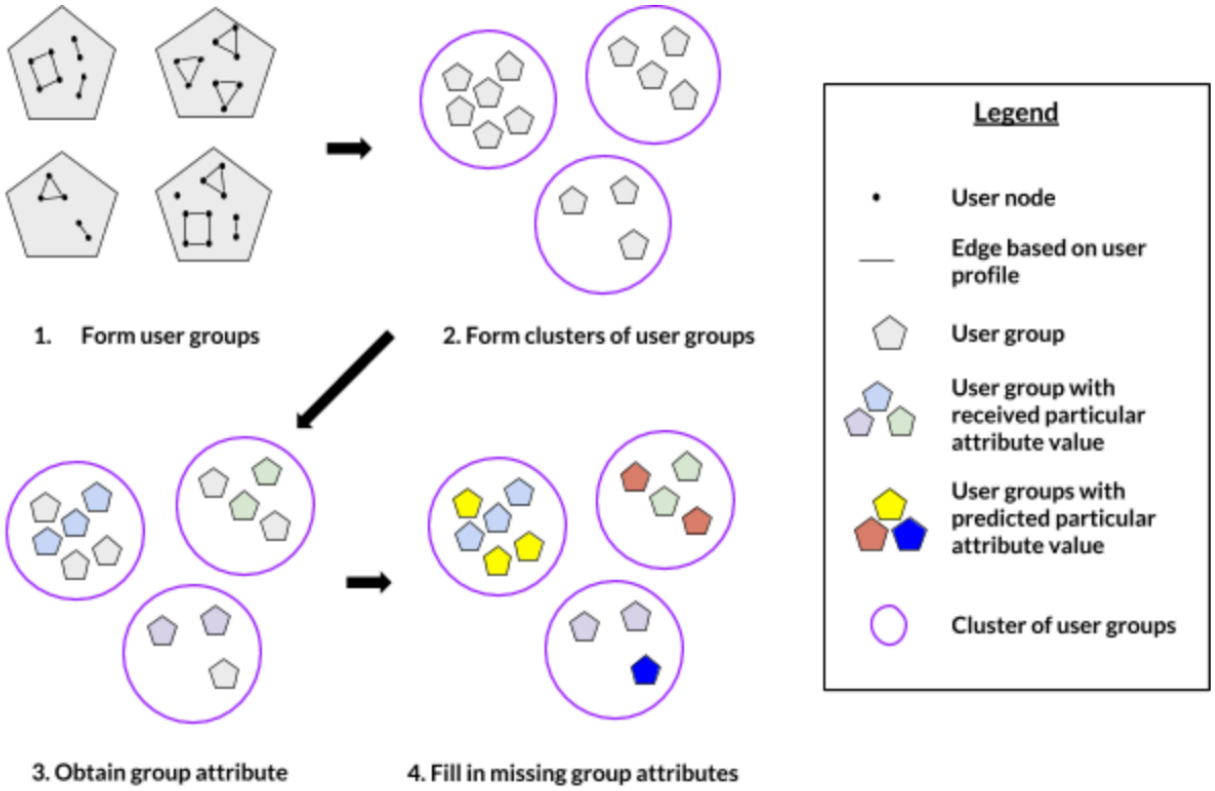

Fig. 1: Attribute prediction for user groups based on forming clusters

A process of attribute prediction for user groups based on forming clusters is illustrated in Fig. 1. Users of an online service are represented as nodes (dots) and connections between users are represented via edges (lines). Users are associated with user profiles that include attribute values, e.g., based on information provided by the user or determined by a server or system that maintains user data. For example, such information can include demographic information,

biographic information, user-provided data, prior user interaction with the server, etc. User profiles can be associated with individuals or organizations.

Nodes are aggregated into user groups (illustrated as pentagons) based on the types of edges between nodes, e.g., users that have similar types of connections with each other are selected into a group together. For example, the type of connection can be a relationship, e.g., spouse, family, friend, co-worker, etc. Groups can also be formed based on interaction between users and content provided by a server that maintains the user data, e.g., content items provided by the server and viewed by the user, user actions such as responding to content or sharing content with other users, etc. Group formation can also take into data indicative of a rate and/or recency of interaction between users, e.g., by assigning weights based on a rate/recency of interaction or relationship between users.

The user group is obtained by selecting nodes that are connected to another node of the user group by a particular type of connection and excluding other user nodes that are unconnected or have different types of connections with nodes in the group. For example, a user group can include users that are related such that the type of connection is “family” or users that live in the same household. Other types of user groups, such as college friends, club members, or an interest-based group etc. can also be formed based on respective types of connections.

In the example shown in Fig. 1, the user group excludes nodes with other types of connections to nodes in the user group. Fig. 1 illustrates 4 user groups. An individual user can be a member of a single user group (e.g., a group with the highest match with other users) or of multiple user groups. Matching can be based on any number of attributes, e.g., last name, last

name+address, etc. Attribute values are scores based on matching attributes and weights associated with respective attributes.

To obtain a user group, a first node is selected. A second node is added to the user group based on the second node having a particular type of connection to the first node. A third node is added to the user group if the third node is connected to the first node and/or the second node by the particular type of connection. In this repeating manner, nodes are added to the user group.

For each user group, a group-wide attribute value is determined based on individual attribute values for users within the user group. For example, an attribute value is obtained from a user profile of each node in the group. An aggregate value is determined based on the obtained values and is set as the group-wide attribute value. More than one group-wide attribute values can be determined and be stored in a group profile. The group profile can include aggregated values, e.g., an age attribute of the group may be a range (or average) based on age attributes of individual users in the group, a location attribute of a group may be based on locations of individual users, etc.

Clusters are formed by selecting user groups that have matching group-wide attribute values into a cluster. In Fig.1, each purple circle corresponds to a cluster. User groups are determined as belonging to the same cluster when the group-wide attribute values of the user groups are within a threshold of each other. For example, matching attributes of user groups are assigned weights and processed using a clustering technique such as k-means clustering, soft clustering, etc. The number of clusters and the number of social groups per cluster can be tuned by the clustering technique. Determination of matching attributes can be based on user

interaction of users within the user group with each other or with a content-provider system, such as a server. In some situations, a user group may be part of a single cluster.

For example, when user groups are households, clustering may be based on similar shopping patterns, e.g., a shopping spend per household (or per user per household) can be determined and used for clustering. Other group-specific demographic parameters, e.g., education level, etc. can be used in combination with the shopping spend. Weights are assigned to individual parameters when multiple parameters are used. A particular user group can be part of a single cluster, or multiple clusters based on respective weight. Fig. 1 shows 3 clusters.

A respective identifier (identifying attribute) is determined for each different user group in a cluster. The identifying attribute uniquely identifies each user group. For example, when addresses associated with a user profile in a user group match, the address is selected as an identifier of the user group. In another example, the identifier can be an attribute of a selected node that is in the user group. For example, when the user group is a household, the identifier can include last name, address, phone number, etc. In another example, the identifier can be a numeric code that identifies the user group.

The identifiers of different groups are sent to an external provider (separate from the online service that the users are associated with). Exchange of information with the external provider is performed with verification and uses encryption techniques. In response, value of a particular attribute for the user groups is received from the external provider. For example, the external provider can provide a data source, e.g., a household income database. The external provider obtains the particular attribute, e.g., household income, by querying the database using the received identifier. The external provider provides the particular attribute values for a subset

of the user groups within each cluster, illustrated as pentagons with pastel shades of blue, green and purple in Fig. 1.

After particular attribute values are received for a subset of the user groups, these values are utilized to predict the value for the particular attribute for other user groups in the cluster that are not in the subset. These are user groups with predicted attribute values, as illustrated in Fig. 1. One technique for prediction of the particular attribute value includes determining an aggregate value of the particular attribute for the cluster (e.g., a cluster-average) based on the received particular attribute values of the subset. The aggregate value is then assigned as the particular attribute value for a user group that is not part of the subset. The aggregate value is scaled prior to assignment, e.g., based on a ratio of cluster-wide average for particular attributes (e.g., an average 3 users per user group in the cluster) to the specific user group (e.g., that has 6 users).

Another technique for prediction of the particular attribute value includes obtaining a value of a different attribute for a user group that is in the cluster, but outside the subset (target group) for which the particular attribute value is to be predicted, and for a source user group that is in subset (source group). The predicted particular attribute value for the target user group is then determined by scaling the particular attribute value of the source group based on a ratio of the value of the different attribute for the source group and the target group. For example, household income or spending for a target user group may be scaled from the source group based on a ratio of the number of users in the household. Other parameters such as age, shopping spend, etc. can also be used to determine the ratio. The predicted value is stored as part of user group profile. User groups with predicted attribute values are illustrated as blue, yellow, and orange pentagons in Fig. 1.

Users are provided with content items, e.g., audio/video, sponsored content/ advertisements, web pages, games, content provided by other users such as photos, text posts, comments or reactions, etc. based on the particular attribute value of the user group to which the user belongs. As described above, the particular attribute value, which was previously unknown, is predicted using the foregoing method. Thus, customization of content delivery to users is feasible even with incomplete user data.

In some examples, the content items can be provided on a third-party website distinct from the system that predicts attribute values. For example, the third-party website can request content items such as advertisements. In these cases, the predicted attribute values can be utilized to select and provide advertisements or other content on the third-party website.

CONCLUSION

Techniques to predict attributes for a user or user group are described. User profiles and online connections to other users are analyzed to determine user groups such as a household. Clusters are formed based on matching attributes of similar user groups. Available databases are queried using the user group identifiers to obtain a group-wide attribute value, e.g., household income, for a subset of user groups of a cluster. The obtained values are used to predict missing attribute values that were not found in the database, after adjusting for variations between user groups, such as a number of users in the group. User/group profiles are updated with the predicted attribute values. The predicted attribute values are utilized to customize content delivery to users.