

# Technical Disclosure Commons

---

Defensive Publications Series

---

May 23, 2017

## EXPANDING USER LISTS USING LOCATION DATA

Jianjun Qiu

Follow this and additional works at: [http://www.tdcommons.org/dpubs\\_series](http://www.tdcommons.org/dpubs_series)

---

### Recommended Citation

Qiu, Jianjun, "EXPANDING USER LISTS USING LOCATION DATA", Technical Disclosure Commons, (May 23, 2017)  
[http://www.tdcommons.org/dpubs\\_series/534](http://www.tdcommons.org/dpubs_series/534)



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

## **EXPANDING USER LISTS USING LOCATION DATA**

### **BACKGROUND**

In online advertising, advertisers use remarketing lists to define a group of users to whom the advertiser would like to present ads. Users may be added to an advertiser's remarketing list for various reasons. In some instances, an initial remarketing list is populated at least in part by users that have recently interacted with a resource associated with a particular advertiser. For example, a computing system may automatically add a user to an advertiser's remarketing list if the user has recently visited a website associated with the advertiser, used an application developed or published by the advertiser, or completed a transaction (e.g., an online sale) with the advertiser. An online ad service may use the remarketing list to serve advertisements to users on the list, even if the users not presently interacting with a website, application, or other resource belonging to the advertiser. For example, a first user may visit the website for a first advertiser and, as a result, is added to a remarketing list for the first advertiser. For a period of time thereafter, when the first user accesses other websites or applications that display third-party advertisements, the ad service may select to present advertisements to the first user that are part of an ad campaign developed by the first advertiser.

### **SUMMARY**

This document describes computer-based techniques by which location data is analyzed to predict users that exhibit similar online behaviors, or that share other characteristics with, one or more initial users. In some implementations, a computing system can apply these techniques to expand remarketing lists for advertisers participating in an online advertising service.

## DESCRIPTION OF DRAWINGS

Figure 1 is a conceptual illustration of a computing system for using location data to identify users that likely share similar characteristics to users in an initial users list.

Figure 2 is a flowchart of an example process for generating a location profile for a user.

Figure 3 is a flowchart of an example process for generating a location-secondary characteristics model using location profiles and secondary characteristic data for a group of users.

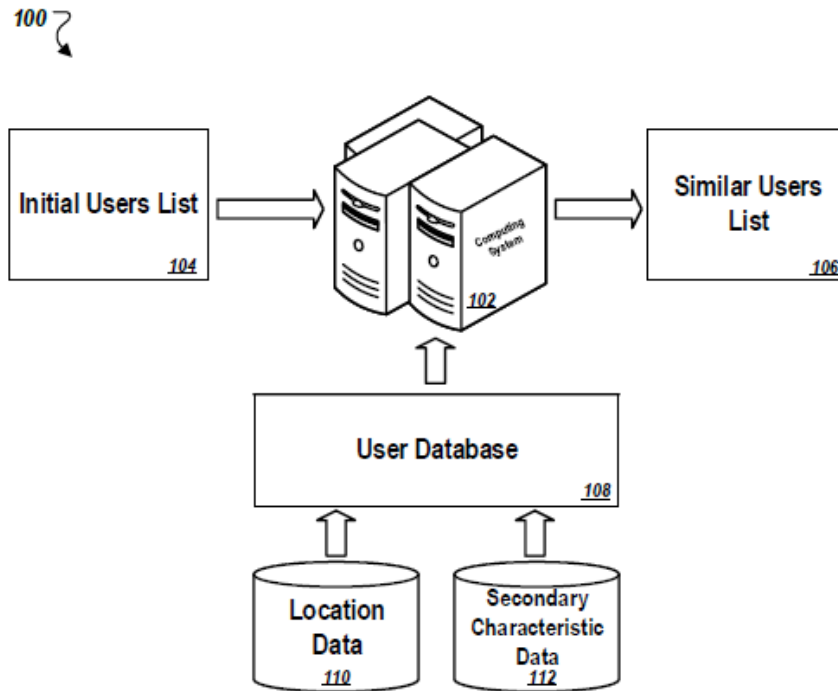
Figure 4 is a flowchart of an example process for identifying a target set of users that are similar to one or more users in an initial set of users.

Figure 5 is a flowchart of an example process for applying information about secondary user characteristics to weight or select particular location features when identifying a target set of users that are similar to one or more users in an initial set of users.

## DETAILED DESCRIPTION

This document describes systems, methods, devices, and other techniques for analyzing user location data and other information about users to determine a set of users that share similar interests to an initial set of users. In some implementations, the techniques may be applied to expand a remarketing list.

Figure 1 depicts an environment 100 of an example computing system 102 for using location data 110 to identify users that likely share similar characteristics to users in an initial users list 104.

**FIG. 1**

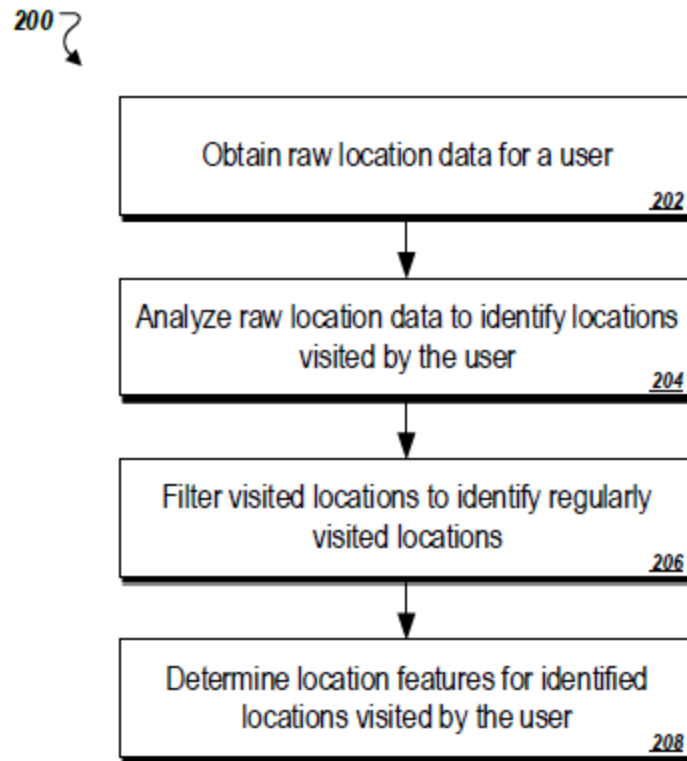
The computing system 102 may include one or more computers in one or more locations. The system 102 processes an initial users list 104 to generate a similar users list 106. The similar users list 106 identifies users that are predicted to be similar to all or some of the users identified by the initial users list 104. In some implementations, the computing system 102 is part of an online advertising service that selects and serves ads for presentation in electronic resources (e.g., websites, applications, or rich-media documents) published by a range of third-party publishers. Further, the ads may be served on behalf of various third-party advertisers that have partnered with the online advertising service to distribute ads to users according to parameters of advertiser-developed campaigns.

The similar users list 106 identifies users who are determined to be similar to users in the initial users list 104. In some implementations, the system 102 identifies users for inclusion in the similar users list 106 based on the users having location profiles that are sufficiently similar

to location profiles of users in the initial users list 104. A user's location profile generally indicates locations that the user has visited with regularity or sufficient frequency over a recent period of time. Further, in some implementations, the system 102 identifies users for inclusion in the similar users list 106 based on the users having one or more secondary user characteristics that are sufficiently similar to secondary user characteristics of users in the initial users list 104. The system 102 can sometimes infer secondary characteristics of users in the initial users list 104 and other users who are candidates for inclusion in the similar users list 106 based upon location profiles of the users and a model that correlates secondary user characteristics with information about locations frequently visited by users having known secondary characteristics.

The system 102 may access a user database 102 that stores location data 110 and secondary characteristic data 112 for one or more users. Secondary user characteristics generally include any user characteristics that relate directly or indirectly to a user's interest in particular topics or categories of content (e.g., advertisements). In some implementations, the secondary characteristic data 112 may represent characteristics of a given user's online activity such as the user's browsing history, search history, click history, or conversion history. For example, the secondary characteristic data 112 may include browsing history that indicates a particular user frequently visits websites related to sporting gear and luxury automobiles. The system 102 may then select advertisements to present to the user that relate to sporting gear, luxury automobiles, and associated topics.

Figure 2 shows a flowchart of an example process 200 for generating a location profile for a user.

**FIG. 2**

In general, a location profile is a model of a user's travel or location history over a period of time. The location profile identifies one or more locations that a user has visited during the period of time. As described further below, in some implementations the location profiles of users can be used to identify users who share the same or similar secondary characteristics, and thus, to identify users who may be interested in receiving similar topics or categories of advertisements. In some implementations, the process 200 is carried out by a computing system, e.g., system 102.

At stage 202, the system obtains raw location data for a user. The raw location data can include time-location pairs that indicate the current location of a device associated with a user (e.g., a smartphone, a tablet computing device, a wearable computing device, an onboard vehicle system, or other device that typically travels with a user) at different times. In some

implementations, the raw location data for a user is Global Positioning System (“GPS”) data reported by one or more of the user’s devices. In some implementations, the raw location data is derived from information about the locations of network access points that communicated with one or more of the user’s devices.

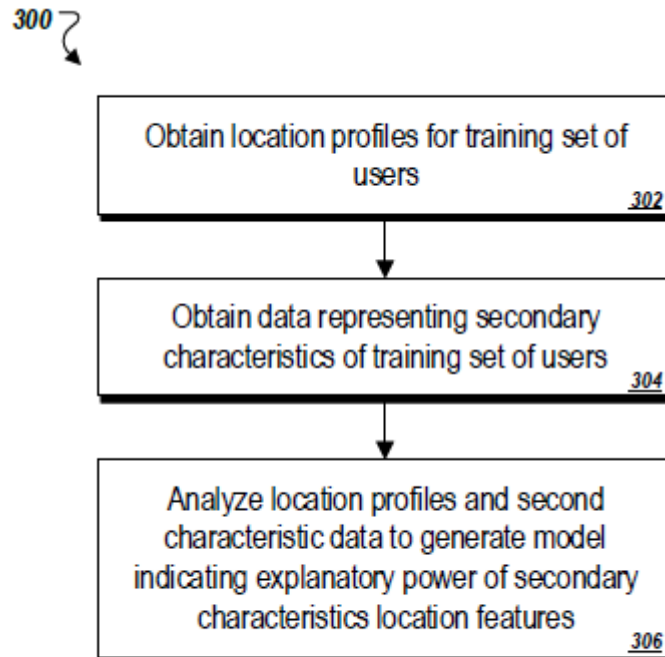
At stage 204, the system analyzes the raw location data to identify one or more locations visited by the user during the spanned period of time. In some examples, the identified locations visited by the user are physical landmarks or venues such as schools, restaurants, stores, work places, transportation hubs (e.g., subway or train stations), or the like. The system may identify visited locations by cross-referencing a database that indicates geographic coordinates of various landmarks and venues to determine which landmarks or venues are located at the geographic coordinates indicated by the raw location data. Optionally, at stage 206, the system filters the list of visited locations based on visit frequencies. For example, the system may discard information about locations that the user visited less than two times in a week-long period or in a month-long period. Different thresholds may be applied for different location categories. In some implementations, the location profile distinguishes between locations visited at different times. For example, a church visited twice daily during the weekdays is more likely indicative of a user dropping and picking up kids from a church school, whereas a church visited on Sunday mornings more likely indicates the user is attending weekend services.

At stage 208, the system determines location features for the locations visited by the user. The location features for a location can include characteristics such as a vertical classification or category for a landmark or venue at the location (e.g., park, restaurant, school, residence, transportation hub), a unique identifier of a location visited by the user, geographic coordinates of the visited location, or other characteristics of the location. In some implementations, the

location profile of a user may include aggregated information about user visits to multiple locations that are grouped by one or more location features. For example, the location profile for a first user may indicate that he or she visits department stores an average of more than three times per month, whereas the location profile for another user may indicate that he or she visits department stores an average of less than one time per month. The system can generate location profiles for a population of users, including users on an initial remarketing list (e.g., initial user list 104) and candidate users for inclusion on an expanded remarketing list (e.g., similar users list 106).

After generating users' location profiles, the system may process an initial remarketing list to generate an expanded remarketing list identifying users similar to all or some of the users on the initial remarketing list. In some implementations, the system determines a group of similar users to add to the initial remarketing list using a location-secondary characteristics model ("LSC model"). The LSC model can be a statistical model that indicates correlations between locations visited by users (or location features of locations visited by users) and secondary user characteristics. As an example, considering secondary user characteristics that represent aspects of a user's online activities, an LSC model may indicate the respective explanatory powers of one or more visited locations, or location features, to the tendency of a user to visit certain categories of websites, to click a link on certain categories of websites, to complete a transaction (e.g., purchase) on certain categories of websites, and/or to perform one or more other online activities. Figure 3 is a flowchart that depicts an example process 300 for generating an LSC model using location profiles and secondary characteristic data for a group of users.

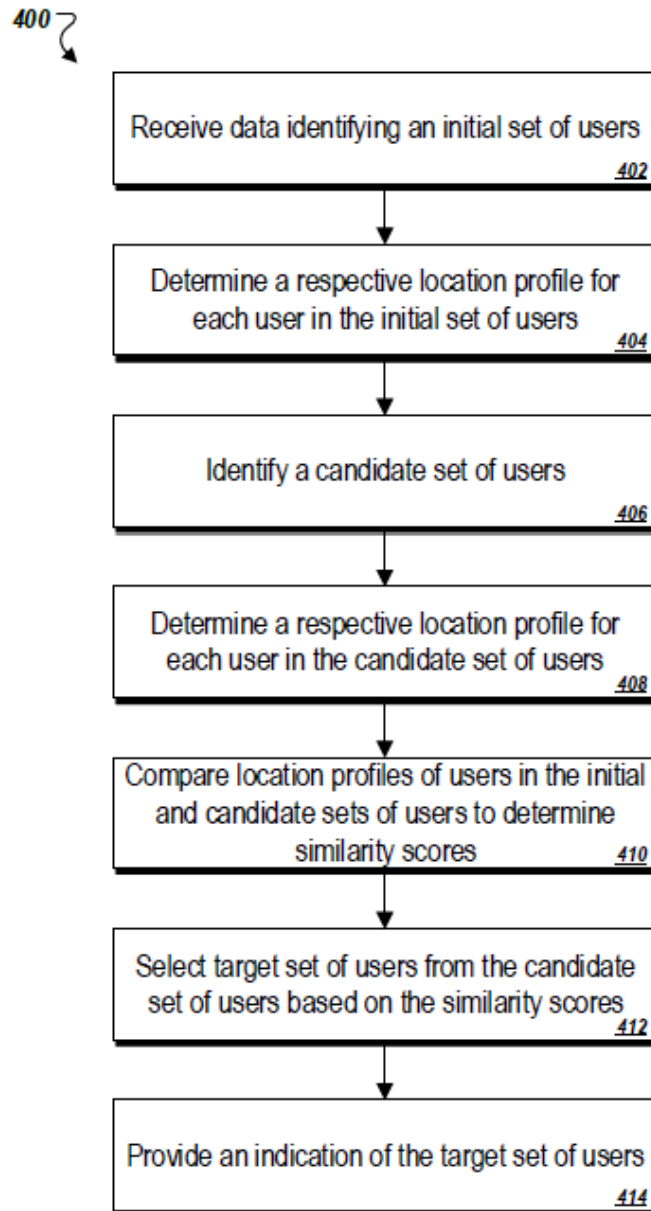


**FIG. 3**

The process 300 can be carried out by a computing system, e.g., computing system 102. At stage 302, the system obtains a respective location profile for each user in a group of users (referred to here as a training set of users). At stage 303, the system then obtains data representing secondary characteristics of each user in the training set of users. Then, at stage 306, the system performs statistical analysis on the location profiles and secondary characteristic data of the training set of users to generate the LSC model. In some implementations, the LSC model can indicate, for each of multiple different secondary user characteristics, respective explanatory powers for each of multiple different visited location features with respect to the particular secondary user characteristic. In some implementations, the LSC model can indicate, for each of multiple different visited location features, respective explanatory powers of the particular secondary user characteristic to each of multiple different secondary user characteristics. In some implementations, the LSC model may classify visited location features as either a predictor or a non-predictor for a given secondary user characteristic based on the

strength of the explanatory power for the visited location feature to the secondary user characteristic. If the explanatory power for a visited location feature to a secondary user characteristic meets a threshold value, the visited location feature may be classified as a predictor for the secondary user characteristic. If the explanatory power for a visited location feature to a secondary user characteristic does not meet the threshold value, then that feature may be classified as a non-predictor for the secondary user characteristic.

Figure 4 shows a flowchart of an example process 400 for identifying a target set of users that are similar to one or more users in an initial set of users. In some implementations, the process 400 may be carried out to identify similar users to add to an initial remarketing list. A computing system may be performed by a system of one or more computers in one or more locations, e.g., by computing system 102.

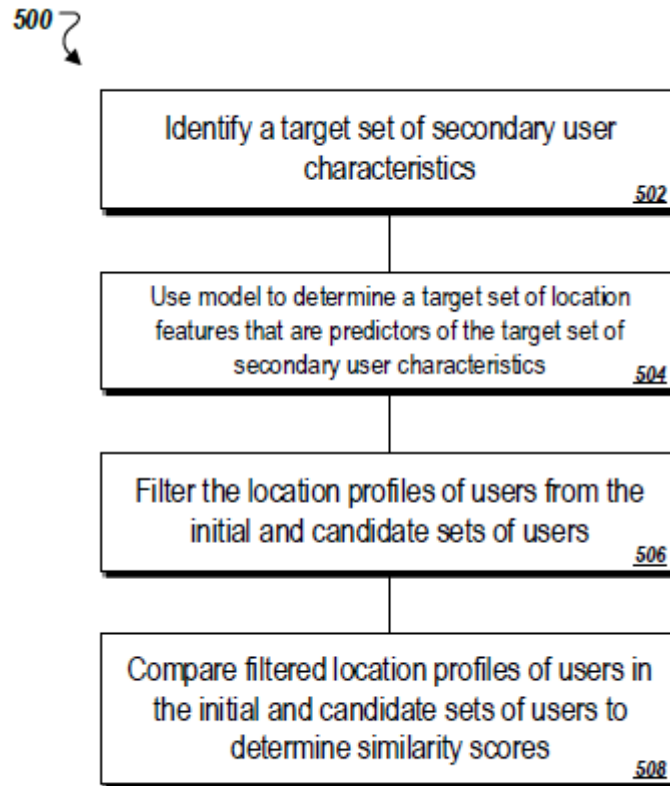
**FIG. 4**

At stage 402, the system receives data identifying an initial set of users. The initial set of users may, for example, be users on an advertiser's remarketing list. At stage 404, the system determines a respective location profile for each user in the initial set of users. In some implementations, the location profiles are generated according to the process 200 described with respect to Figure 2. At stage 406, the system identifies a candidate set of users. Location data

may be available for each user in the candidate set of users, although data representing secondary characteristics of the candidate set of users may or may not be available to the system. At stage 408, the system determines respective location profiles for users in the candidate set of users. As with the initial set of users, the location profiles can be generated for the candidate set of users according to the process 200 described with respect to Figure 2. The system 410 may then compare the location profiles of one or more users in the initial set of users with the location profiles of one or more users in the candidate set of users to determine similarity scores between each candidate user and one or more users in the initial set of users. By way of example, two users who both frequently stopped by pre-schools in the morning and mid-day during two or more days of the week may have similarity scores representing a high level of similarity of the users (e.g., since the travel patterns of the users indicate that the users likely share common interests related to care of pre-school aged children).

At stage 412, the system then selects one or more target set of users as a subset of the candidate set of users. The selected target users are deemed sufficiently similar to users in the initial set of users such that, for example, the target users can be added to an initial remarketing list. The similar location profiles indicate the target users are likely to be interested in third-party content (e.g., advertisements) similarly to users on the initial list of users. At stage 414, the system provides an indication of the target set of users, e.g., by presenting the target set of users to an advertiser or system administrator. In some implementations, the system may automatically add the target set of users to the initial users list, although a system administrator or a party that submitted the request to expand the initial users list (e.g., an advertiser) may be provided an option to confirm or reject the addition of automatically identified users to the initial users list.

In some implementations, the system can further apply information about secondary user characteristics to weight or select particular location features when comparing location profiles. For example, the location profiles of users in an initial remarketing list may indicate that these users regularly visit a business district on weekday mornings and also regularly visit grocery stores on the weekend. If the LSC model indicates that frequent visits to a business district is a strong predictor of users visiting banking websites, but frequent visits to grocery stores are not predictors of whether users will visit banking websites, then the system may select a target set of users that similarly visit business districts on weekday mornings regardless of whether these users also regularly visit grocery stores on the weekend. Thus, the system may compare the location profiles of users from the initial set of users to the location profiles of users from the candidate set of users primarily, or only, with respect to location features that have sufficiently strong explanatory power for one or more secondary user characteristics (e.g., the user's propensity to visit banking websites). An example process 500 for analyzing information about secondary user characteristics to weight or select particular location features is shown in Figure 5.

**FIG. 5**

At stage 502, the system identifies a target set of secondary user characteristics. The target set of secondary characteristics represents user behaviors or characteristics that are sought in the target set of users (e.g., users that may be added to a remarketing list). For instance, in the preceding example, the target secondary characteristic was users whose online activities included visiting banking or financial websites. In some implementations, the target set of secondary characteristics is derived from the initial users list by identifying one or more of the most prominent or distinctive secondary characteristics of all or some of the users on the initial users list. In some implementations, the target set of secondary characteristics is derived from information about an entity that is requesting to expand the initial users list. For example, the target secondary characteristics may include users that frequently interacted with websites or applications related to a same or similar topic of a website or application of an advertiser that is

seeking to expand a remarketing list. In some implementations, the target set of secondary characteristics may be manually selected by the advertiser or a system administrator

At stage 504, the system evaluates an LSC model to determine one or more location features that are predictors of characteristics in the target set of secondary characteristics. For example, the LSC model may indicate that regular visits to a fitness center is a predictor (e.g., has significant explanatory power) for users that visit health and wellness websites. At stage 506, the system may filter the location profiles of users from the initial and candidate sets of users based on the target set of location features, and at stage 508 the system compares the filtered location profiles of users in the initial and candidate sets of users to determine a measure of similarity between the users. As a result of the filtering, the comparisons of location profiles may be performed with respect to the target set of location features. For example, to identify additional users that are likely to visit health and wellness websites, the system may select users from the candidate set of users whose location profiles indicate that they regularly visit fitness centers. In this way, users' location data and travel patterns may be used as a proxy for online activity data or other secondary user characteristics, which may not be accessible to the system.

For situations in which the systems and techniques discussed herein collect personal information about users, or may make use of personal information, the users may be provided with an opportunity to control whether programs or features may collect personal information (e.g., information about a user's location or online activities), or to control whether or how to receive content from a content server or other data processing system that may be more relevant to the user. In addition, certain data may be anonymized in one or more ways before it is stored or used, so that personally identifiable information is removed when generating parameters. For example, a user's identity may be anonymized 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, postal code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over how information is collected about him or her and used by the content server or other computing systems.



## ABSTRACT

This document describes computer-implemented systems, methods, devices, and techniques for identifying users whose location histories or travel patterns indicate shared interests or characteristics of the users. In some implementations, the techniques are applied to expand a remarketing list by adding users to the remarketing list who have similar location histories or travel patterns to one or more users that were initially identified on the remarketing list. The similar location histories or travel patterns can be identified with respect to particular location features that are determined to have sufficient explanatory power in predicting specified interests or other characteristics of a user.