September 24, 2018

# INTELLIGENT DATA-DRIVEN USER POPULATION SUBSET SELECTION FOR PROGRESSIVE FEATURE ROLLOUTS

Michael Procopio



# INTELLIGENT DATA-DRIVEN USER POPULATION SUBSET SELECTION FOR PROGRESSIVE FEATURE ROLLOUTS

When a new feature is launched in a software product, a feature rollout service decides how to expose or 'roll out' the new feature to users. Typically, the feature rollout service adopts multiple phases to gradually introduce the new feature to more and more users with an improved version based on user feedback from each phase. In each phase, the feature rollout service collects information (e.g., user feedback or bugs) from a selected subset of users and updates the new feature on the basis of user feedback or bugs. In an initial phase, the new feature is provided to a subset of users that are randomly selected. For example, a user interface component (e.g., a button) can be provided to a graphical user interface (GUI) of the software for a selected user to access the new feature, but not force the user to interact with the new feature. In another example, a user interface component (e.g., a pop-up window) may be provided to advertise the new feature and receive consent from the user to access the new feature. As the selected users use the new feature, any feedback or data derived from the usage are collected for improvement to the new feature. Such information is used to prevent potential data loss and/or data corruption due to the new feature, and/or malfunction of the new feature.

Conventionally, after the initial phase, the improved new feature becomes available to a greater number of users that are also randomly selected. Feedback is collected again and the latest version of the new feature undergoes another phase of the rollout. In a later phase of the rollout, the latest version of the new feature can be automatically provided to the randomly selected users. Moreover, the randomly selected users are informed of an addition of the new feature. In some cases, a user can choose to remove the new feature. In other cases, a user may not be given such a choice. Accordingly, quality of the new feature would gradually improve

with additional and more feedback from the greater number of randomly selected users. The feature rollout service can finally make available the new feature that has been improved in multiple phases, to the general population of users.

Given the nature of the feature rollout service, feedback collected during early phases of a rollout may be critical to an improvement of a new feature, as well as understanding potential users of the new feature. Depending on quality of feedback, critical bugs can be detected early on to prevent any negative impact to the general population of users. Therefore, randomly selecting a subset of users to test a new feature may jeopardize success of the new feature in the early phases. Some users may not be familiar with a new feature, or accustomed to any features similar to the new feature. Therefore, they may have negative experiences with the new feature and accordingly, may not try to use the new feature at all in the future. This means that some users would not provide enough feedback to make improvements on the new feature.

Accordingly, a mechanism is proposed for predicting a subset of users for a feature rollout in a cloud storage system such that feedback that is useful for improvement of a new feature may be collected. A cloud storage system used herein refers to a system including a cloud-based environment (including a server and a data store) connected to a user device via a network. The server may host a cloud-based content management platform. The cloud-based content management platform may enable a user to access and collaboratively edit documents via a GUI. The cloud-based content management platform can additionally provide one or more applications such as an online calendar, an email, and a messenger application. Thus, the users can schedule meetings for a project and communicate with each other over email messages or instant messages to facilitate online collaboration. The GUI of the cloud-based content management platform may be provided by a web browser or a mobile or desktop application. A

new feature may be added to the cloud-based content management platform so that an addition of the new feature may be visible on the GUI.

The data store may include a cloud storage that stores documents associated with each user account and metadata associated with the documents. Such metadata may include an access control list of users (e.g., who the creator or owner is, who can view and edit, who the document has been shared with), a key text, and events associated with the document. A key text used herein refers to a text that represents a topic of content of the document. An event used herein refers to a user activity incurred to access a document (e.g., an open event, view event, edit event, or a comment event). For each event, the data store may store metadata associated with the event, such as a type of the event, a creator of the event, content, and a timestamp. Alternatively, the method may be performed in any system outside the cloud storage system.

In an initial phase of the rollout, the mechanism identifies users who are satisfied with using the new feature (e.g., who frequently use the new feature and/or provide positive feedback) and determines common characteristics amongst the users. The mechanism then uses the common characteristics to predict a set of users who are likely to be satisfied with an improved version of the new feature for the next phase of the rollout.

Figure 1 illustrates a flow diagram of a method for intelligently predicting a subset of users for a feature rollout in a cloud storage system. First, at block 100, the feature rollout service may identify a first set of users having user accounts associated with the cloud storage system on a random basis. The feature rollout service may identify a predetermined number of users (e.g., 5%) as a first set of users on a random basis.

Subsequently, at block 110, the feature rollout service may provide access to a new feature of the cloud storage system to user accounts associated with the first set of users.

Specifically, when the user logs into the cloud-based content management platform, the feature rollout service may provide a GUI representing a home screen of the cloud-based content management platform. On the GUI, the feature rollout service may provide an option to allow access to the new feature to the user accounts associated with the first set of users. Once the feature rollout service receives consent from the user for an addition of the new feature, the feature rollout service may allow access to the new feature to a corresponding user account. Consequently, the new feature may be visible and available for use on the GUI and highlighted as a new feature. Alternatively, the feature rollout service may automatically provide the new feature to the user accounts associated with the first set of users without an explicit action by the first set of users and highlight or advertise the new feature on the GUI.

Next, at block 120, the feature rollout service may determine, for each user in the first set of users, user satisfaction with the new feature. The feature rollout service may determine the user satisfaction implicitly and/or explicitly. In an implicit approach, the feature rollout service may measure the user satisfaction based on user interactions with the new feature. For example, the feature rollout service may determine a frequency and type of the user interactions (i.e., how substantively the user interacts with the new feature), whether the new feature has been requested to be turned off (i.e., indication of no satisfaction), and any bugs or malfunctions detected during the user interactions (i.e., negative experience with the new feature). In an explicit approach, the feature rollout service may measure the user satisfaction by surveying users. For example, the feature rollout service may collect any feedback submitted in association with the user interactions. In particular, the feature rollout service may collect comments or ratings by providing survey questions or polls having comment boxes or scales for multiple criteria to evaluate the new feature such as usability and functionality.

Subsequently, at block 130, the feature rollout service may train a machine learning model using common characteristics of users in the first set of users who are satisfied with the new feature. Examples of the common characteristics are described below with respect to the heuristics approach. The machine learning model may correspond to a model artifact that is created by the feature rollout service using training data that includes training inputs and corresponding target outputs (i.e., correct answers for respective training inputs). The feature rollout service may generate training data that includes one or more training inputs and one or more target outputs. The training input may include a list of users in the first set. The target output may include an indication of satisfaction (or no satisfaction) of each user in the first set.

The feature rollout service may find patterns (i.e., common characteristics of users in the first set of users who are satisfied with the new feature) in the training data that map the training input to the target output (i.e., an indication of satisfaction) and provide the machine learning model that captures these patterns. The machine learning model may be composed of, e.g., a single level of linear or non-linear operations (e.g., a support vector machine [SVM] or may be a deep network, i.e., a machine learning model that is composed of multiple levels of non-linear operations. An example of a deep network is a neural network with one or more hidden layers, and such machine learning model may be trained by, for example, adjusting weights of a neural network in accordance with a backpropagation learning algorithm or the like.

Once the feature rollout service determines that the machine learning model is ready for use, at block 140, the feature rollout service may predict a second set of users who are likely to be satisfied with an improved version of the new feature by using the trained machine learning model. That is, the feature rollout service may provide a list of all users (excluding the users in the first set) as an input data to the trained machine learning model. Then, the feature rollout

service may obtain output from the trained machine learning model that indicates which user(s) is likely to be satisfied with an improved version of the new feature.

Alternatively, the feature rollout service may adopt a heuristics approach to predict the second set of users who are likely to be satisfied with the improved version of the new feature. For example, the feature rollout service may first determine model users from the first set of users. A model user may be a user who has provided a positive feedback on the new feature. The feature rollout service may analyze the feedback collected in association with the user interactions to identify a model user who has provided more than a predetermined threshold percentage of positive feedback on each feedback criterion such as usability and functionality. Alternatively, or additionally, a model user may be a user who has frequently used the new feature during a predetermined time period. The feature rollout service may select a user who has used the new feature, for example, every day or above any number of times during the last month as a model user.

Once the feature rollout service identifies the model users from the first set of users, the feature rollout service may determine common characteristics of the model users. With respect to the common characteristics, the feature rollout service may determine, for each model user, a number of documents stored in a respective user's cloud storage, type of documents (e.g., an image, word document, spreadsheet, and slides), topicality of the documents, and/or frequency of each type of access (e.g., open, edit, share, comment) to documents. The feature rollout service may determine topicality of a document based on a key text associated with the document. The key text may be indexed and represent a topic of content of the document. The key text may be extracted from a title or content of the document. The key text may be a word, phrase, or sentence. The key text may be manually entered by users associated with the document or

automatically generated. The topicality of a document may be determined from a particular term repeated in a document or content of the document. The topicality may be determined based on knowledge graph entities extracted from the document.

Regarding the common characteristics of the model users from the first set of users, the feature rollout service may determine common characteristics shared by more than a predetermined threshold number of model users. For example, there may be more than the predetermined threshold number of model users (e.g., fifty model users) who have more than a hundred documents stored in a respective model user's cloud storage with more than 50% of the documents being images and all documents being related to healthcare (according to topicality of documents), and who spend more than 80% of the time accessing documents by editing via the GUI of the cloud-based content management platform. The feature rollout service may identify, as common characteristics, the number of documents stored to be at least a hundred documents, the type of documents to be at least 50% images, the topicality to be healthcare, and/or frequency of editing to be at least 80% of the time documents are accessed.

Then, the feature rollout service may then query users who have the same common characteristics (e.g., a user who has more than a hundred documents stored, more than 50% of the documents being images, and/or all documents about healthcare and/or who spends more than 80% of the time accessing documents by editing), and identify them as the second set of users. In this way, the feature rollout service may select users that are likely to have positive experiences and/or frequently use the improved new feature. Additionally, the feature rollout service may impose a maximum number of users to be included in the second set of users. However, the second set of users should be larger than the first set of users. As such, the feature

rollout service may predict the second set of users based on a heuristics or machine learning approach.

Next, at block 150, the feature rollout service may provide access to the improved version of the new feature to user accounts associated with the second set of users. Specifically, the feature rollout service may provide an improved version of the new feature that reflects the user interactions with the older version of the new feature that was provided to the first set of users. Specifically, the feature rollout service may analyze the feedback collected in association with the user interactions and update the new feature to address any negative feedback, fix any bugs, or solve any malfunctions. The feature rollout service may provide the new feature in a similar manner described in details with respect to the block 110 above.

Further to the description 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 collection of user information (e.g., information about user activities on the cloud-based content management platform, information about content of documents stored in the cloud storage, a user's preferences, or a user's current location), and if the user has 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. 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.

**ABSTRACT**

A mechanism is proposed for predicting a subset of users for a feature rollout in a cloud storage system. The method includes identifying a first set of users having user accounts associated with the cloud storage system on a random basis. The method also includes providing access to a new feature of the cloud storage system to user accounts associated with the first set of users. Further, the method includes determining, for each user in the first set of users, user interactions with the new feature. The method furthermore includes predicting a second set of users who are likely to be satisfied with an improved version of the new feature. The second set of users is larger than the first set of users. The improved version of the new feature reflects the user interactions by the first set of users with the new feature. The method includes providing access to the improved version of the new feature to user accounts associated with the second set of users.

**Keywords**: progressive rollout, phased rollout, feature rollout, software deployment, machine learning.

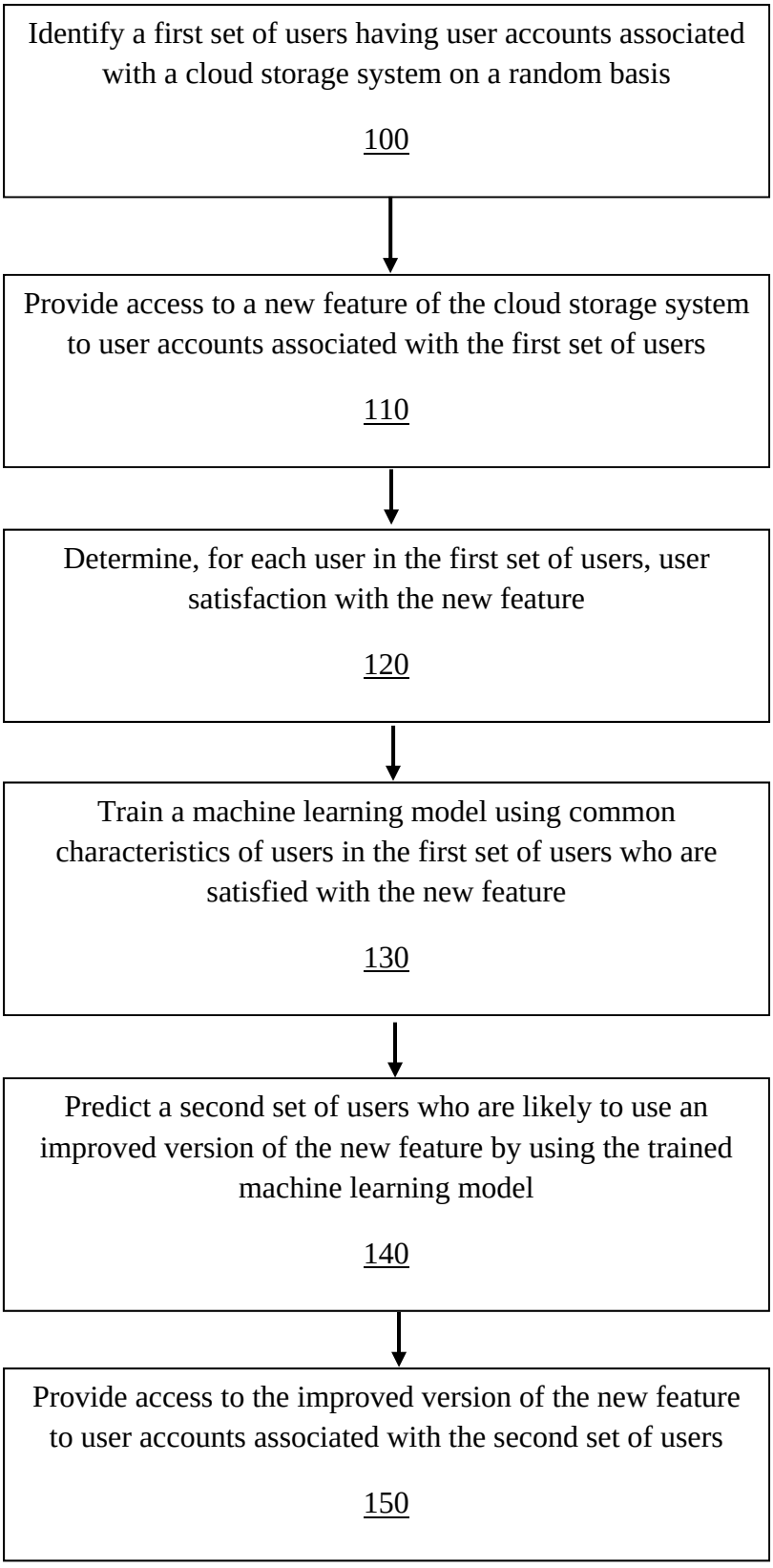
Identify a first set of users having user accounts associated with a cloud storage system on a random basis
100
Provide access to a new feature of the cloud storage system to user accounts associated with the first set of users
110
Determine, for each user in the first set of users, user satisfaction with the new feature
120
Train a machine learning model using common characteristics of users in the first set of users who are satisfied with the new feature
130
Predict a second set of users who are likely to use an improved version of the new feature by using the trained machine learning model
140
Provide access to the improved version of the new feature to user accounts associated with the second set of users
150

**FIG. 1**