**“Many Grantees that Selected Replacement Buses Also Selected Replacement Vans”: Putting AI (Artificial Intelligence) into ALI (Activity Line Item) Selection**

**Abstract**

This paper explores the idea of using a collaborative filtering recommendation system to make the process of applying for FTA funding more efficient and accurate. Artificial Intelligence algorithms similar to the ones used to recommend products to shoppers online could also be employed to recommend specific budget Activity Line Item codes based on which codes frequently co-occur with one another. This paper presents a co-occurrence matrix of ALIs used in FTA applications from FY 2017-2020, identifies the ALI codes that are most frequently used together in recipients’ applications, and describes how AI algorithms could also be used for anomaly detection.

**Introduction to Recommendation Systems and Collaborative Filtering**

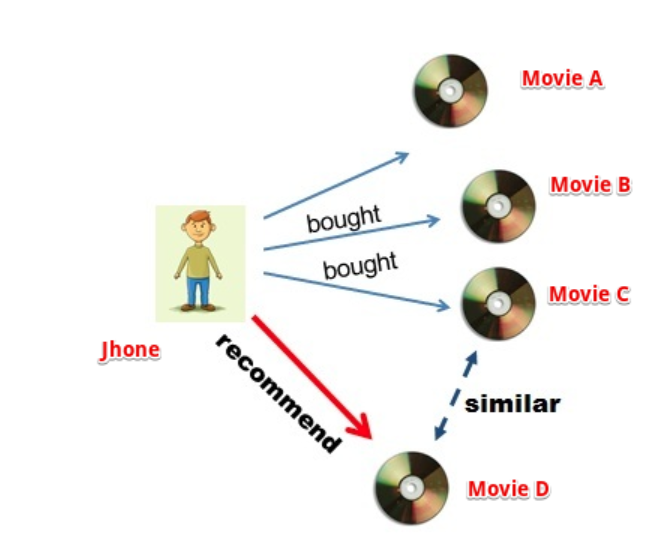
Recommendation systems are commonplace in online. Amazon uses them suggest products to customers, YouTube uses them to decide which video to play next on autoplay, Facebook uses them to recommend pages to like and people to follow, and Netflix uses them to recommend movies you may want to watch.

One of the most common types of recommendation systems, called collaborative filtering, attempts to predict a user’s preference based on preferences expressed by similar users. In the example below User 1 and User 2 purchase an apple and an orange. When User 3 purchases an apple, an algorithm would recommend an orange (vs a banana or a strawberry) because other (presumably similarly situated) users also purchased the orange.



*Source:* <https://medium.com/@tomar.ankur287/item-item-collaborative-filtering-recommender-system-in-python-cf3c945fae1e>

A related method, called content-based filtering would recommend items that are similar to items a user has previously purchased or interacted with. In the example below, Jhone has previously watched Movie A, Movie B, and Movie C. An algorithm identifies Movie C as being similar to Movie D and recommends Movie D.



Source: <https://towardsdatascience.com/comprehensive-guide-on-item-based-recommendation-systems-d67e40e2b75d>

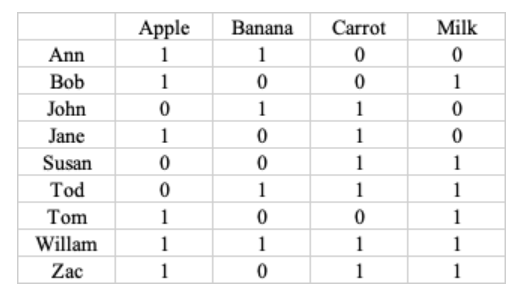
One way in which FTA grant recipient are similar to Amazon shoppers or Netflix users is when they develop a grant or cooperative agreement budget in TrAMS and select budget Activity Line Items (ALIs). A grant application budget is essentially a basket of goods and services that contains at least one ALI, and often includes multiple ALIs. Each ALI is assigned a unique six digit code. (A directory of ALI names and codes is at <https://www.transit.dot.gov/funding/grantee-resources/teamtrams/scope-activity-line-item-tree>. When applicants select ALIs, they are adding items to their “shopping cart.”

The goal of incorporating a recommendation engine to help with ALI selection is *not* to “upsell”, encourage “impulse purchases”, or promote a particular budget activity line item. Transit application budgets are set by state and local planning documents, policies, and budgets and are not subject to the whims of the person developing a grant application. The value proposition of an ALI recommendation engine would be simply to allow grantees to develop application budgets more efficiently and with fewer errors.

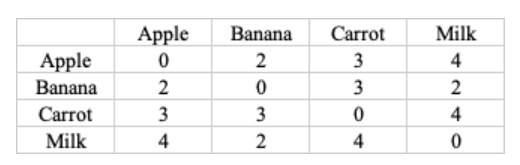
An ALI recommendation algorithms could apply either collaborative filtering (which compares grantees’ ALI selections to one another) or content-based filtering (where a grantee is suggested relevant ALIs based on their selection history). The remainder of this paper discusses a collaborative filtering method.

**Collaborative Filtering Using a Simple Dataset**

Here’s some simple data that may help further explain the concept behind the first item based collaborative filtering example (the one with the fruits). Let’s assume we have data on ten shoppers and we have got some user-history data about what they bought from a grocery store. As shown on the table below, Ann bought an apple and a banana, Bob bought an apple and milk, John bought a banana and a carrot, etc.



The goal of a collaborative filtering algorithm is to find how many times two foods have appeared together in the user data. For example, apple and banana appeared together twice since both Ann and William purchased apples and bananas. Apples and carrots appear together three times because Jane, William, and Zac purchased apples and bananas. The table below transposes the user grocery purchase information into a co-occurrence table.

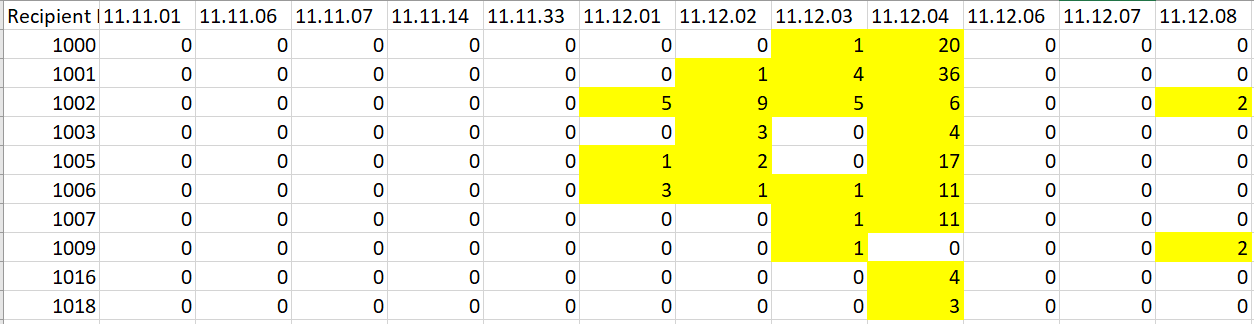


An online grocery store could use this simple data set to make recommendations to shoppers, based on the first item they select. For example, if a shopper selected an apple, the store would recommend milk, which has the highest weighting value of 4, followed by carrot with the weighting value of 3, and banana with the weighting value of 2;

Likewise, for banana, the recommendations are carrot (3), milk and apple (2). For carrot, the recommendations are milk (4), followed by apple, and banana (3). For milk, the recommendations are apple and carrot (4) then banana (2).

**The FTA Activity Line Item Tree**

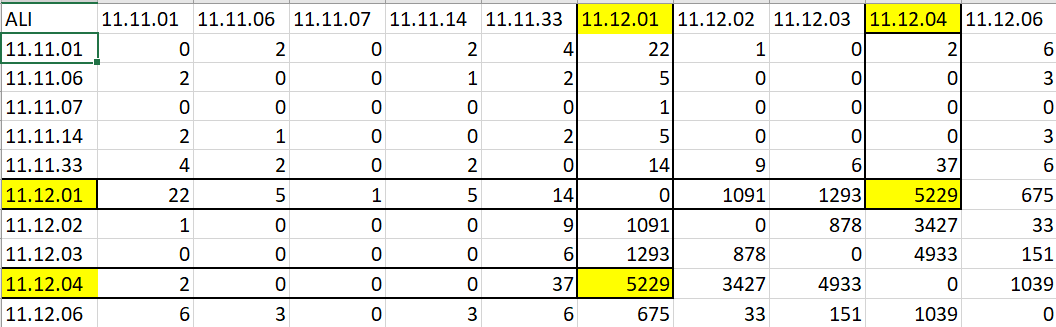
FTA Activity Line Item data is not all that different from the toy data set in the example above, but instead of a table of nine grocery shoppers and four foods, the data includes of 890 organizations that received one or more FTA award from FY 2017 through FY 2020 and 633 unique activity line item codes.[[1]](#footnote-1) The entire table is too large to be shown here, but below is an excerpt of the first ten recipients (from the Alabama Department of Transportation, ID 1000 to the South Alabama Regional Planning Commission, ID 1018) and the first 15 ALIs (from Engineering/designing a 40 foot bus, code 11.11.01 to Buy Replacement Intercity bus, Code 11.12.08).



If a recipient did *not* include an ALI in any of its awards, the cell shows a 0. If a recipient’s awards *did* include an ALI, the cell shows the number of times the ALI was listed. For example, as shown in the first row of the table the Alabama Department of Transportation (ID 1000) did not include engineering and designing a 40-foot bus (11.11.01), articulated bus (11.11.06), commuter bus (11.11.07), dual-mode bus (11.11.14), or ferry boat (11.11.13). The recipient did not include replacing a 40-foot bus (11.12.01), a 35 foot bus (11.12.02), an articulated bus (11.12.06), commuter bus (11.12.07), or intercity bus (11.12.08). It *did* list replacing 30 foot buses once (11.12.03) and included replacing <30 foot buses 20 times in its FY 2020 awards (11.24.04). The ALIs selected are either in all in a single award for recipients receiving one award between FY 2017-FY 2020 or (more likely) spread out across multiple awards in the case of recipients that received more than one grant or cooperative agreement over the course of four fiscal years.

Even though some ALIs show zero in the table extract shown above, at least one of the 890 recipients selected the ALI in order for it to be included on the table. Conversely, if an ALI in FTA’s database was not selected at all in FY 2020, it does not appear on the table (for example, ALI 11.11.02, engineering and designing a standard 35 foot bus was not selected by any recipient and therefore doesn’t appear on the extract shown above).

As is the case with the supermarket example, the ALI data can be transposed into a co-occurrence matrix which shows how frequently two ALIs are paired together. The excerpt below shows the first ten rows and columns from the matrix:



The co-occurrence matrix is also quite sparse. In many instances two ALI codes co-occur with one another rarely or not at all. (ALI codes also do not co-occur with themselves so 0 is shown in the cell where, for example, 11.11.01 and 11.11.01 cross paths). However, there are some ALIs with relatively high co-occurrences. As highlighted above, ALI code 11.12.01 (replacing a 40-foot bus) and ALI code 11.12.04 (replacing an <30 foot bus) are paired together 5,229 times. These two codes appeared together in a recipients’ applications 5,229 times. This means either that many recipients used the two codes across multiple applications and/or the two codes were used multiple times in the same application. Co-occurrence results also appear twice in the matrix since co-occurrences between two items are equal regardless of their order (i.e 11.12.01 co-occurs with 11.12.04 and vice versa).

**The Most Frequently Co-occurring Activity Line Item Codes**

When looking at the dataset as a whole, the greatest number of co-occurrences took place between various ALI codes and the operating assistance ALIs (such as 30.09.01). Operating assistance is a frequently selected ALI in typical fiscal years and it became even more frequently used in FY 2020 as the CARES Act allowed transit agencies that typically are not allowed to spend Federal funds for operating expenses to include it. However, when an item is very widespread, it’s not very useful to include in a recommendation engine, because it doesn’t provide new analytical information and users don’t need a recommendation for something they plan to select anyways.

The chart below shows the 22 ALI pairs that have 5,000 or more co-occurrences (excluding the operating assistance ALIs):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Target ALI Code** | **Target ALI Name** | **Co-occers with (ALI Code):** | **Co-occers with (ALI Name):** | **Number of Co-occurrences** |
| 12.72.11 | Rail Force Account (Other) | 12.79.00 | Project Administration (Rail) | 46,480 |
| 12.72.03 | Project Management (Rail) | 12.79.00 | Project Administration (Rail) | 21,882 |
| 12.72.11 | Rail Force Account (Other) | 12.72.03 | Project Management (Rail) | 20,684 |
| 12.72.08 | Rail Construction (Force Account) | 12.79.00 | Project Administration (Rail) | 14,043 |
| 11.12.04 | Buy Replacement Buses <30 ft | 11.7A.00 | Bus Preventive Maintenance | 13,318 |
| 12.72.11 | Rail Force Account (Other) | 12.72.08 | Rail Construction (Force Account) | 13,082 |
| 11.12.04 | Buy Replacement Buses <30 ft | 11.12.15 | Buy Replacement Vans | 11,509 |
| 12.34.02 | Rehab/Renovate Rail Station | 12.79.00 | Project Administration (Rail) | 10,594 |
| 12.72.11 | Rail Force Account (Other) | 12.34.02 | Rehab/Renovate Station (Rail) | 10,170 |
| 11.12.04 | Buy Replacement Buses <30 ft | 11.80.00 | State Program Administration | 9,204 |
| 12.24.03 | Rehab/Renovate Line Equipment | 12.79.00 | Project Administration (Rail) | 8,927 |
| 12.24.03 | Rehab/Renovate Line Equipment | 12.72.11 | Rail Force Account (Other) | 8,270 |
| 12.72.08 | Rail Construction (Force Account) | 12.72.03 | Project Management (Rail) | 6,359 |
| 11.7L.00 | Mobility Management (Bus) | 11.12.04 | Buy Replacement Buses <30 ft | 6,311 |
| 11.12.15 | Buy Replacement Vans | 11.80.00 | State Program Administration (Bus) | 5,971 |
| 11.12.01 | Buy Replacement Bus STD 40 ft | 11.7A.00 | Preventive Maintenance (Bus) | 5,774 |
| 11.12.15 | Buy Replacement Vans | 11.13.15 | Buy Expansion Vans | 5,662 |
| 11.42.20 | Acquire Miscellaneous Support Equipment (Bus) | 11.12.04 | Buy Replacement Buses <30 ft | 5,649 |
| 11.7A.00 | Preventive Maintenance (Bus) | 11.12.15 | Buy Replacement Vans | 5,447 |
| 11.42.20 | Acquire Miscellaneous Support Equipment (Bus) | 11.7A.00 | Preventive Maintenance (Bus) | 5,446 |
| 11.12.01 | Buy Replacement Bus STD 40 ft | 11.12.04 | Buy Replacement Buses <30 ft | 5,229 |
| 11.80.00 | State Program Administration (Bus) | 11.7L.00 | Mobility Management (Bus) | 5,058 |

The 22 ALI codes fall within two main categories:

1) Rail construction ALIs showing many co-occurrences between rail construction force account, rehab/renovate stations, rehab/renovate line equipment, project management and project administration.

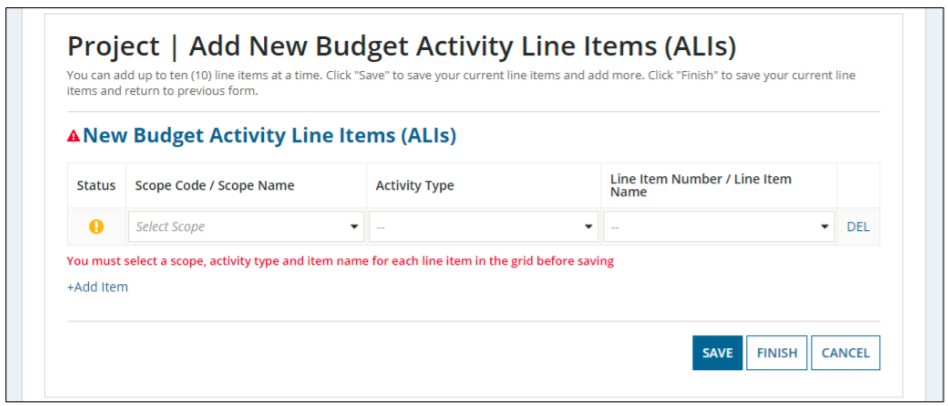
These co-occurrences likely exist due to FTA guidance on the use of force account (i.e. using a transit agency’s own labor force in lieu of a 3rd party contractor). According to FTA’s Grants Management Circular 5010.1E, force account can include design, construction, overhaul inspection, and construction management activities, but cannot include project administration. However, since project administration is often part of a construction project, it is listed as a separate budget activity line item.

2) Bus ALIs showing many co-occurrences with different types of rolling stock purchases (i.e. purchasing 40 foot and <30 foot buses and vans) as well as co-occurrences between rolling stock purchases and preventive maintenance, acquiring support equipment, state program administration, and mobility management. This may be due to transit agencies with different types of vehicles that need to be replaced, as well as the need to do preventive maintenance on vehicles that have been purchased. In addition, State Departments of Transportation are eligible to use a portion of their Federal funds for program administration and may also be more likely to implement mobility management.

**The Value Proposition for a Recommendation Engine**

Adding a recommendation engine that uses collaborative filtering, content based filtering, or some other method to the TrAMS budget module could make it easier and less time consuming to develop application budgets and may reduce the odds that application budgets include incorrect ALIs (and the time it takes to make corrections) due to human error. The algorithms could also be used for anomaly detection, flagging ALI codes that are rarely selected and prompting a message to the applicant to double-check their selection to confirm.

In order to add ALI codes in TrAMS, users must navigate to their application’s budget module where they access the fields in the screen shown below. To add an ALI, users select information from all three drop downs (scope code/scope code name, activity type, and line item number/line item name). In order to add additional ALIs, users click the “add item” button and repeat the process. Once users click “add item” a new set of drop-downs appears that are defaulted to the blank fields shown below.



Adding a single ALI is not very time consuming, however most applications contain multiple ALIs with an average of 11 per award and 237 awards contained 20 or more ALIs. Developing an application budget using the existing design can become tedious and repetitive.

On the other hand, a recommendation algorithm that prompts users to select from a short list of ALIs based on initial selections could provide a useful shortcut. For example, once a user uses the existing module to select ALI 11.12.04 (Purchase a <30 foot bus for replacement) the recommendation engine would automatically provide the ALIs that are most likely to co-occur with 11.12.04, which are:

30.09.01 (Operating Assistance)

11.7A.00 (Bus Preventative Maintenance)

11.12.15 (Purchase vans)

11.80.00 (Bus Project Administration)

In another example once a user selects ALI 12.27.11 (Rail Force Account), the engine would provide the ALIs that most frequently co-occur:

12.79.00 (Rail Project Administration)

12.72.08 (Rail Construction)

12.34.02 (Rehab/Renovate Rail Station)

12.24.03 (Rehab/Renovate Line Equipment)

In either case, the user should be able to click on the ALI(s) recommended instead of having to work through the drop-downs in the module, or the user could ignore the recommendations.

The shortcut provided by the recommendation engine could reduce “fat finger” errors that may result in applications being returned to the recipient after an FTA reviewer spots the mistake. (FTA staff cannot change budget information on their own, instead they need to route the application back to the recipient with a comment and have the applicant re-send the corrected version to FTA)

Machine learning algorithms could also be used for anomaly detection (spotting information that is rarely present in the data and could be an indication of a mistake). The ALI dataset is a good candidate for this practice because many ALI codes have been used only sparingly, if at all. The chart below shows the distribution of the most frequently selected ALIs.

Operating assistance, preventive maintenance, and rolling stock purchase ALI codes are used frequently, but only 57 out of 633 ALIs (9%) were included in applications 100 times or more. And there are 373 ALI codes that were selected 10 times or less. It’s possible that some of these codes are correct and represent budgets for relatively unique applications, but it may also be the case that the codes were incorrectly selected. What’s more there are a total of 1314 ALI codes available for selection, which means 681 codes have never been selected even once in the past four fiscal years.

An anomaly detection algorithm could flag infrequently selected ALI code and provide an “are you sure you want to select this ALI code?” message to the application developer. It could also be used to weed the ALI table of codes that have become obsolete, duplicative or cause confusion.

**Next Steps**

Implementing an ALI code recommendation engine is not a particularly urgent issue, but FTA may want to investigate further (including the feasibility and cost of adding an engine in the Appian platform) in the course of thinking about IT strategic priorities or future projects. This topic could also be brought up if FTA receives questions from the US DOT CIO or other offices about the potential for incorporating AI into agency operations.

1. The dataset used in this analysis is the Application Budget by ALI Report for 9-30-2020 which can be found in the Static Reports section in TrAMS. It includes awards made from November 2016—September 2020, approximately 5,800 awards containing 63,600 ALIs. [↑](#footnote-ref-1)