Doing groceries again: Towards a decision augmentation for grocery stores selection: Supplement

Daniyal Kazempour, Melanie Oelker, Peer Kröger

1 On the Utility Score

In this supplmentary we provide additional information on the utility score including examples and elaborations on the boundaries of each term. Given a store location loc_i the utility score is computed as:

$$\mu(loc_i) = \underbrace{\frac{|loc_{visited} \in \mathcal{N}(loc_i)|}{|sm_{topk}|}}_{①} + \underbrace{\sum_{i=0}^{k-1 \in \mathcal{N}(loc_i)} \frac{f(loc_{visited_i})}{f(loc_{visited_{total}})}}_{②} + \underbrace{\sum_{i=0}^{k-1 \in \mathcal{N}(loc_i)} (d_{lastvisited}(loc_{visited_i}))^{-1}}_{③} \tag{1}$$

Part ① Example: we determine from the Google timeline data of a user the top-5 visited supermarkets over the past (denominator). In the vicinity of a store location loc_i from our query, we observe 3 stores from 3 distinct companies. This means that the value of ① would be in this case $\frac{3}{5}$. This part of the equation is bounded within the interval $[0, 1]$, namely 0 for the case that there are no stores from other companies of the top-k visited supermarkets in the vicinity of loc_i and 1 in case that all top-k visited supermarkets are in the vicinity of loc_i.

Part ② Example: we have as top-k $k = 3$ with the 3 different stores stor1, stor2 and stor3. Stor1 has been visited (based on the user's timeline data) 12 times, while stor2 has been visited 18 times and stor3 10 times. In total the user visited all three (top-3) stores $12 + 18 + 10 = 40$ times (denominator). In neighborhood of the store loc_i we have only stor1 and stor2, meaning that the result of ② is $\frac{12}{40} + \frac{18}{40} = \frac{30}{40}$. This part of the sum is as well-bounded to $[0, 1]$. In the case of 0 no stores of a particular company are in the neighborhood of loc_i, and in the case of 1 all top-k stores are in its vicinity. Contrary to ①, in ② it is no longer the number of stores of different companies that are accounted for but the frequency. As a consequence, we may have only a fraction of stores from the top-k in the vicinity of loc_i but still have a high value close to 1 given the case that those few stores of particular companies in the vicinity of loc_i are visited most frequently.

Part ③ Example: stor1 was visited 6 days ago, meaning that we have $\frac{1}{6}$ for stor1. For stor2 it has been 11 days since the last time a store of company stor2 was visited by the user, yielding $\frac{1}{11}$. The sum from the values of stor1 and stor2 yield $\frac{1}{6}+\frac{1}{11} = \frac{17}{66}0.26$. The inverse of the number of days is taken in order to create a bounding and to provide a meaningful semantic behind the score as we shall elaborate in the upcoming lines. This part of the sum is bounded to the interval of $[0, |loc_{visited} \in \mathcal{N}(loc_i)|]$. Given the case that for all top-k stores of particular companies in the neighborhood of loc_i the number of days since last visited is infinite, the term would converge towards 0. In the case that for each of the stores in neighborhood the number of days since last visited is 1, we would have the number of stores in the neighborhood $|loc_{visited} \in \mathcal{N}(loc_i)|$.

Overall the boundary of $\mu(loc_i)$ is $[0, |loc_{visited} \in \mathcal{N}(loc_i)| + 2]$.