RightFit: Reduce Product Returns Attributed to Size Issue

Your Name

November 25, 2015

1 Notation

Let $X = \{x_1, x_2, \ldots, x_{|X|}\}$ be the set of all "SKU simples". Set $S = \{s_1, s_2, \ldots, s_{|S|}\}$ be the set of all SKU's, such that, associated with each $x \in X$ is a SKU $sku(x) = s$, where, $s \in S$. Also, $X_s = \{x \in X | sku(x) = s\}$ is the set of all SKU simples belonging to SKU s. Let $U = \{u_1, u_2, \ldots, u_{|U|}\}$ be the set of all users. For a user $u \in U$ and a SKU simple $x \in X$, we say that $buy(u, x) = 1$, if x was bought by u. We define the set $X_u = \{x \in X | buy(u, x) = 1\}$ and $X_u \subset X$, as the set of all SKU simples bought by user u.

Let $C = \{c_1, c_2, \ldots, c_{|C|}\}$ be the set of all clusters of X, $|C| < |X|$ and let $\{\mu_1, \mu_2, \ldots, \mu_{|C|}\}$ be the corresponding cluster centers (mean or medoid). Each $x \in X$ belongs to a cluster $c \in C$, and is given by $c = cluster(x)$.

Then, the number of times a user $u \in U$ bought a SKU simple from a cluster $c \in C$ is given by:

$$cnt(u, c) = |\{x \in X_u | cluster(x) = c\}|$$

We say that a cluster c is a "preferred cluster" of a user u if $cnt(u, c) >= 1$, that is, the user has bought at least one SKU simple belonging to the cluster c. We also define the set $C_u = \{c \in C | cnt(u, c) >= 1\}$ as the set of all preferred clusters of user u. Next, we define the support between a user u and a cluster c as the adjusted count:

$$sup(u, c) = \frac{cnt(u, c) + 2}{|X_u| + 4} \tag{1}$$

Let $dist(x, c)$ be the distance of a SKU simple x from a cluster c. For instance, $dist(x, c) = euc(x, \mu_c)$ could be the standard euclidean distance between x and the cluster center of c. We define the "closeness" of x to cluster c as the normalized distance:

$$close(x, c) = 1 - \frac{dist(x, c)}{\max_{\forall x_i \in X_s} dist(x_i, c)} \tag{2}$$