

# Supplementary Material-An Efficient Person Clustering Algorithm for Open Checkout-free Groceries

Junde Wu<sup>1</sup>, Yu Zhang<sup>2</sup>, Rao Fu<sup>2</sup>, Yuanpei Liu<sup>3</sup>, and Jing Gao<sup>1\*</sup>

<sup>1</sup> Purdue University, West Lafayette, IN, 47907, USA [jinggao@purdue.edu](mailto:jinggao@purdue.edu)

<sup>2</sup> Harbin Institute of Technology, Harbin, 150080, China

<sup>3</sup> Beijing Institute of Technology, Beijing, 102213, China

**Abstract.** This supplementary material provides more detailed theoretical explanations of Cluster Feature vector.

## 1 Cluster Feature vector

As mentioned in the manuscript, Cluster Feature vector (CF vector) is a commonly used data structure for summarizing large amounts of data. In this part, we explained its incrementality and additivity properties in detail (For the convenience of reading, we start with the definition of CF vector).

The CF vector has three components:  $K$ , the number of data objects,  $LS$ , the linear sum of the data objects, and  $SS$ , the sum of squared data objects. The structures  $LS$  and  $SS$  are  $n$ -dimensional vectors. These three components allow to compute cluster measures, such as cluster mean  $\mu$  (Eq. (1)) and radius  $\sigma$  (Eq. (2)).

$$\mu = \frac{LS}{K} \quad (1)$$

$$\sigma = \sqrt{\left(\frac{SS}{K} - \left(\frac{LS}{K}\right)^2\right)} \quad (2)$$

where  $(\cdot)^2$  and  $\sqrt{\cdot}$  represent element-wise square and square root. The CF vector is widely used for its incrementality and additivity properties, as described follows:

The CF vector is widely used for its incrementality and additivity properties [1], as described follows:

---

\* Junde Wu and Yu Zhang have equal contributions.  
Jing Gao is the corresponding authors

(1) *incrementality*: A new object  $x_j$  can be easily inserted into CF vector by updating its statistic summaries as follows:

$$\begin{cases} LS \leftarrow LS + x_j \\ SS \leftarrow SS + (x_j)^2 \\ K \leftarrow K + 1 \end{cases} \quad (3)$$

(2) *additivity*: Two disjoint vectors  $CF_1$  and  $CF_2$  can be easily merged into  $CF_3$  by summing up their components.

$$\begin{cases} K_3 \leftarrow K_1 + K_2 \\ LS_3 \leftarrow LS_2 + LS_1 \\ SS_3 \leftarrow SS_2 + SS_1 \end{cases} \quad (4)$$

## References

1. Silva, J.A., Faria, E.R., Barros, R.C., Hruschka, E.R., Carvalho, A.C.d., Gama, J.: Data stream clustering: A survey. *ACM Computing Surveys (CSUR)* **46**(1), 1–31 (2013)