Novel trajectory clustering method based on distance dependent Chinese restaurant process

Reza Arfa1, 2, Rubiyah Yusof1, 2, Parvaneh Shabanzadeh1, 2

1 Centre for Artificial Intelligence and Robotics, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia

2 Centre for Artificial Intelligence and Robotics, Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia, 54100, Kuala Lumpur, Malaysia

Corresponding Author:

Rubiyah Yusof1, 2, Reza Arfa1, 2

Email address: rubiyah.kl@utm.my, rezaarfa@gmail.com.

**Supplementary Material**

This supplementary material provides more detail about collapsed Gibbs sampling procedure given in Equation 6.

The steps involved in resampling are removing ’s outgoing link and resample its new outgoing link using Equation 6. There are only two scenarios that might happen when resampling . The first one happens when the new outgoing link does not change the customer partition. This scenario happens if links to itself (i.e. ) or other customers belonging to the same partition of. In this case, the partition remains the same. Therefore, no further action is required. Another possible scenario is when a new outgoing link joins two partitions. Since in this scenario the customer partition has changed, the values of prior and likelihood of partitions will also change and need to get updated.

By substituting Equations 2 and 5 to Equation 6 the collapsed sampling to resample can be compactly represent as follow:

(1A)

Where

(2A)

Using the Dirichlet-Multinomial Conjugacy, can be written as (Teh, Newman, and Welling 2007)

(3A)

Where in this equation is the gamma function, is the concentration parameter of symmetric Dirichlet distribution, is the total number of grid cells, is the number of times th grid cell is observed in the table , and is the number of all grid cells assigned to table .