Mathematical formula for the metric Relative expression across cell type clusters

Anonymous Authors[1]

Abstract

This document provides the Suggestion of Asli and Katelyn for the Relative expression across cell type clusters metric the idea for the similarity metric: Relative expression levels/probe efficiency is the same by converting the pairwise differences per gene for each cell to pairwise differences per cell for each gene.

0.1. Relative expression across cell type clusters

Let n_c be the different numbers of clusters. Let $U = U_1, ..., U_{n_c}, V = V_1, ..., V_{n_c}$ be the set of sets of cells in each cluster for single-cell data and the spatial data. We have $|U| = |V| = n_c$.
Denote the shared genes as $I = I_1, ..., I_{n_c}$ as $I = U \cap V$ and subset U to V with $U = U \cap I$.

Define the mean expression levels for each modality as M_{sc} and M_{sp} :
$M_{sc} = \frac{(\sum_{i=0}^{n_c} U_i)}{n_c}$ and $M_{sp} = \frac{(\sum_{i=0}^{n_c} V_i)}{n_c}$

Calculate the pairwise difference between cell types for each column using the following idea:

For each unique pair of columns (col_i, col_j) from both M_{sc} and M_{sp} such that $i < j$. We calculate the difference in between and store the values in P_{sc} which represents the pairwise differences per gene for single cell data and analog for the spatial data: P_{sp}:

$P_{sc} = (p_{kj})$ with $(p_{kj})_{1 \leq k \leq \frac{c!}{(c-2)!*2!}, 1 \leq j \leq g}$

where c= Number of cells and g= Number of genes in M_{sc}. For $M_{sc} = (m_{il})_{1 \leq i \leq c, 1 \leq l \leq g}$ we have

$p_{kj} = \sum_{k=1}^{\frac{c*(c-1)}{2}-1} \sum_{j=i+1}^{g} [m_{j,col_i} - m_{k,col_l}]$

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

and analog for P_{sp}.

To define the mean normalized relative expression matrix for each gene, we define:

$$\mathfrak{M}_{i,j}^{sc} = \left(\frac{p_{i,j}}{\sum_{s=1}^{\frac{c*(c-1)}{2}} p_{s,j}} \right)_{1 \leq i \leq \frac{c*(c-1)}{2}, 1 \leq j \leq g}$$

Analog for $\mathfrak{M}_{i,j}^{sp}$.

As last step we define $S_i = \frac{|\mathfrak{M}^{sc} - \mathfrak{M}^{sp}|}{|\mathfrak{M}^{sc} - \mathfrak{M}^{sp}|}$