



## CPS Paper

### Dimensionality reduction using the ordered label for trajectory inference

**Author:** Mr Masaaki Okabe

**Coauthors:** Hiroshi Yadohisa

**Submission ID:** 794

**Reference Number:** 794

#### Brief Description

We propose a dimensionality reduction method that assumes that labels with an ordering structure are given as external information.

We assume a situation in which information from which samples were obtained is given as ordered labels.

The noise may be added to the labels.

In this situation, the label does not necessarily have a relationship with the true trajectory.

Our method solves this problem by performing supervised dimensionality reduction robust to label perturbations.

#### Abstract

Dimensionality reduction is a method to represent high-dimensional data in a lower dimension and is used in various places for interpretation. Trajectory inference is one of the tasks using dimensionality reduction methods. It is a method used to determine the dynamic process patterns of cells and to position cells based on the progression of these processes. In this task, dimensionality reduction is used on biological data such as genomic data and single-cell RNA sequence data. More generally, trajectory inference can be thought of as a problem of putting the samples in order. For example, Monocle3, a typical method used in trajectory inference, requires constructing a tree structure after dimensionality reduction using UMAP. These methods do not use the label data as external information. Therefore, the final result depends on the accuracy of the dimensionality reduction. On the other hand, another method uses existing supervised trajectory estimation, for example, pspertime. This method can find variables associated with labels. However, the information on the label does not necessarily represent the true sample order. In this research, we propose a dimensionality reduction method that assumes that labels with an ordering structure are given as external information. For example, we assume a situation in which information on clusters from which samples were obtained is given as ordered labels. In this situation, perturbations such as noise may be added to the labels. In other words, in this situation, the label does not necessarily have a relationship with the true trajectory. Therefore, if we perform dimensionality reduction to represent the label information better, we may overfit the label information and fail to estimate the trajectory that we want to estimate. In trajectory inference, the label information is expected to be used as a supplement to obtain information related to biological processes. This method solves this problem by introducing a weight parameter for label information and performing supervised dimensionality reduction robust to label perturbations. Numerical experiments using the ordinal correlation between the visualization of the dimensionality reduction results and the true sample order as an evaluation index demonstrate the method's usefulness.