

Learning a metric for clustering structured data

Marek Dědič¹²

¹ČVUT v Praze, Fakulta jaderná a fyzikálně inženýrská, Matematická informatika

²Cisco Systems Inc., Karlovo náměstí 10, Praha 2

October 11, 2019



- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss

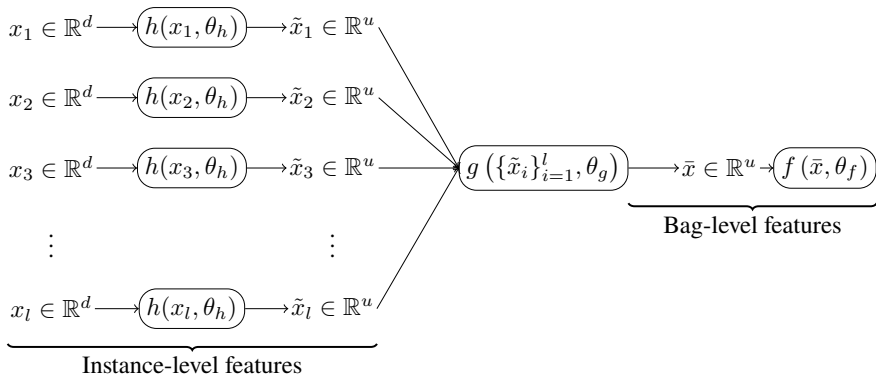
Motivation

- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss

Motivation

- Enable clustering of IPs from SwFlows - enhance Correlation
- Take advantage of the structure of the data using MIL
- Try out unsupervised learning

- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss



Contrastive predictive coding

- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss

Contrastive predictive coding

The ideas of contrastive predictive coding were originally adapted giving a loss function

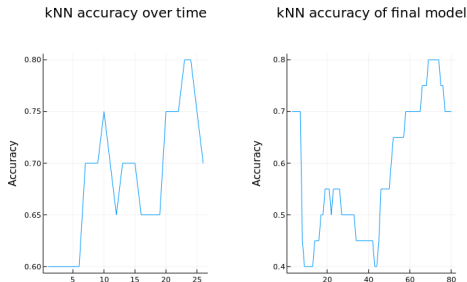
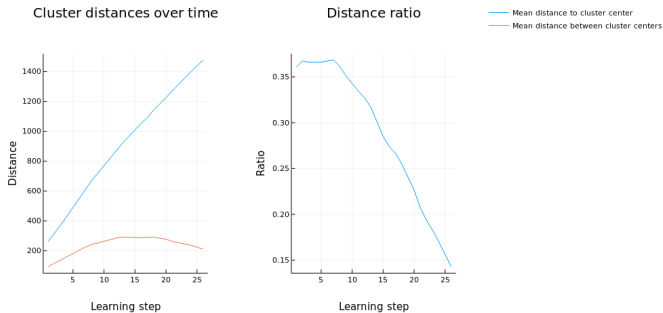
$$\log \left\| f \left(B_n^{(1)} \right) - f \left(B_n^{(2)} \right) \right\|^2 - \log \sum_{j=1}^K \left\| f \left(B_n^{(1)} \right) - f \left(B_j' \right) \right\|^2$$

Contrastive predictive coding

Later on, the loss function was simplified to

$$D_{ij} = \left\| f \left(B_i^{(1)} \right) - f \left(B_j^{(2)} \right) \right\|_2^2$$
$$\frac{1}{n} \sum_{i=1}^n \left(\log (D_{ii}) - \log \left(\sum_{i \neq j} D_{ij} \right) \right)$$

Contrastive predictive coding - preliminary results (Musk)



Triplet loss

- Motivation
- MIL
- Contrastive predictive coding
- Triplet loss

Triplet loss

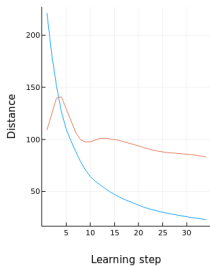
The triplet loss is another alternative, this time requiring supervised learning

$$y_{ij} = \begin{cases} 1 & \text{for } y_i = y_j \\ 0 & \text{otherwise} \end{cases}$$

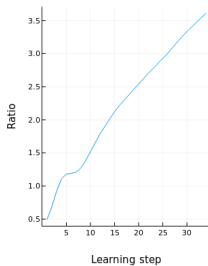
$$\sum_{ij} y_{ij} D_{ij} + c \sum_{ijl} y_{ij} (1 - y_{il}) \max(0, 1 + D_{ij} - D_{il})$$

Triplet loss - preliminary results (Musk)

Cluster distances over time

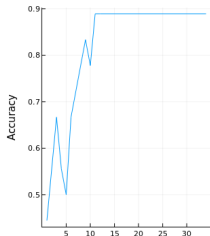


Distance ratio



— Mean distance to cluster center
— Mean distance between cluster centers

kNN accuracy over time



kNN accuracy of final model

