IEEE WCCI 2020

IEEE World Congress on Computational Intelligence

Virtual Conference – July 19-24, 2020

Deep Categorization with Semi-Supervised Self-Organizing Maps

Pedro H. M. Braga, Heitor R. Medeiros and Hansenclever F. Bassani

{**phmb4**, hrm, hfb}**@cin.ufpe.br**

Centro de Informática - Universidade Federal de Pernambuco (CIn - UFPE)

July, 2020

Outline

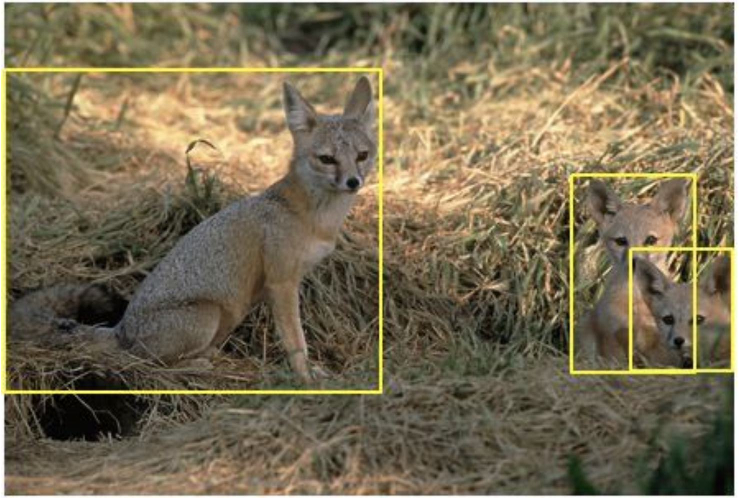
1. Introduction
2. Background
3. Batch SS-SOM
4. Experiments and Results
5. Conclusion and Future Work

Introduction

- Recent research on Artificial Neural Networks with **supervised learning** has shown **significant advances**.
  - A **key** to the **success** is the **sufficiently** large **labeled** training **data**.
- Unfortunately, **labeling is a difficult task**.
- The use of supervised learning became **impractical** in many applications:
  - **Medical field**.
  - **Robotics**: new categories of elements may frequently arise.

Introduction

- However, there is a plentiful amount of unstructured data available.
- Competitions like [ImageNet Object Localization Challenge]() from Kaggle tries to encourage this kind of practice.
- The challenge is to identify all objects within an image, so those images can then be classified and annotated.

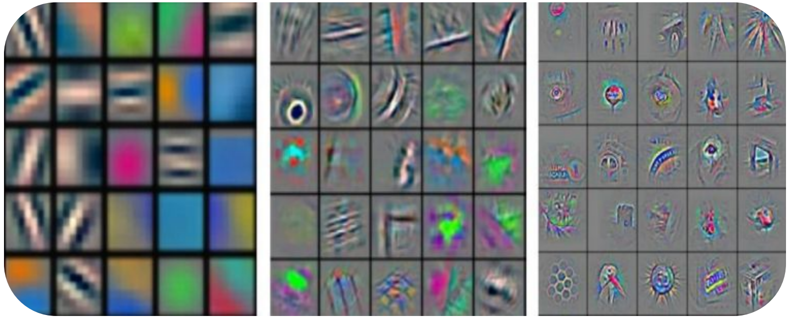


Introduction

- However, there is a plentiful amount of unstructured data available.
- There has been a growing interest in **Semi-Supervised Learning (SSL).**
- **Combine** both types of **data** to improve the performance of models.
- A **halfway between supervised** and **unsupervised** learning, that can be used to both **clustering** and **classification** tasks.

Introduction

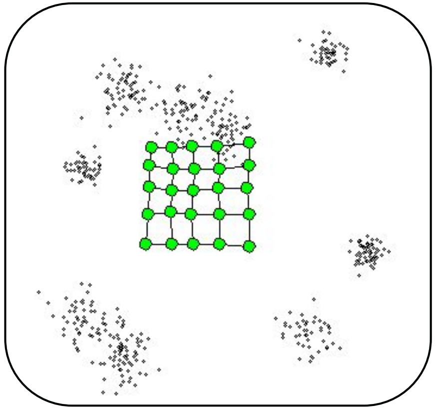
- Moreover, not only the raw data itself can be used, but also its characteristics (features), or learned representations (Bengio et al., 2013).
- Techniques based on Deep Learning have been very successful in yielding good high-level representations (Aljalbout et al., 2019).

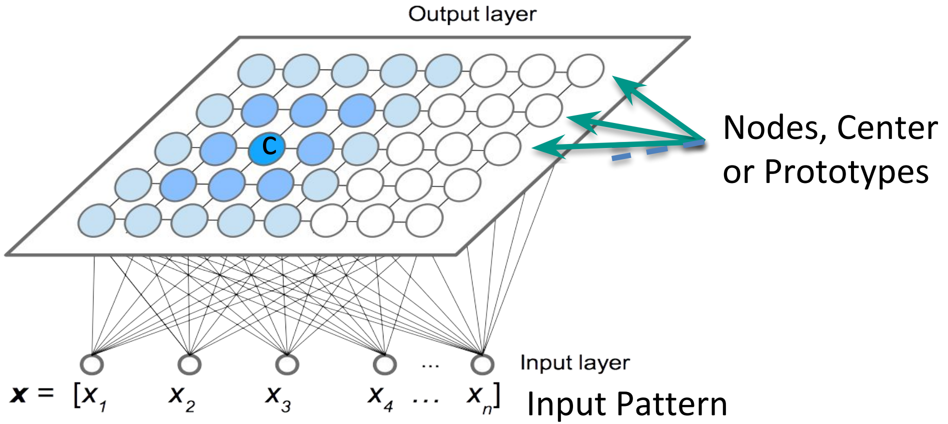


Introduction

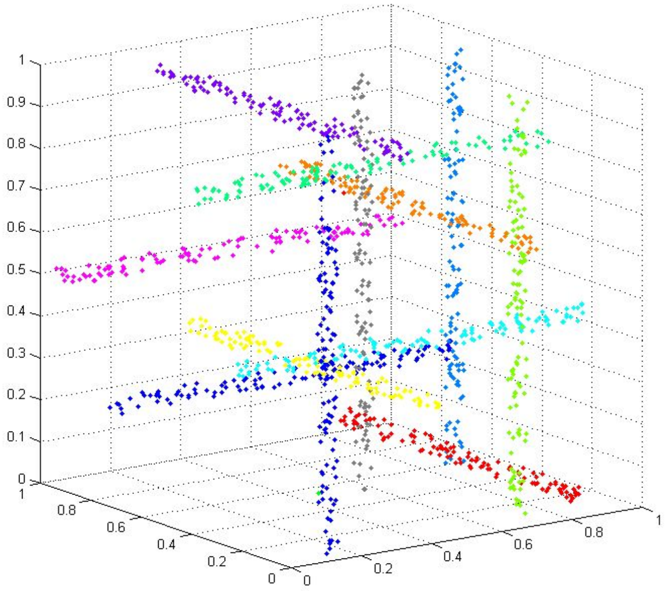
- So, why not scale traditional models to a new variety of applications?
- We highlight **SOM-based models** to start doing that (Liu et al., 2015; Braga & Bassani, 2018).

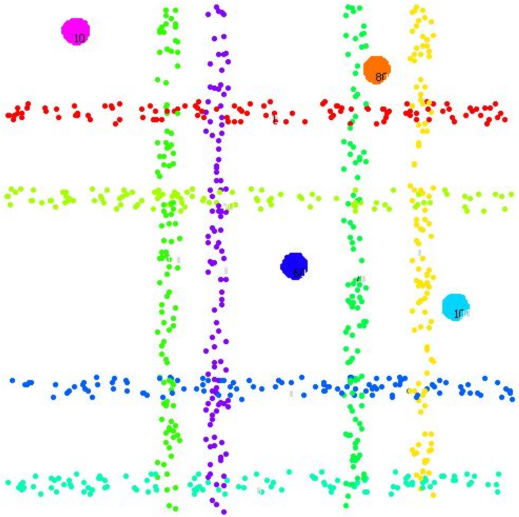
Background
[Self-Organizing Maps (SOM)]





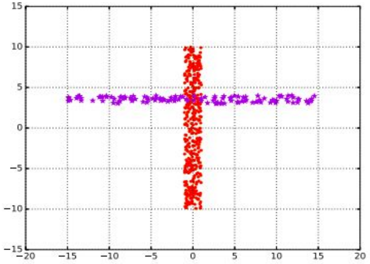
Background
[Relevance Learning]



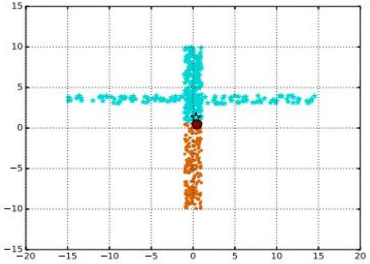


Background
[State of the Art Self-Organizing Maps]

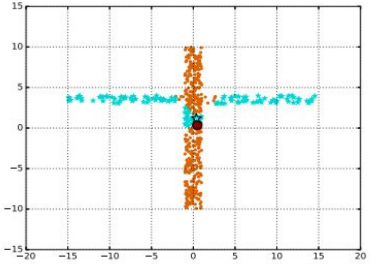
- **LARFDSSOM [Bassani and Araujo, 2015]**
  - Purely Unsupervised.
  - Applies local relevances to the input dimensions.
  - Time Varying-Structure.



Input



SOM



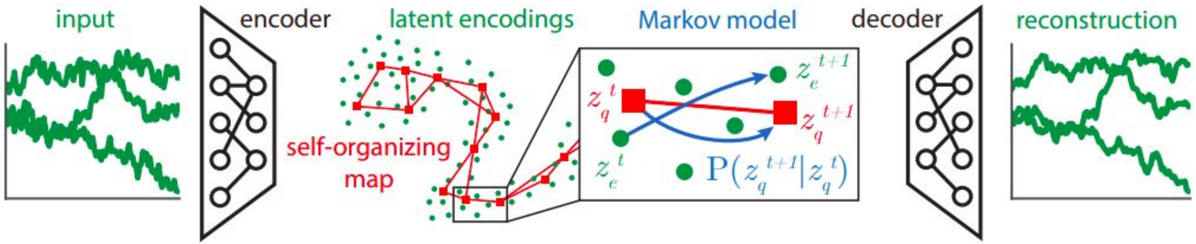
LARFDSSOM

Background
[State of the Art Self-Organizing Maps]

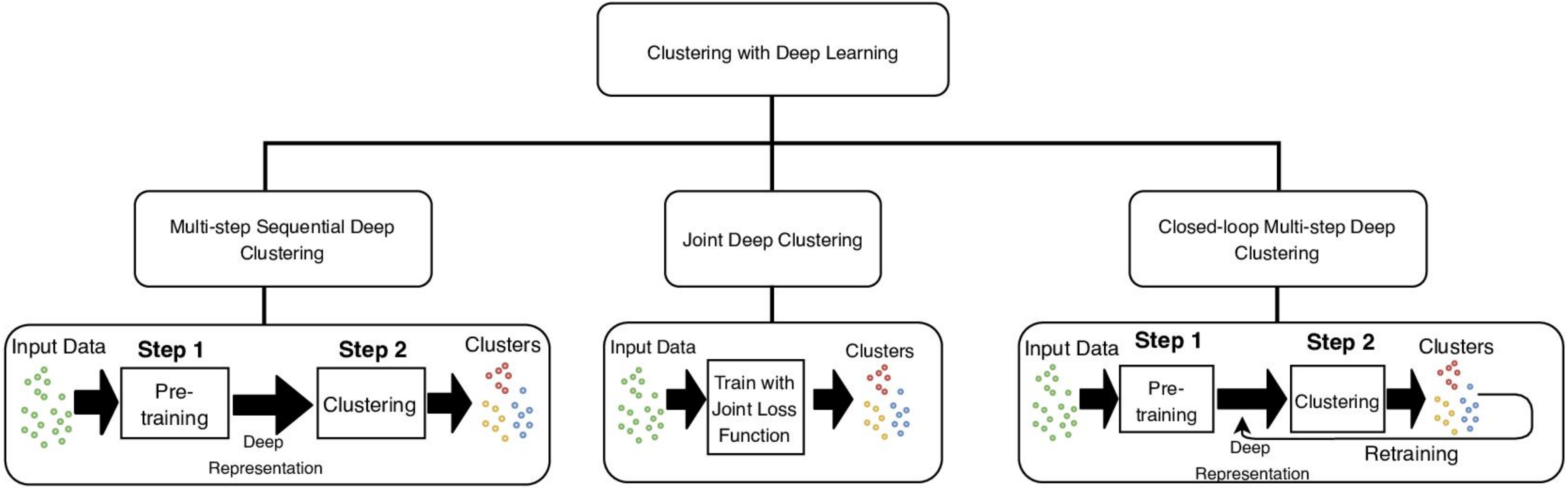
- **SS-SOM [Braga and Bassani, 2018]**
  - A Semi-Supervised Model.
  - **Includes concepts from LVQ** when the **class label** of some input pattern **is given**
  - It can switch between a **supervised** or **unsupervised** learning procedure during the self-organization process

Background [Representation Learning and SOM]

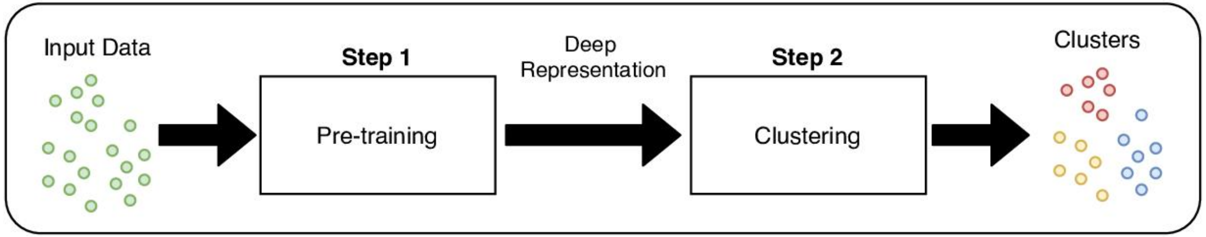
- Aljalbout et al. (2018) have shown that it is possible to learn interpretable latent representations when combining an autoencoder or generative models with SOM.
- For instance, **SOM-VAE** (Fortuin et al., 2018).



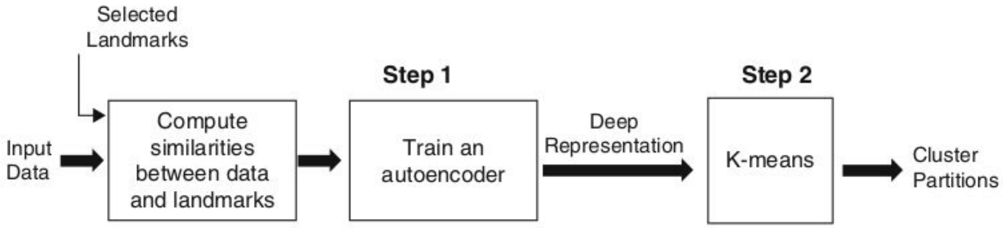
Background
[Deep Clustering]



Background
[Multi Step Sequential Deep Clustering]

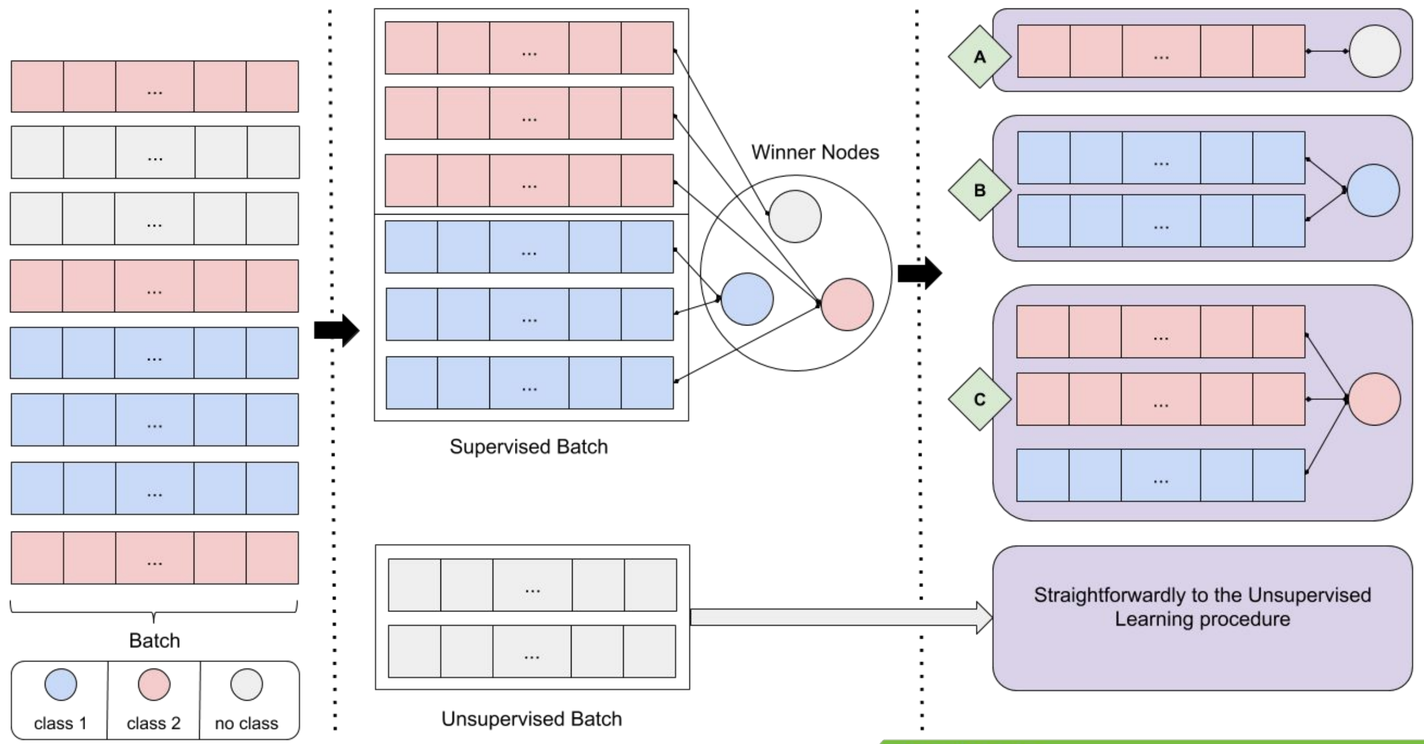


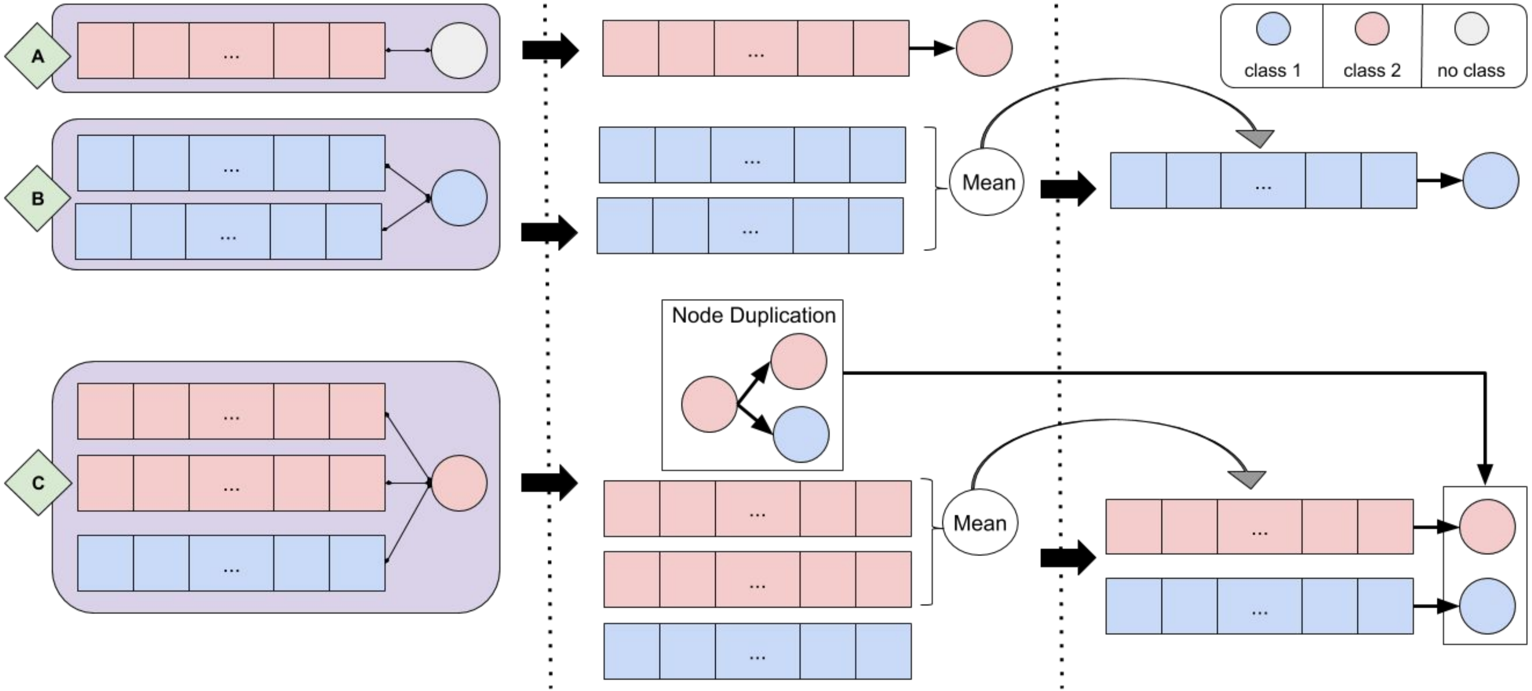
Fast Spectral Cluster



Batch SS-SOM

- Deals with Batches of Samples.
- Can be easily coupled to Deep Learning Architectures.
- Takes advantage of Graphics Processing Units (GPUs)
- Three important modifications
  1. When a batch is given to the model, it is separated into the unsupervised and supervised batch.
  2. The computation of average vectors, of the samples with the same label, for each prototype.
  3. A new node duplication scheme.


Batch
class 1
class 2
no class
Supervised Batch
Winner Nodes
Unsupervised Batch
A
B
C
Straightforwardly to the Unsupervised Learning procedure


A
B
C
Node Duplication
Mean
Mean
class 1
class 2
no class

Experiments [Setup]

- Parameter Sampling:
  - **LHS**
  - Batch Size of 32
- Datasets
  - UCI
  - MNIST, Fashion-MNIST and SVHN
- Runs
  - 500 over UCI Datasets
  - 10 over Image Datasets

TABLE I: Parameter Ranges of BATCH SS-SOM

Parameters	**min**	**max**
Activation threshold (a_t)	0.90	0.999
Lowest cluster percentage (lp)	0.001	0.01
Relevance rate (β)	0.001	0.5
Max competitions (age_wins)	$1 \times S^*$	$100 \times S^*$
Winner learning rate (e_b)	0.001	0.2
Wrong winner learning rate (e_w)	$0.01 \times e_b$	$1 \times e_b$
Neighbors learning rate (e_n)	$0.002 \times e_b$	$1 \times e_b$
Relevance smoothness (s)	0.01	0.1
Connection threshold ($minwd$)	0	0.5
Number of epochs ($epochs$)	1	100

* S is the number of input patterns in the dataset.

Experiments [Parameters Sensitivity]

- A sensitivity analysis presented in Bassani and Araujo (2015) revealed that only $\boldsymbol{a_t}$ and ***lp*** presented a high impact on the results.

- So that, we can **keep** the **other parameters** values **fixed inside** the **ranges** defined in **Table 1**, given their **marginal influences.**

TABLE I: Parameter Ranges of BATCH SS-SOM

Parameters	**min**	**max**
Activation threshold (a_t)	0.90	0.999
Lowest cluster percentage (lp)	0.001	0.01
Relevance rate (β)	0.001	0.5
Max competitions (age_wins)	$1 \times S^*$	$100 \times S^*$
Winner learning rate (e_b)	0.001	0.2
Wrong winner learning rate (e_w)	$0.01 \times e_b$	$1 \times e_b$
Neighbors learning rate (e_n)	$0.002 \times e_b$	$1 \times e_b$
Relevance smoothness (s)	0.01	0.1
Connection threshold ($minwd$)	0	0.5
Number of epochs ($epochs$)	1	100

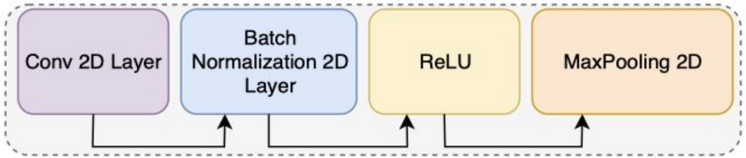
* S is the number of input patterns in the dataset.

Results - Unsupervised Learning [UCI Datasets]

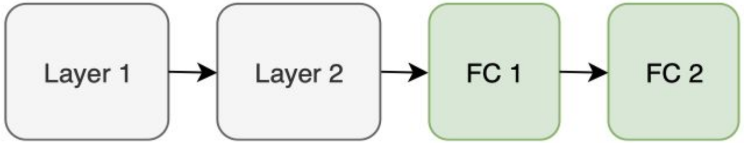
TABLE II: CE Results for Real-World Datasets. Best results for each dataset are shown in bold.

CE	Breast	Diabetes	Glass	Liver	Shape	Vowel
DOC [30]	**0.763**	0.654	0.439	0.580	0.419	0.142
PROCLUS [31]	0.702	0.647	0.528	0.565	0.706	0.253
LARFDSSOM [8] / SS-SOM [15]	**0.763**	**0.727**	**0.575**	0.580	0.719	0.317
ALT-SSSOM [17]	**0.763**	0.697	**0.575**	**0.603**	**0.738**	**0.319**
Batch SS-SOM	**0.763**	0.723	0.537	0.580	0.693	0.301

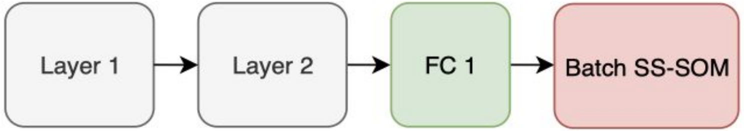
Experiments [MNIST]



(a) Custom layer block for MNIST.



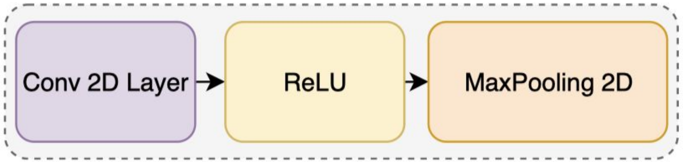
(b) MNIST CNN Model: Two custom layer blocks (Fig. 8a) followed by two fully-connected (dense) layers.



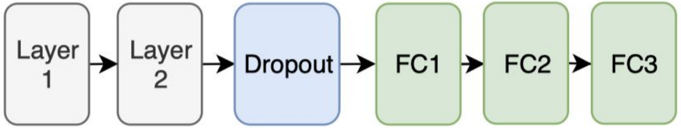
(c) BATCH SS-SOM training pipeline: The previous FC2 is removed, the features are extracted from FC1 and then fed to BATCH SS-SOM.

Fig. 8: MNIST Training Pipeline.

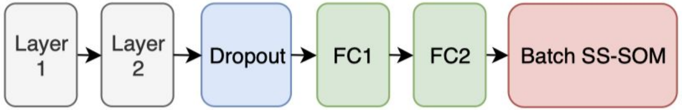
Experiments [Fashion-MNIST]



(a) Custom layer block for Fashion-MNIST.



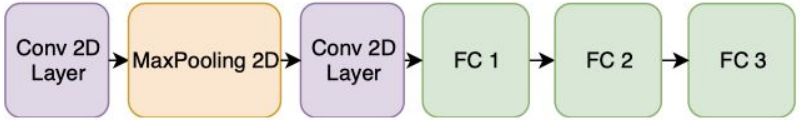
(b) Fashion-MNIST CNN Model: Two custom layers (Fig. 9a), and a dropout layer followed by 3 fully-connected (dense) layers.



(c) BATCH SS-SOM training pipeline: The previous FC3 is removed, the features are extracted from FC2 and then fed to BATCH SS-SOM.

Fig. 9: Fashion-MNIST Training Pipeline.

Experiments [SVHN]



(a) SVHN CNN Model: One convolutional 2D layer followed by a max-pooling 2D, other convolutional 2D layer and 3 fully-connected (dense) layers.

(b) BATCH SS-SOM training pipeline: The previous FC3 is removed, the features are extracted from FC2 and then fed to BATCH SS-SOM.

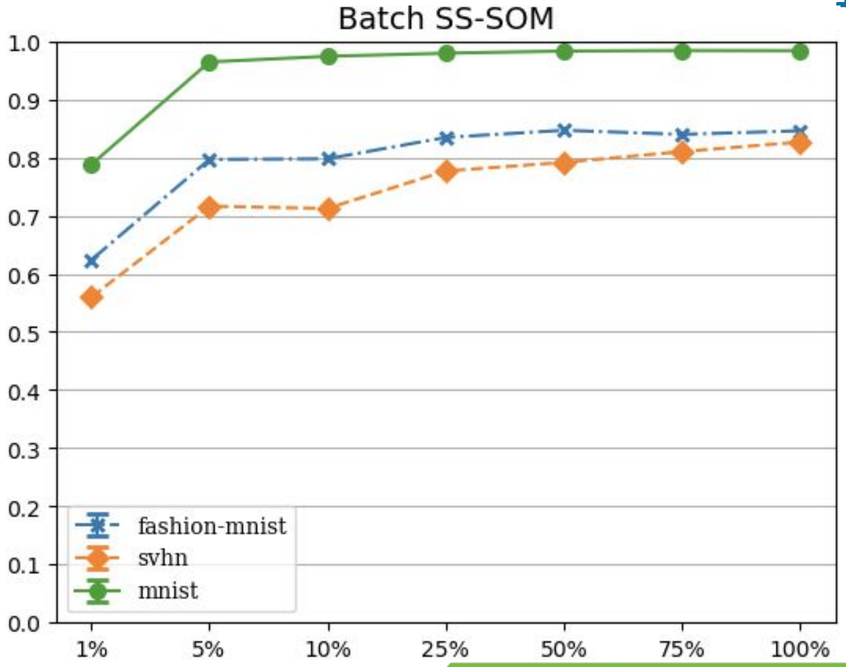
Fig. 10: SVHN Training Pipeline.

Results - Semi-Supervised Learning [MNIST, SVHN and Fashion-MNIST]

TABLE III: The Accuracy results obtained with BATCH SS-SOM on each dataset according to a percentage of labeled data.

%	MNIST	SVHN	Fashion-MNIST
1%	0.788	0.560	0.624
5%	0.9643	0.716	0.797
10%	0.974	0.713	0.798
25%	0.9793	0.777	0.834
50%	0.983	0.792	0.847
75%	0.9839	0.810	0.840
All	0.9836	0.826	0.846

Results - Semi-Supervised Learning [MNIST, SVHN and Fashion-MNIST]



Conclusion & Future Work

- BATCH SS-SOM: an approach that can be applied to both classification and clustering tasks in combination.
- Good performance in comparison with other traditional clustering models.
- Deals with more complex datasets and its representations.
- A promising path to follow: the first step towards more SOM-based models that can work effectively in non-traditional scenarios, such as image classification, transfer-learning, clustering latent representations, and so on.

IEEE WCCI 2020

IEEE World Congress on Computational Intelligence

Virtual Conference – July 19-24, 2020

Thank you!

SPONSORS:

IEEE Advancing Technology for Humanity

IEEE Computational Intelligence Society

IET

The International Neural Network Society (INNS)

EPS Evolutionary Programming Society

Thanks!

Any questions?

You can find us at:

- github.com/phbraga/batch-sssom
- @pedro_mbraga
- {**phmb4**, hrm, hfb}@cin.ufpe.br