The DBSCAN Clustering Algorithm on P Systems

György Vaszil

Faculty of Informatics

Sevilla, February 7, 2019

UNIVERSITY of DEBRECEN

Motivation

- **Internet of Things** (IoT) infrastructure for the creation of **Smart Cities**:
  - internet connected sensors,
  - devices,
  - citizens, …

→Enormous amount of “raw”data (“**Big Data**”), needs **interpretation**

- A technique: **data clustering**

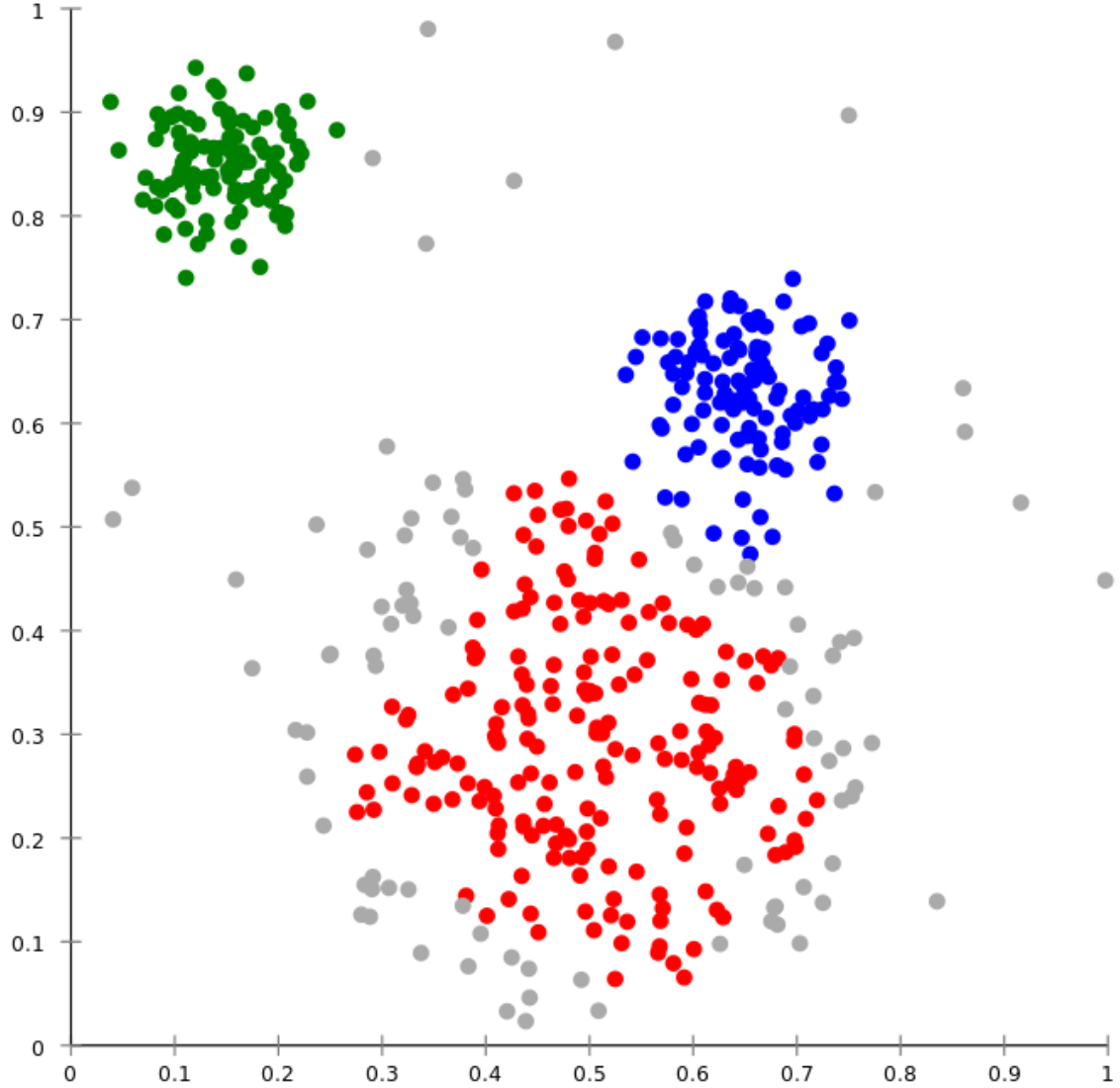
UNIVERSITY of DEBRECEN

Data clustering

- A sub-field of **data mining**
- The aim is to **partition** a data set into **clusters**, where
  - **intra-cluster** items are **similar**,
  - **inter-cluster** items are **dissimilar**
- To discover **implicit patterns** or **knowledge**

UNIVERSITY of DEBRECEN

- Data set – **points**
- A measure to define the “**distance**” of points
- Clusters – points that are **close** to each other



UNIVERSITY of DEBRECEN

Clustering with P systems, examples…

- Jie Xue, Xiyu Liu: A K-nearest Based Clustering Algorithm by P Systems with Active Membranes. JSW 2014 Vol.9(3): 716-725
- Xu J., Liu X., Xue J. (2014) Cluster Analysis by a Class of Splicing P Systems. In: Park J., Pan Y., Kim CS., Yang Y. (eds) Future Information Technology. Lecture Notes in Electrical Engineering, vol 309. Springer, Berlin, Heidelberg
- Yang Jiang, et al: A novel clustering algorithm based on P systems. International journal of innovative computing, information & control: IJICIC 10(2):753-765, 2014
- Zhao Y, Liu X, Li X. An improved DBSCAN algorithm based on cell-like P systems with promoters and inhibitors. PLoS One. 2018; 13(12):e0200751. Published 2018 Dec 17. doi:10.1371/journal.pone.0200751

UNIVERSITY of DEBRECEN

Clustering with P systems, examples…

- Jie Xue, Xiyu Liu: A K-nearest Based Clustering Algorithm by P Systems with Active Membranes. JSW 2014 Vol.9(3): 716-725
- Xu J., Liu X., Xue J. (2014) Cluster Analysis by a Class of Splicing P Systems. In: Park J., Pan Y., Kim CS., Yang Y. (eds) Future Information Technology. Lecture Notes in Electrical Engineering, vol 309. Springer, Berlin, Heidelberg
- Yang Jiang, et al: A novel clustering algorithm based on P systems. International journal of innovative computing, information & control: IJICIC 10(2):753-765, 2014
- **Zhao Y, Liu X, Li X. An improved DBSCAN algorithm based on cell-like P systems with promoters and inhibitors. PLoS One. 2018; 13(12):e0200751. Published 2018 Dec 17. doi:10.1371/journal.pone.0200751**

UNIVERSITY of DEBRECEN

The DBSCAN algorithm

Density **B**ased **S**patial **C**lustering of **A**pplications with **N**oise

- One of the **most common** clustering algorithms
- Groups together points with many **neraby neighbors**
- Marks points that **lie alone** in low density regions as "**noise**"

Ester, M; Kriegel, H-P; Sander, J; Xu, X (1996). Simoudis, E; Han, J; Fayyad, U M., eds. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231.

UNIVERSITY of DEBRECEN

The DBSCAN algorithm

Parameters:

- A **radius**, and
- the **minimal number of points** inside the radius required to form a dense region.

A point is **dense**, if it has enough close neighbors.

UNIVERSITY of DEBRECEN

The DBSCAN algorithm

1. Pick a **starting point** that is not yet visited
2. Mark it as core if it is **dense** (has enough close neighbors)
3. Create a **new cluster** for the point
4. Put all **neighboring points** in the cluster
   - If they are **dense**: -mark them **as core**
     -put them in the **same cluster**
     -go **back to 4**.
5. If the cluster **cannot be grown** any more, go **back to 1**.

UNIVERSITY of DEBRECEN

The membrane system for DBSCAN

- the points — $(i, \bar{p}_i)$ $\quad 1 \leq i \leq n$
- the result — $(i, \bar{p}_i)_j$

↖ the i-th point belongs to the j-th cluster

the cluster marked with j

UNIVERSITY of DEBRECEN

The membrane system for DBSCAN

$$\Pi = (V, [\,], w_1, R)$$

- $V = \{(i,\bar{p}_i), (c_i,\bar{p}_i), (c_i,\bar{p}_i)_j, (c_i,\bar{p}_i)_{j2}, (c_i,\bar{p}_i)_j^{core}, (c_i,\bar{p}_i)^{noncore}, (c_i,\bar{p}_i)_{j2}^{core}, (c_i,p_i)_j^{core}, (c_i,p_i)_{j2}^{nocor}, (c_i,\bar{p}_i)_j^{nocor} \mid 1 \le i,j \le n\}$
- $w_{s1} = A\,(c_1,\bar{p}_1)(2,\bar{p}_2)\ldots(n,\bar{p}_n)$
- promotes and priorities

UNIVERSITY of DEBRECEN

The membrane system for DBSCAN

1. pick a point:

$$A(i, \bar{p}_i) \rightarrow B(i, \bar{p}_i)'$$

2. Is it dense?

$$(k, \bar{p}_k) \rightarrow E_i(k, \bar{p}_k)i? \mid (i, \bar{p}_i)'' \quad |\bar{p}_k - \bar{p}_i| < \epsilon$$

We get as many E_i-s, as the number of neighbors of point i

UNIVERSITY of DEBRECEN

The membrane system for DBSCAN

1. pick a point:

$$A(i,\bar{p}_i) \rightarrow B(i,\bar{p}_i)'$$

2. Is it dense?

$$(\varepsilon,\bar{p}_\varepsilon) \rightarrow E_i(\varepsilon,\bar{p}_\varepsilon)_i^2 \;\Big|\; (i,\bar{p}_i)'' \quad |\bar{p}_\varepsilon - \bar{p}_i| < \epsilon$$

$$(i,\bar{p}_i)' \rightarrow (i,\bar{p}_i)_i^{core} \;\Big|\; \underbrace{E_i \cdots E_i}_{\text{min.}} > (i,\bar{p}_i)' \rightarrow (i,\bar{p}_i)^{non\ core}$$

UNIVERSITY of DEBRECEN

The membrane system for DBSCAN

3. Put the neighbors in the cluster

$$(i,|\bar{p}i)_{j?} \rightarrow (i,|\bar{p}i)j \mid (j,|p_j)^{core} > (i,|\bar{p}i)_{j?} \rightarrow (i,|\bar{p}i)$$

UNIVERSITY of DEBRECEN

The membrane system for DBSCAN

In addition: similar rules in several "versions"

and: $B \to C$, $C \to D$, $D \to A$

When A is present again, the process can start with a new point.

UNIVERSITY of DEBRECEN

Is it possible…

- … to increase the parallelism of the system?
- … to decrease the running time?

UNIVERSITY of DEBRECEN

Acknowledgments

Research supported in part

- by project no. **K 120558**, implemented with the support provided from the **National Research, Development and Innovation Fund** of Hungary, financed under the **K 16** funding scheme, and
- by the **EFOP-3.6.1-16-2016-00022** project, co-financed by the European Union and the European Social Fund.

UNIVERSITY of DEBRECEN