UDC 004.94(045) **doi: 10.20998/2078-7405.2021.94.02**

P. Veres, Miskolc, Hungary

THE IMPORTANTS OF CLUSTERING IN LOGISTIC SYSTEMS

Abstract. *Nowadays, the development of higher efficient processes and procedures is the key for success in industrial environment. The companies have machines, production lines, software and hardware tools with high level principles of efficient working. Example: the Industry 4.0 concept use the machines and methods of the near past, upgrade them, and gave them new purpose, as a more efficient tool. Some of the bases of those tools are not as efficient as which many would think, like in group generating or in other word clustering. Clustering is a very hard process, and it is in almost every decision making in every company's lives. It is important to sometimes examine its significance and flaws. This paper presents the clustering briefly and shows its errors through an example.*
Keywords: *Industry 4.0; clustering, heuristics; decision-making; routing.*

1. INTRODUCTION

Nowadays, everything is about working efficiently [1]. Globalization was the first step for giant logistic networks to emerge across the world [2]. Most of the times it is cheaper to produce or buy components or full products from the other side of the planet. Lots of multination companies realises this, and made connections. Our world is and was so interconnected, that if an error or a failure occured somewhere, it was hard to find, and even harder to fix. Our sensore technology, tracking devices and mobil computers help to eliminate the long seaching, but in most case they cant help in the recovery. This is where old and new technologies combines. The Industry 4.0 concept uses the existing technolgies and connects them into a system, that can do more then individually [3]. It creates new inforamtion from exist data and information. One of the most used princeple of the Industry 4.0 concept is machine decision making without human interaction in a situation, which has not yet occured for either the machine or a human. This makes the machine, the production or the whole system smart. This is a smart factory, and it contains some sort of Artificial Intelligence[4]. There are multiple IT companies, that want to create the most advanced AI, but in our situation, in the assembly or production industry, a basic algorithm that able to „learn“ with minimal computing power is enough [5]. One of this minimalistic AI is clustering algorithms, that can create groups based on measured or calculated parameters.

In almost every part of a any logistic system there are some clustering task. It used in route planning, package and unit load planing, location determination, warehouse allocation, high bay storage service, order and delivery management, and so more. It is a very important subtask, which is neglected by many manager

and engineer, because they said it is so „basic“, it’s implemented in everything perfectly. But most of these algorithms of the early ERP-s, are not capable of handeling effciently the increased databases[6].

In this paper I would like to show the importance of clustering with small examples, and show how much a clustering error affects an optimization process.

2. CLUSTERING ALGORITHMS IN GENEREAL AND LOGISTICS

Clustering is a NP-Hard problem in mathematics, which trnalates to: it has no ekzakt mathematical method to solve the problem. Therefor there are numerous clustering algorithm to which tries to produce a reasonable good solution in a short time. There are commonly used algorithms, such as: the most used K-mean and other K algorithms, nearest neighbor, Cobweb, CURE, Fuzzy cluster, BIRCH, DBSCAN, Human intuition, and so on [7]. Most of them must follow the same rules of their data tables and parameters. If two or more paramters share some convergence, that make the progress faster or more reliable, but this is not necesarry for the clustering algorithms. There are two important conditions for realible usage of these algorithms:

- Every individual object needs to have values in every parameter that we want to invole into the clustering process,
- If the parameters are not share the same unit of measure, it need to be weighted.

There are exceptions in both condition: Some advenced algorithms have AI to fill out blank parameters, if there are correlations in the dataset (big dataset is necessary); and if there are few parameters to consider some clustering process can be done with great success without weighting [8].

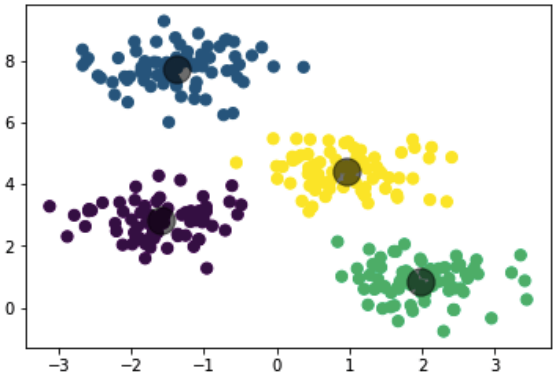

Figure1 – K-mean used in 2D clustering [9]

One of the most important and spectecular use of the algoritms in logistics is in route planning with multiple vehicles. In this situation to get a good solution, first you have to determine which vehicle visits which point and then optimise the route seperatly in the groups. Most of the clustering algorithms today use this method, which gives us a fairly good result but there is a chance, that it won‘t be the optimal soulution. To get the optimal solution with a greater chance the two optimization process (clustering and route planning) have to work paralell, because both have influence over on the other. This is where Heuristics can help. Heuristic algorithms handles all parameters at once, where grouping is just another parameter. Besides that, no other parameter has to change, and if it’s not necessary, it doesn’t need weighting the parameters. However, an evaluation function, like the fitness function in evolutionary algorithms, is necessary for this method. In my personal studies I use both methods and I recommend using them in this division:

- Use the Clustering algoritmhs, if there are just a few parameters (under 10) and only need mild weighting.
- Use Heuristics, if there are big databases, with high number of parameters.

3. CLUSTERING EXAMPLE IN MULTI-VEHICLE ROUTING

The multi-vehicle routing problem is a combination of at least two problem, but most of the time there are four basic logistics problem, which plays a role in the difficulty [10]:

- routing problem
- clustering problem
- capacity problem
- location determination problem

In this paper the first two has already been expalined. The capacity problem is a real life problem of transporting. beacuse, the capacity of the vehicles and the locations are predetermined, and can‘t be violated. The capacity acts like a limit, that can determine the quantity of the vehicles, storages, docks, etc. in a company. The location determination problem creates „floating objects“, until the optimization problem rootes them and gives them their own parameters. It can be a parking slot or a charging station for our vehicles. Also it can be a propective warehouse in the future or any object, that will be part of the route, but the optimization process should decide where it needs to be. These are the four fundamental problems of the multi-vehicle routing problem. In the next example besides, the location determination problem, all other has a role.

In the example there are 20 objects in a circle in the map giving the same volume and weight of wares. In the middle of the circle, there will be the company that collects the wares. We like to get the least amount of driven lenght, with the data in Table 1. This is a highly specific data, because we know what need to come

out. With this setup, there will be 3 routes, and the perfect solution should look like a radiation warning sign, which can be seen in figure 2. The solution is created by a Heuristical method called Evolutionary alhorithm, and the total lenght of the route, that the 3 vehicles have to take is 965,85. It has no dimensions, because we don‘t define it in the coordinates.

Table 1 – Coordinates and basic data of the artficial problem

#	X	Y	Weight	Volume
Center	0	0		
1	-60,079	52,82534	2	2
2	-79,3423	-10,2371	2	2
3	-44,7031	-66,3448	2	2
4	20,30587	-77,38	2	2
5	71,51973	-35,8459	2	2
6	74,14548	30,04077	2	2
7	26,39927	75,51873	2	2
8	-39,2817	69,69179	2	2
9	-78,276	16,51858	2	2
10	-64,0922	-47,8768	2	2
11	-6,36629	-79,7463	2	2
12	55,68468	-57,4388	2	2
13	79,90534	3,89068	2	2
14	49,84098	62,57697	2	2
15	-14,0837	78,75056	2	2
16	-68,4403	41,42365	2	2
17	-76,3009	-24,0451	2	2
18	-32,3252	-73,1784	2	2
19	33,61117	-72,5968	2	2
20	76,71326	-22,6953	2	2
SUM			40	40

			Max Weight	Max Volume
Routes:	**3**		**27**	**15**

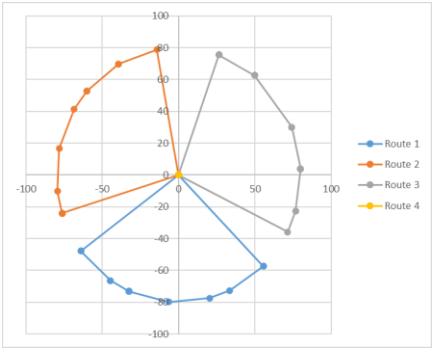

Figure 2 – Perfect solution of artificial routing problem

To get the perfect solution, there were advanced programing, fine tuning and multiple re-runs of the optimization method. This is not acceptable in an industrial environment. Most of the time the resoult was a semi-perfect circle or 1-2 object in the wrong group as shown in figure 3-5.

Figure 3 shows a routing optimization problem, where the solution in 1216,85. It is 26% higher than the optimal.

In figure 4. there are a mild clutering solution with a value 1072,49. It is 9% higher than, the optimal.

In figure 5 we can see a mild clustering and route optimization error, where thesolution is 1238,53, which is 28% higher, the the perfect score.

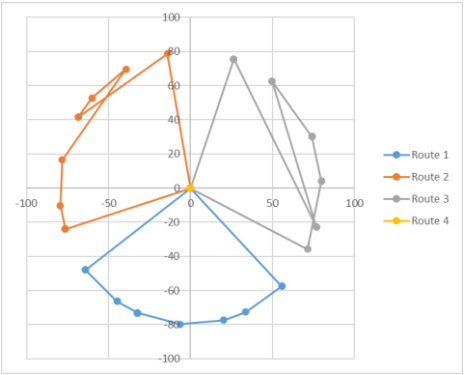

Figure 3 – Problem with routing optimization in the solution of artificial routing problem

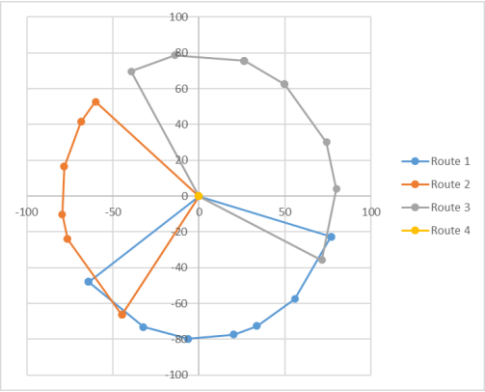

Figure 4 – Problem with clustering in the solution
of artificial routing problem

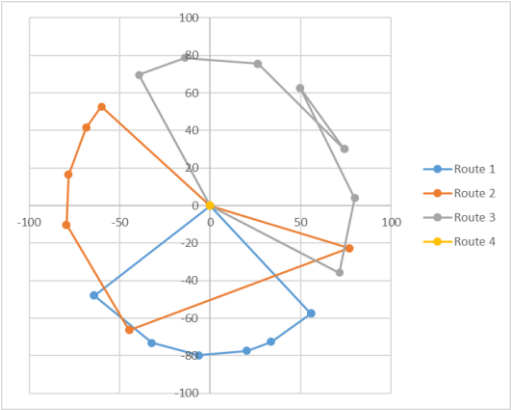

Figure 5 – Problem with routing and clustering optimization
in the solution of artificial routing problem

4. SUMMARY

Clustering algorithms and other methods to create groups and gather more information of a database is necessary in the industry today to be as efficient as the competitors. If you can get information or save money with little or no efforts, that is a win for everybody. Nowadays a few percent increase or decrease in efficiency

can mean that a company can turn in profit or not. The example shows us that a little error in clustering and/or a connecting problem can decrease efficiency with more than 10%, which is a huge loss. This is a very important reason to look after which tools, we use in our calculations, software, and databases.

ACKNOWLEDGEMENT

The research was financially supported by the Higher Education Excellence Program (FIKP) of the Ministry of Human Capacities in the framework of the "Optimization of natural resources Research on the efficiency improvement procedures and methods of logistics systems based on modern technologies" research topic in logistics system design of the University of Miskolc.

References: 1. *Bányai, T.:* Introductory Chapter: Industry 4.0 and Its Impact on Logistics - A Retrospective Review, INDUSTRY 4.0 - IMPACT ON INTELLIGENT LOGISTICS AND MANUFACTURING, pp 1-9, 2020; https://doi.org/10.5772%2Fintechopen.89387 **2.** *Dobos, P., Tamás, P., Illés, B., Balogh, R.:* Application possibilities of the Big Data concept in Industry 4.0, IOP CONFERENCE SERIES: MATERIALS SCIENCE AND ENGINEERING 448: 1 Paper: 012011, 2018; **3.** *Illés, B., Tamás, P., Dobos, P., Skapinyecz, R.:* New challenges for quality assurance of manufacturing processes in industry 4.0, Solid State Phenomena. Vol. 261. Trans Tech Publications Ltd, 2017; https://doi.org/10.4028/www.scientific.net/SSP.261.481 **4.** *Borodavko, B., Illés, B., Bányai, Á.* Role of artificial intelligence in supply chain, ACADEMIC JOURNAL OF MANUFACTURING ENGINEERING 19: 1 pp. 75-79. , 2021; **5.** *Timothy, R.*: Google's DeepMind AI discovers physics, NEW SCIENTIST, Volume 232, Issue 3100, Page 25, 2016; https://doi.org/10.1016/S0262-4079(16)32121-2 **6.** *Saraswati J., Brian F., Rafiq A.:* Lean OR ERP – A Decision Support System to Satisfy Business Objectives, PROCEDIA CIRP Volume 70, Pages 422-427, 2018, https://doi.org/10.1016/j.procir.2018.02.048 **7.** *Bryar A.H., Tarik A. R., Seyedali M.:* Performance evaluation results of evolutionary clustering algorithm star for clustering heterogeneous datasets, DATA IN BRIEF Volume 36, 2021, p. 107044, 2021, **8.** *Zun-yang L., Feng D., YingX.X.H.:* Background dominant colors extraction method based on color image quick fuzzy c-means clustering algorithm, DEFENCE TECHNOLOGY 2020. https://doi.org/10.1016/j.dt.2020.10.002 **9.** *Muthukrishnan*: Mathematics behind K-Mean Clustering algorithm ARTIFICIAL INTELLIGENCE DATA SCIENCE, 2018 **10.** *Tamás P., Illés B., Bányai T., Bányainé T. Á., Umetaliev A., Skapinyecz R.*: Intensifying Cross-border Logistics Collaboration Opportunities Using a Virtual Logistics Center, JOURNAL OF ENGINEERING RESEARCH AND REPORTS 13, Paper :3, 2020.

Петер Вереш, Мішкольц, Угорщина

ВАЖЛИВІСТЬ КЛАСТЕРИЗАЦІЇ У ЛОГІСТИЧНИХ СИСТЕМАХ

Анотація. *Концепція Індустрії 4.0 використовує існуючі технології і об'єднує їх в систему, яка може робити більше, ніж індивідуально. Вона створює нову інформацію з існуючих даних.*

Одним з найбільш часто використовуваних принципів концепції Індустрії 4.0 є прийняття рішень машиною без втручання людини в ситуації, яка ще не відбулася ні для машини, ні для людини. Це робить машину, виробництво або всю систему розумними. Це розумна фабрика, і в ній є свого роду штучний інтелект (ШІ). Є кілька IT-компаній, які хочуть створити найбільш просунутий ШІ, але в даній ситуації, в сфері збирання або виробництва, досить простого алгоритму, який може «вчитися» з мінімальною обчислювальною потужністю. Один з цих мінімалістичний II - алгоритми кластеризації, які можуть створювати групи на основі виміряних або розрахованих параметрів. Кластеризація - це NP-важке завдання в математиці, яка зводиться до того, що у неї немає спеціального математичного методу для її вирішення . Тому існує безліч алгоритмів кластеризації, які намагаються в короткі терміни дати розумне гарне рішення. Одним з найбільш важливих і специфічних способів використання алгоритмів в логістиці є планування маршруту з декількома транспортними засобами. У цій ситуації, щоб отримати гарне рішення, спочатку визначається, який транспортний засіб відвідує яку точку, а потім оптимізується маршрут окремо по групах.Більшість алгоритмів кластеризації сьогодні використовують цей метод, який дає досить хороший результат, але є шанс, що він не буде оптимальним рішенням. Щоб отримати оптимальне рішення з більшою ймовірністю, два процеси оптимізації (кластеризація і планування маршруту) повинні працювати паралельно, тому що обидва мають вплив один на одного. Тут може допомогти евристика. Евристичні алгоритми обробляють всі параметри відразу, а угруповання - це просто ще один параметр. Крім того, ніякі інші параметри не повинні змінюватися, і, якщо в цьому немає необхідності, не потрібно зважувати параметри. Однак для цього методу необхідна функція оцінки, така як функція пристосованості в еволюційних алгоритмах. В даний час збільшення або зменшення ефективності на кілька відсотків може означати, що компанія може отримувати прибуток чи ні. Невелика помилка в кластеризації і / або проблема з підключенням можуть знизити ефективність більш ніж на 10%, що є величезною втратою. Це дуже важлива причина, щоб визначити, які інструменти повинні використовуватися в наших розрахунках, програмному забезпеченні і базах даних.

Ключові слова: *Індустрія 4.0; кластеризація; евристика; прийняття рішень; маршрутизація.*