Issue 7, 2020

https://doi.org/10.23939/sisn2020.07.078

УДК 004.4

AN INFORMATION SYSTEM FOR COMPARING PARTNER EMPLOYMENT SCHEDULES

[1]Bohdan Pliasun, [2]Nataliia Kunanets, [3]Nataliia Veretennikova

[1, 2, 3] Lviv Polytechnic National University
[1]bohdan.pliasun.kn.2016@lpnu.ua, ORCID 0000-0002-6912-5982
[2]nek.lviv@gmail.com, ORCID 0000-0003-3007-2462
[3]nataver19@gmail.com, ORCID 0000-0001-9564-4084

There are various approaches to describing, comparing and converting data nowadays, but there are challenges that require new methods. This article describes the combination of some cluster comparison approaches for the flexibility of design systems. This approach can reduce the impact of missing data when comparing clusters. This method is used in the system of comparison of partner employment schedules, where there is no constant flow of data from the user and the data itself has big differences. In most cases, this data is incomplete, which complicates the cluster comparison process by simple methods. Combining optimistic comparison and realistic methods, you can change the coefficient of similarity for each individual pair of users (a realistic comparison is used for users with a close data set, and an optimistic comparison is for a different set).

The article also describes the possibility of converting point data over a long period (using time) by gluing a factor. This allows to play a long-lasting event only in a few moments. Also, this conversion allows to work with one element at a specific time in the same way as with a time interval.

Thus, using the described approaches reduces the impact of lack of data from users, as well as conditional specific values can be converted into segments close to reality, which allows the projected system to become more flexible.

Key words: cluster, partner employment schedules.

Introduction

In today's dangerous society, there are situations when it is necessary to determine the location of a certain person. Modern information technologies make it easy to carry out such search. These technologies are very convenient for parents who find it necessary to have information about the location of their children and to ensure the safety of the elderly. Such information is provided by almost all mobile operators.

These technologies are used to establish effective communications with partners, to agree working schedules, based on technologies of location tracking. The analysis of the obtained data helps to determine convenient locations for business meetings, briefings and other events.

State of the problem research

Geo-location is the process of binding, editing, and using metadata that describes the geographical location of a location [1].

Geolocation technologies are used in marketing activities [2, 3] and help to increase sales. Modern urban transport systems are provided with geolocation systems. This approach is generated by the need to determine the location of vehicles at a certain point in time [4], when transporting important goods [5]. Geolocation technologies are actively used in other sectors of the economy.

The aim of the paper is to analyze the functionality of an information system that ensures the coordination of working meetings with the use of geopositioning technologies.

Data preparation

This system is designed to compare the employment of individuals over a period of time, to take into account their common interests, as well as the possibility of cooperation or other aspects of communication.

To begin with, it is necessary to generate employment data, so a module for processing user location data using Google Maps was created. Receiving GSP data from a user were processed as follows:

– Search near the user's location data of historical places or institutions (cafes, pools, entertainment centers, hospitals, etc.).

– Google Maps identifies the type of location. Some types of Google Maps locations can be combined, such as hospital and sanatorium can be joined into healthcare or simply health. These combinations can be done in any way.

As a result, it is obtained data such as the type of location and time of navigation (the transmission time of location data). This is followed by a time merging. Time merging is the unification of single data (with a specific time) whose parameters are close to each other (i.e., where the time difference is less or equal to the time T and is set by the user) into a time group. The merging contains several steps:

– An array of data (location type and time) of the objects is received at the input.

– This array searches for nearby items (an order specifies a time attribute) that have the same place type.

– If the time difference between the found objects is less or equal to T, they are combined into a complete conditional object that has time attributes of beginning and end.

– Merging is performed until any difference in the time attributes of next objects is larger than T.

– Objects that remain outside the specified time (t) are transferred to a new object with a start of t – T/2 and end of t + T/2.

Following the above actions, it is obtained grouped data, which can be presented graphically (Fig. 1).

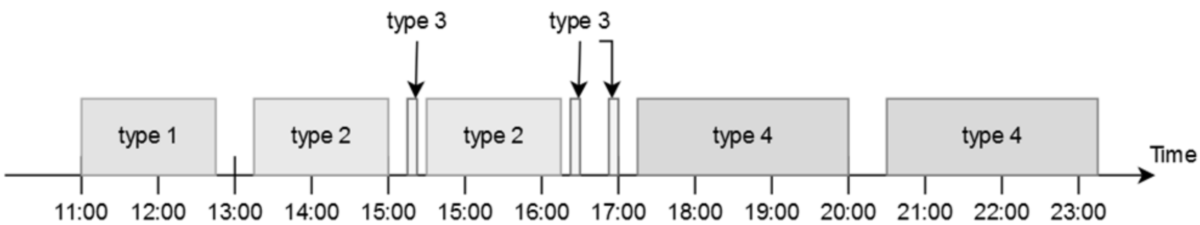

Fig. 1. User data

Cluster

Clusterization is performed on certain objects called clusters. The cluster model is written in C # presented below.

```
public class Cluster
{
        // for each interval there is a dataset of probabilities
        public Dictionary<DateTime, CountingType> TimeDatasets { get;
        set; } = new Dictionary<DateTime, CountingType>();
```

```
        public int CountWeeks { get; set; } = 0;
        public int EveryMinutes { get; set; }
}
public class CountingType
{
        public double GeneralCount { get; set; } = 0;
        public Dictionary<int, double> TypeFrequency { get; set; } =
             new Dictionary<int, double>();
}
```

In `Cluster` Class, there are attributes such as `EveryMinutes` – the number of minutes that will define a single time span along the time axis, `CountWeeks` – the number of weeks that are entered into the cluster (determine the cluster value), `TimeDatasets` – the set of key-value pairs, where the key is a single unit of time, and the value is a `CountingType` Class.

`CountingType` Class displays the set of occurrence probabilities for each place type. This class contains the following attributes as `GeneralCount` defines the amount of entered data; `TypeFrequency` is a set of key-value pairs, where the key is the location type and the value is the time parameters for the object.

Setting data to the cluster

For our system, the number of minutes that will determine a single time interval was chosen equal to 60 (minutes).

In order to be able to cluster the data, they are divided into weeks, since it is mainly the days of the week that form a human schedule. A cluster is formed every week in which the fields `EveryMinutes`, `CountWeeks` are filled with 60 and 1 respectively. The `CountingType` attribute is formed as follows:

- Items from the array of grouped data (full objects) are selected in turn.
- The appropriate time intervals of `TimeDatasets` attribute are searched for the selected array element by time (day of the week and time of a day).
- The `GeneralCount` and `TypeFrequency` fields are filled in for each period as follows:
  - If the time period (field) is entirely the time interval of the array element, the `TypeFrequency` field records the location type with value 1.
  - If the time period (field) belongs to the time interval of an element from an array only partially, then the `TypeFrequency` field records the type of place with a value equal to the fraction of interval intersection. For example, the element period of an array is 13:40-15:00, and the selected field time is 13:00-14:00, then the value for a time period of the array will be entered with a value of 0.33.
  - In both cases, the `GeneralCount` field will be initialized with the number 1.

The result is a set of clusters that correspond to the appropriate weeks. As an example, three individuals and data of four weeks were selected. Following the above operations, it was obtained data shown in Fig. 2–13.

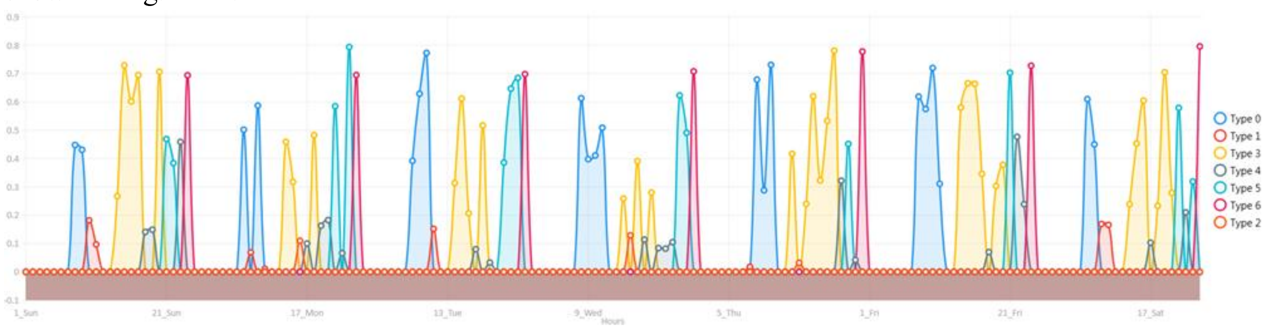

Fig. 2. User Cluster 1, Week 1

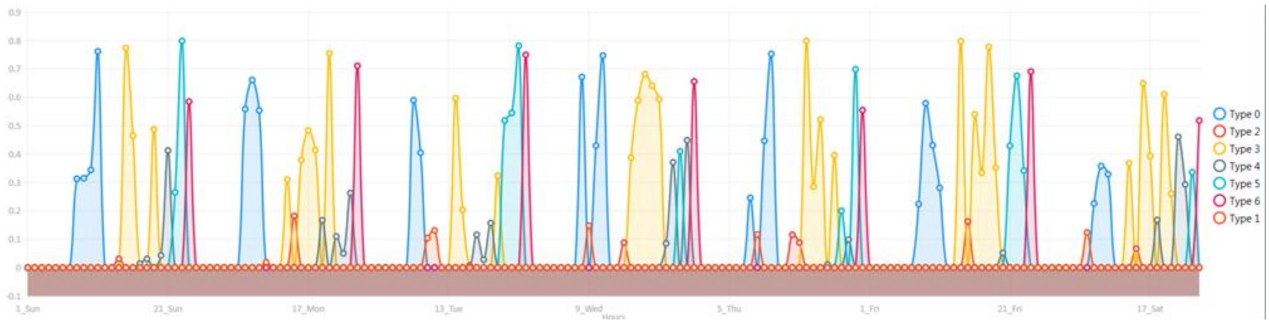

Fig. 3. User Cluster 1, Week 2

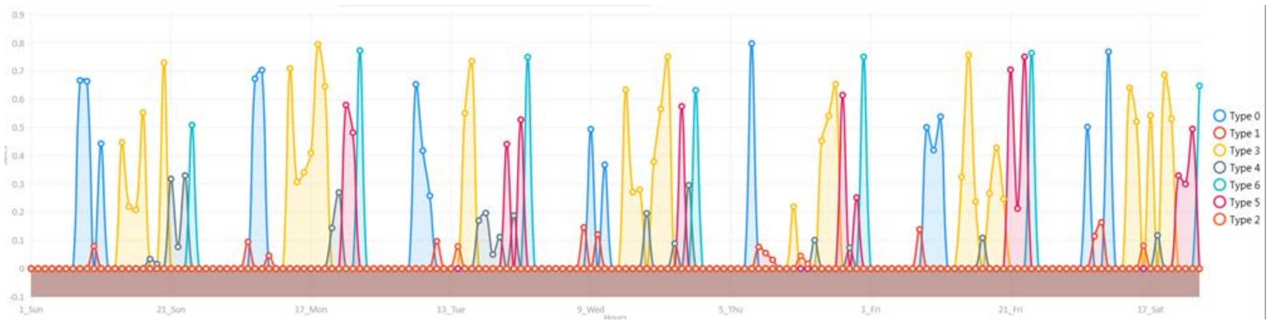

Fig. 4. User Cluster 1, Week 3

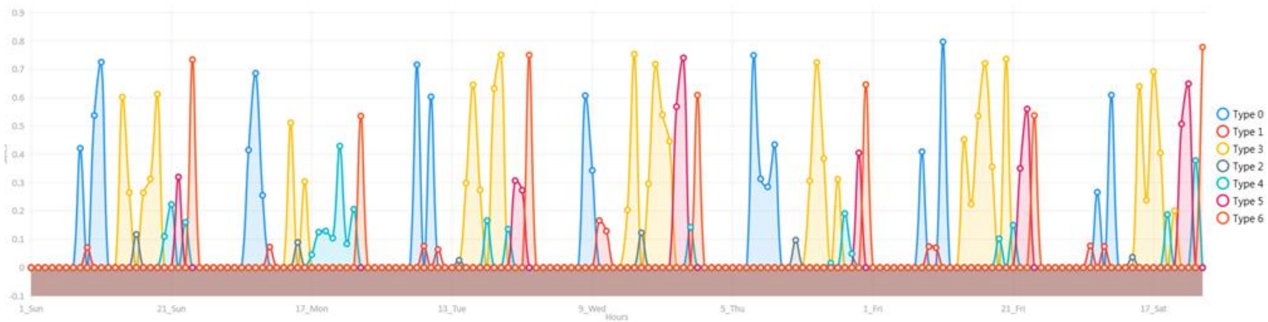

Fig. 5. User Cluster 1, Week 4

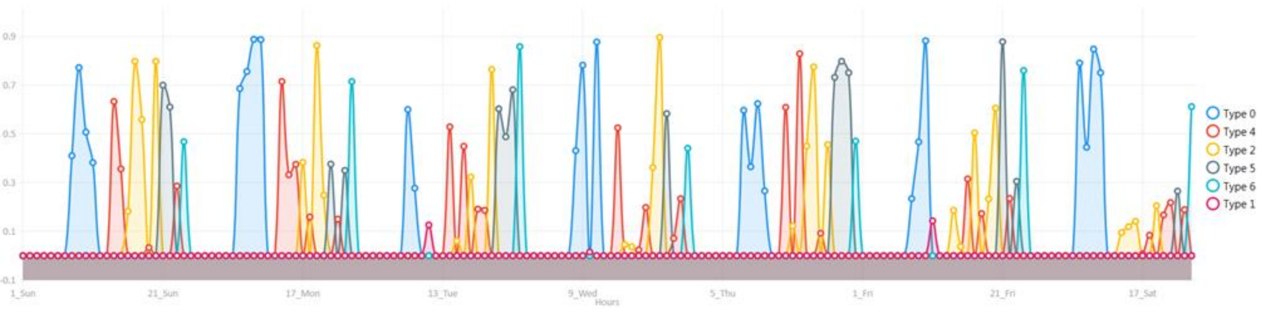

Fig. 6. User Cluster 2, Week 1

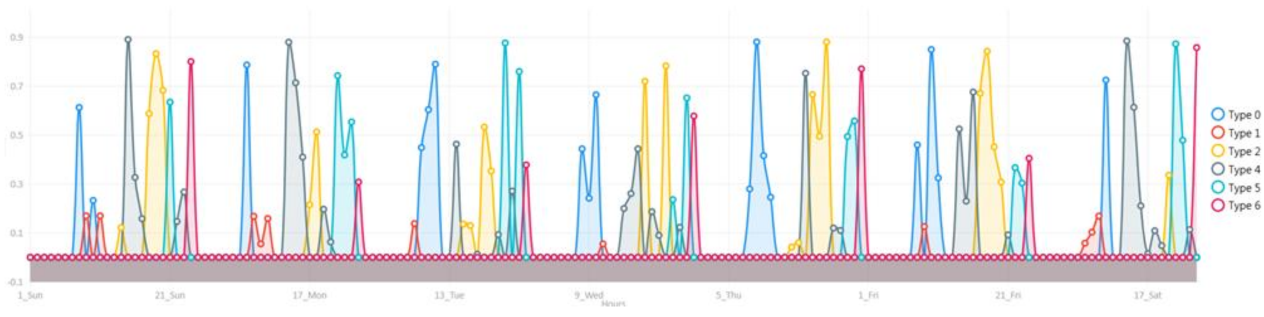

Fig. 7. User Cluster 2, Week 2

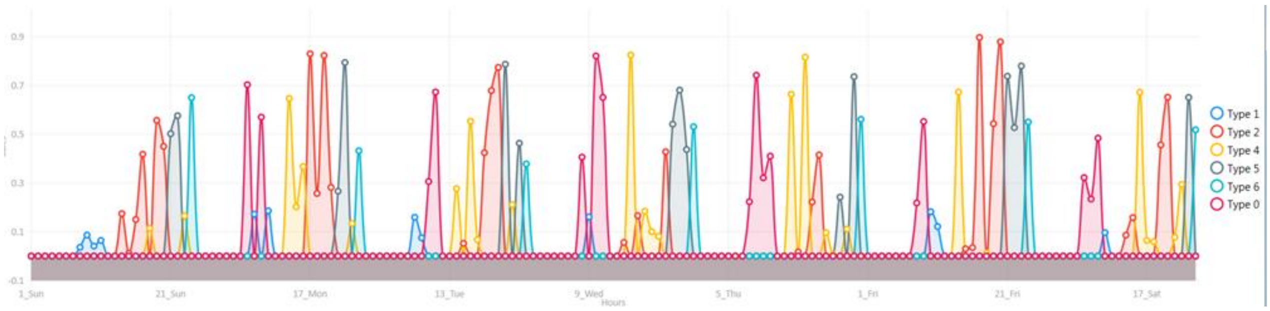

Fig. 8. User Cluster 2, Week 3

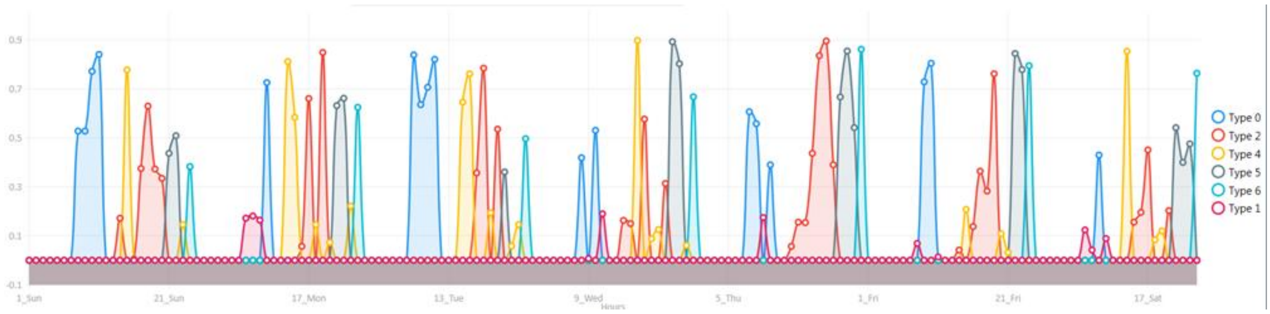

Fig. 9. User Cluster 2, Week 4

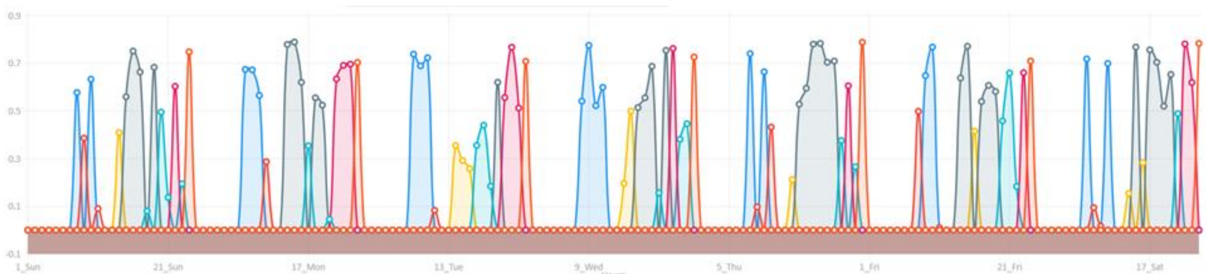

Fig. 10. User Cluster 3, Week 1

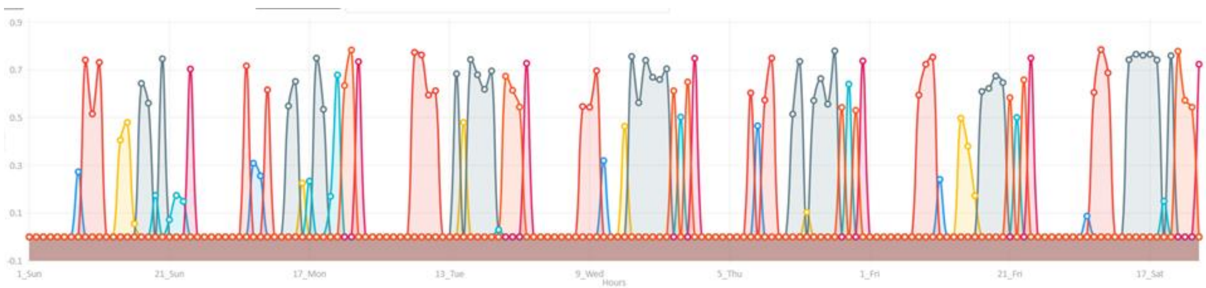

Рис. 11. User Cluster 3, Week 2

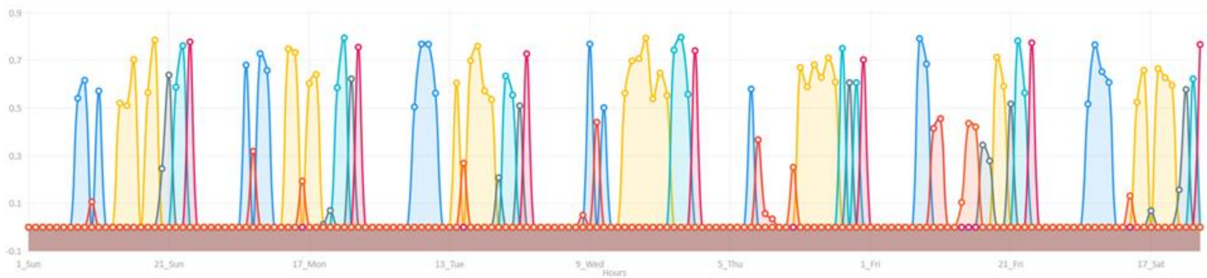

Fig. 12. User Cluster 3, Week 3

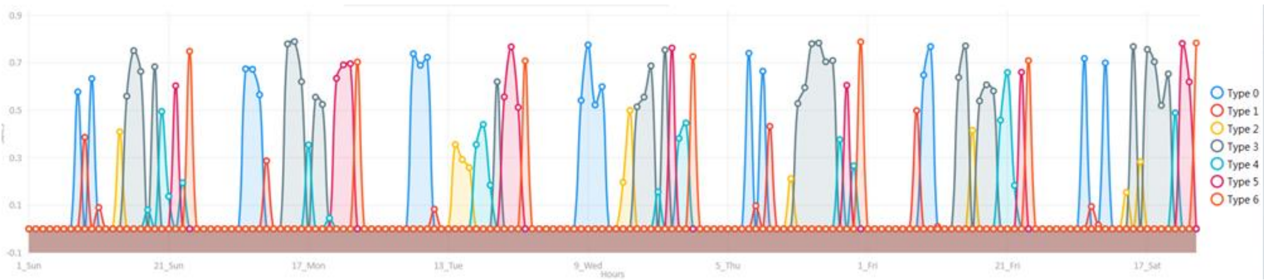

Fig. 13. User Cluster 3, Week 4

Cluster comparison and integration

Comparison of clusters gives an opportunity to get percent similarity and difference between certain clusters. The comparison is based on the principle of similarity of relatively smaller, difference of relatively larger. The functioning of the method is shown in Figures 14a and 14b.

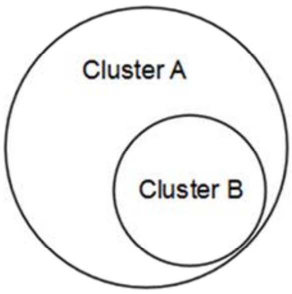

Fig. 14a. Absorption

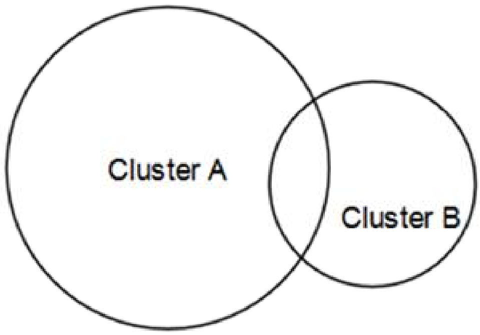

Fig. 14b. Intersection

If the smaller cluster is completely absorbed by the larger one (Fig. 14a), then the SMALLER cluster is similar to the LARGER one by 100 %, but the LARGER cluster is NOT similar to the SMALLER one by 100 %! If you need to find out the difference between clusters, it should be calculated concerning the larger one. When there is a cluster intersection (Fig. 14b), the coefficients are calculated similarly.

This approach is an optimistic solution to the problem of lack of user data, so this method avoids the growing difference between full and incomplete clusters. If there is a need for accurate data, then the difference coefficient is taken into account when comparing clusters.

The result of the comparisons is a similarity matrix, which determines the coefficient similarity between the established clusters.

The above clusters are compared and shown in Fig. 15–17.

```
--------------------_
1 -- { 0, 0,218638095238095, 0,230380952380952, 0,183114285714286 }
2 -- { 0,218638095238095, 0, 0,206438095238095, 0,202571428571429 }
3 -- { 0,230380952380952, 0,206438095238095, 0, 0,199238095238095 }
4 -- { 0,183114285714286, 0,202571428571429, 0,199238095238095, 0 }
--------------------_
```

Fig. 15. The similarity matrix of clusters of user 1

```
--------------------_
1 -- { 0, 0,205419047619048, 0,225314285714286, 0,231752380952381 }
2 -- { 0,205419047619048, 0, 0,209333333333333, 0,235285714285714 }
3 -- { 0,225314285714286, 0,209333333333333, 0, 0,206895238095238 }
4 -- { 0,231752380952381, 0,235285714285714, 0,206895238095238, 0 }
--------------------_
```

Fig. 16. The similarity matrix of clusters of user 2

```
--------------------_
1 -- { 0, 0,321371428571429, 0,307485714285714, 0,317295238095238 }
2 -- { 0,321371428571429, 0, 0,364285714285714, 0,363542857142857 }
3 -- { 0,307485714285714, 0,364285714285714, 0, 0,354628571428571 }
4 -- { 0,317295238095238, 0,363542857142857, 0,354628571428571, 0 }
--------------------_
```

Fig. 17. The similarity matrix of clusters of user 3

As shown in Fig. 15–17, the similarity between clusters varies within 15–35 %, this indicates some inaccuracies in the data, which is often one of the disadvantages of building clusters with a small set of data (in this case, one cluster has 1 week). Therefore, when it is created the cluster for merging firstly, a similarity coefficient is 0.10. The results of merging are shown in Fig. 18–20.

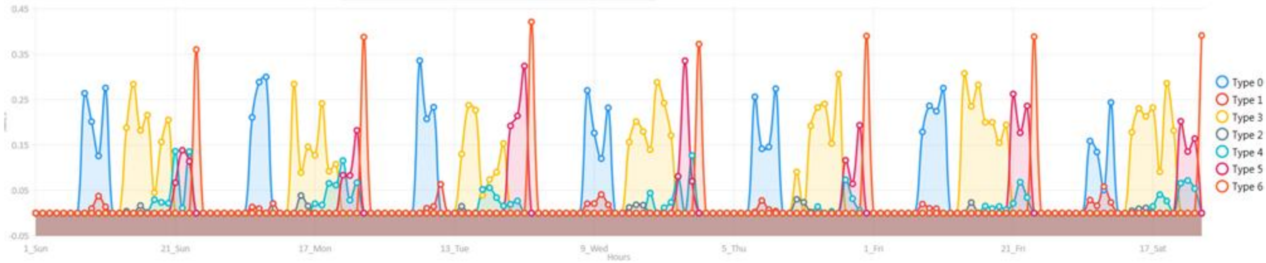

Fig. 18. Four weeks cluster of user 1

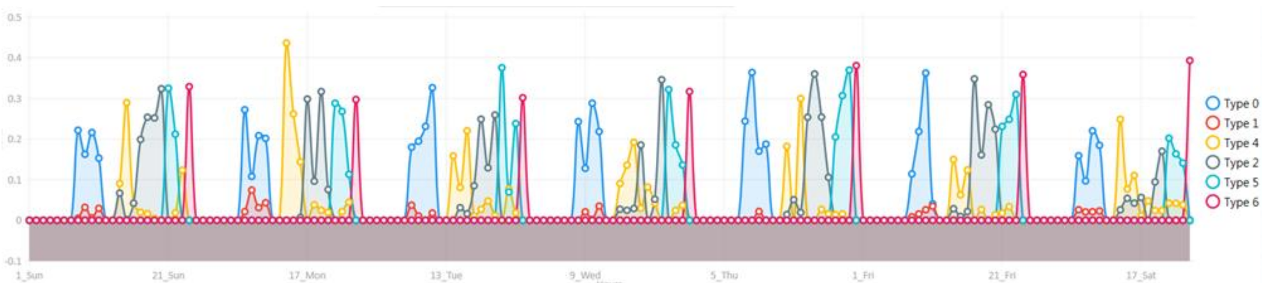

Fig. 19. Four weeks cluster of user 2

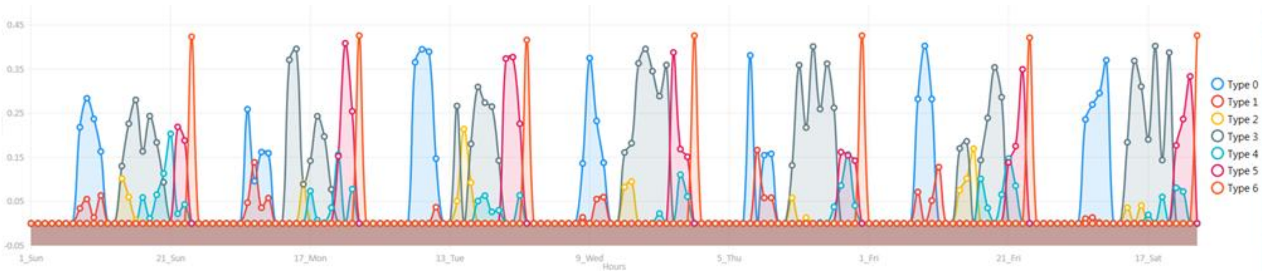

Fig. 20. Four weeks cluster of user 3

The clusters in Fig. 18–20 contain a lot of data that has a low frequency of occurrence. This is data that needs to be cleaned up after a cluster with lots of data has been created. Purification occurs by different methods and coefficients, so the purification method can vary depending on the task. The most common method is to clean up data that is 10 times less frequent (relative to the maximum value in the same period).

After developing user schedules (clusters that contain several weeks), they can be compared with each other. Using the user data, they are compared, as shown in Fig. 21.

```
--------------------_
1 -- { 0, 0,101937755102041, 0,180501870748299 }
2 -- { 0,101937755102041, 0, 0,114083333333333 }
3 -- { 0,180501870748299, 0,114083333333333, 0 }
--------------------_
```

Fig. 21. Comparison matrix of user schedules

Fig. 21 shows that schedules of users 1 and 3 are the most similar, so it is easier for them to schedule meetings.

Conclusions

Nowadays, due to the large amount of data, it is possible to predict events based on the obtained results. These data need to be converted to the required formats to allow them to be further used in calculations and in forecasting. Different types of bonding can be applied when clustering together, which is often described in Cluster Analysis. One of these methods is "nearest neighbor" and "far neighbor". They are all used depending on the needs of the computing industry. Not only Euclidean distance, but square Euclidean distance, Chebyshov distance, power distance, Manhattan distance are used to determine the distance between objects. In this paper, the number difference was used to determine the distance as the main numeric object was only time. Also, when comparing created objects, it was used own comparison method (in terms of frequency and availability), which is described above. This method allows you to choose an optimistic and realistic object comparison approach, which allows each individual user to use the desired method in the future. This method separates the concepts of similarity and difference depending on the object itself. In the described approach, the concept of difference is defined by a relatively larger object (when comparing two different objects in terms of the number of data) and similarities by a relatively smaller one, so the similarity coefficient demonstrated by this method is an optimistic approach to solving the problem, and to obtain realistic data, it is necessary to subtract the unit from the difference coefficient. Using this method, you can maneuver at intervals when there is not enough data to compare clusters.

Список літератури

1. Геолокація. Режим доступу: https://docs.kde.org/trunk5/uk/extragear-graphics/digikam/tool-geolocation.html
2. Гринів Н. Т., Гринів Т. Т. (2012). Геолокаційні сервіси як маркетинговий інструмент. Режим доступу: http://ena.lp.edu.ua:8080/bitstream/ntb/16922/1/56-Hryniv-93-94.pdf
3. Біленко В. О. (2015). Геоконтекстний маркетинг: діалог споживача та рекламодавця. *Економічні проблеми сталого розвитку : матеріали Міжнародної науково-практичної конференції імені проф. Балацького О. Ф.*, 264–265.
4. Delbosc A. (2011). Modelling the causes and impacts of personal safety perceptions on public transport ridership. *Transport Policy,* № 24, 302–309.
5. Онищук В. П. (2014). Передові технології і засоби у перевезенні небезпечних вантажів. *Наукові нотатки*, Випуск № 45, 407–411.

References

1. Heolokatsiia. Access mode: https://docs.kde.org/trunk5/uk/extragear-graphics/digikam/tool-geolocation.html.
2. Hryniv N. T., Hryniv T. T. (2012). Geolocation services as a marketing tool. Access mode: http://ena.lp.edu.ua:8080/bitstream/ntb/16922/1/56-Hryniv-93-94.pdf.
3. Bilenko V. O. (2015). Geo-contextual marketing: a dialogue between the consumer and the advertiser. *Economic problems of sustainable development: Proceedings of the International Scientific and Practical Conference named after prof. Balatsky O. F.*, 264–265.
4. Delbosc A. (2011). Modelling the causes and impacts of personal safety perceptions on public transport ridership. *Transport Policy,* No. 24, 302–309.
5. Onyshchuk V. P. (2014). Advanced technologies and facilities in the transportation of dangerous goods. *Scientific Notes*, Issue No. 45, 407–411.

ІНФОРМАЦІЙНА СИСТЕМА ПОРІВНЯННЯ РОЗКЛАДІВ ЗАЙНЯТОСТІ ПАРТНЕРІВ

Богдан Плясун[1], Наталія Кунанець[2], Наталія Веретеннікова[3]
[1, 2, 3]Національний університет “Львівська політехніка”,
[1]bohdan.pliasun.kn.2016@lpnu.ua, ORCID 0000-0002-6912-5982
[2]nek.lviv@gmail.com, ORCID 0000-0003-3007-2462
[3]nataver19@gmail.com, ORCID 0000-0001-9564-4084

Існують найрізноманітніші підходи до опису, порівняння та конвертування даних, але з часом виникають задачі, які вимагають застосування нових методів. Описано комбінування деяких підходів порівняння кластерів задля гнучкості проектованої системи. Завдяки цьому підходу можна знизити вплив нестачі даних при порівнянні кластерів. Метод застосовують у системі порівняння розкладів зайнятості осіб, де немає постійного потоку даних від користувача та самі дані значно відрізняються. Також ці дані в більшості випадків є неповними, що ускладнює процес порівняння кластерів простими методами. Комбінуючи методи оптимістичного та реалістичного порівняння, можна до кожних окремих пар користувачів змінювати коефіцієнт схожості (для користувачів з близьким набором даних використовувати реалістичне порівняння, а для різного набору — оптимістичне порівняння).

Також описано можливість перетворення точкових даних на протяжні (за часом), використовуючи склейку за певним коефіцієнтом. Це дозволяє відтворити протяжну подію лише за деякими її моментами. Також подібне перетворення дозволяє працювати з одним елементом у конкретний момент часу так само, як і з часовим проміжком.

Отже, з використанням описаних підходів знижується вплив нестачі даних від користувачів; також умовні конкретні значення можна перетворювати на відрізки, близькі до реальності, що робить проектовану систему гнучкішою.

Ключові слова: кластер, розклад зайнятості.