Intelligent information systems

UDC 004. 02:004.9 DOI: https://doi.org/10.20998/2522-9052.2022.4.09

Olena Gavrilenko[1], Oleksandr Khomenko[1], Oksana Zhurakovska[1],
Alla Kohan[1], Andrii Piskun[2], Olena Khalus[1]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2] EU4DigitalUA, Kyiv, Ukraine

APPLICATION OF ASSOCIATION RULES FOR FORMATION OF PUBLIC (ADMINISTRATIVE) SERVICES PORTFOLIO

Abstract. The formation of complex services is carried out exclusively according to the principle of association based on a life (business) event or situation. But as practical experience shows, users often request several services at the same time, and these services are not connected by one life event. This can be seen if you analyze the statistical data on service delivery by service delivery centers. The application of intelligent data analysis allows to reveal the hidden relationship between services. Based on this, a new principle of combining services was formulated. The object of research is the process of forming a portfolio of services based on association rules. An algorithm for forming a portfolio and providing recommendations for a decision-maker has been developed. Recommendations are presented in the form of sets of services that can be included in a portfolio. The application of this principle will allow to expand the set of complex services due to new portfolios. This will simplify and improve the service delivery process for both authorities and users.

Keywords: complex public service; principle of association by life event; service portfolio; correlation of services; association rule.

Introduction

Public service is a legally or socially significant action of the provider of public service (e-service). The result of this action is the acquisition, transfer or termination of rights and/or performance of obligations by the subject of the application, provision of relevant tangible and/or intangible benefits to the subject of the application. It is provided to the subject of application with or without appeal (application) [1].

Further evolution of the public (administrative) services system is associated with the public services reengineering process [2]. Nowadays, the one factor of successful of public (administrative) e-services reengineering is to identify correlated services and combine them into one group. Currently, the following ways of combining services can be distinguished:

- creation of complex services;
- creation of a services portfolio.

The complex public e-service allows the subject of the application to receive several electronic public services from one or more public services (e-services) providing subjects or online (including using the Unified State Web Portal for E-Services) under one request (application) [3].

Currently, the formation of complex services in Ukraine is based on the principle of combining by life event or situation (or business event). Compared to complex services, a service portfolio (SP) is a set of services that are combined on the basis of another principle (excluding a life event or situation).

Defining the SP formation principles is a very important and relevant problem to improve service providing as well as service reengineering. Paper [4] describes the SP formation process based on the service correlation, which is established as a result of analysis of statistical information about the services providing. It shows the relevancy of recognising different SP formation principles.

It is worth noting that in addition to statistical analysis, such areas of mathematics as mathematical logic, set theory and data mining can be used to determine the SP formation principles. Based on the analysis of the subject area, it is proposed to apply the following principles for the SP formation:

- based on the services correlation, which is established by the presence of information dependence between services;
- based on the services correlation, which is established by the presence of logical sequence on the set of services;
- based on the services correlation, which is established by the presence of use of common data;
- based on the services correlation, which is established by mining association rules.

The possibility of applying other principles of SP formation using data analysis and statistical analysis methods depends on the available information about the services providing.

The problem of establishing the relationship between services is considered for the case when information about the list of services in one transaction is known. Namely, the principle of SP formation based on the mining association rules is considered. The aim of mining association rules is to detect relationships (or associations) between objects that often occur together [5 – 10].

The detected relationships are represented as rules and can be used to predict events or the appearance of objects that are dependent. Similarly, this method is proposed to apply for SP formation based on the detection of dependent services that often occur together in users' requests, but are evidently not related, for example, with the common life event.

Problem statement

Given a set of users (subjects of requests), each user requests some set of services (at the same time or sequentially in time). Further, such an application by one user of a set of services will be called a transaction.

It is necessary to establish implicit dependencies based on transaction information between services, which are the most often used ones and which occur together in the majority of transactions. It is necessary to develop recommendations for the combining of public services into SP based on the identified relationships.

An example of a set of transactions is shown in Table 1.

Each row of the table contains a transaction associated with a user. All services that were requested in this transaction are marked with "1".

Table 1 – **A set of transactions**

	service 1	service 2	service 3	service 4	service 5	service 6	service 7	service 8	service 9	service 10
User 1	1	1	1	1	1	0	0	0	1	1
User 2	1	1	0	0	0	1	0	0	0	0
User 3	0	1	1	0	0	0	0	0	1	0
User 4	0	0	0	0	1	1	0	0	0	1
User 5	0	1	1	0	0	0	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...
User *n*	1	0	1	1	0	0	0	0	0	1

Formalized problem statement. Let

$$X = \{ X_1, X_2, ..., X_n \}$$

be a set of services. The outrank relation *R* over the set *X* can be defined as:

R=$\{(X_i, X_j) \mid$ requesting service X_i causes requesting service $X_j\}$ or as:

$$R = \{(X_i,\ X_j) \mid\ X_i \rightarrow X_j\}. \qquad (1)$$

Let`s introduce the following rules.

Rule 1: only the one of two symmetric pairs of services will be included in the relation that has higher confidence. If symmetric pairs have the same confidence value, only one of them will be included in the relation. This will be the pair for which a chain can be built, as shown below in the description of 3th stage.

Rule 2: if there are two association rules:

$$X_i \rightarrow X_j \text{ and } X_j \rightarrow X_k,$$

the rule $X_i \rightarrow X_k$ should be included in the set of rules.

Thus, the relation *R*, according to rules 1 and 2, is irreflexive, antisymmetric, and transitive, i.e. a strict partial order.

Let *T*=$\{t_i\}$, *i*=1, 2,…, *m* be a set of transactions,

where $t_i = (t_{i1}, t_{i2}, ..., t_{in})$ - the transaction,

$$t_{ij} = \begin{cases} 1, \text{if service } X_j \text{ is present in the transaction } t_i, \\ 0, \text{if service } X_j \text{ is absent in the transaction } t_i. \end{cases}$$

It is necessary to build the relation *R* defined in Equation (1) over the set of services *X* based on information about the set of transactions *T* and to formulate recommendations for the including services in the SP.

Building a relation *R* based on association rules

Let`s introduce the following definitions.

supp(X_i) - the support of service X_i. It is defined as the number of transactions that include this service.

$supp_{\min}(X_i)$ - the support threshold, it is the number *s*, above which the service support value, this service can be considered as frequently occurring.

The association rule

$$X_i \rightarrow X_j,$$

where X_i, X_j denote the services, can be interpreted as follows: if the service X_i occurs in a transaction, then the service X_j is also expected to occur in this transaction. The correctness of the formulated association rule is determined by its confidence.

conf ($X_i \rightarrow X_j$) - the confidence of an association rule. It is defined as the ratio of the support of the set $X_i \cup X_j$ to the support of the service X_i. It means the proportion of transactions that contain a service X_i, in which the service X_j also occurs.

The support of an association rule is the proportion of transactions that contain both the condition and the consequence of the rule to the total number of transactions.

Lift is the ratio of the frequency of occurrence of a condition in transactions that contain both the condition and the consequence to the frequency of occurrence of the consequence in the set of transactions.

The aim of the method is to find a set of association rules with high support and confidence (i.e. above predefined thresholds, which are usually depending on the quality of the statistical data).

Basic algorithms for mining association rules that can be used: Apriori, Eclat, PCY (Park, Chen, Yu), Multistage Algorithm, Multihash Algorithm, FP-GROWTH, SON and MapReduce, Algorithm of Toivonen [5, 6, 9, 10].

First stage: Applying the Apriori algorithm for mining association rules.

Mining association rules. The Apriori algorithm is based on the following statement:

if $L \subseteq M$, then $supp(L) \le supp(M)$,

where *L*, *M* are the subsets of a set of services *X*.

Then, the following properties are true.

Property 1: if *M* is frequent, then any subset of *L* is also frequent.

Property 2: If *L* is rare, then any superset *M* is also rare.

The Apriori algorithm traverses the prefix tree level by level. It connects two elements *L* and *M* if *L* is a subset of *M*. The algorithm calculates how often a subset *L* occurs in the set of transactions *T*. Thus, according to the algorithm:

– rarely occurring subsets and all their super sets are excluded;

– *supp*(*L*) is calculated for each appropriate candidate *L* of size *k* at level *k*.

Prerequisite: Data set *T*, which consist of a transaction list and

$$s = supp_{\min}$$

(X_i) - threshold value $supp(X_i)$ set by the user.

Result: A set of association rules.

It should be noted that only candidates of size *k*=1 were considered in this study.

Second stage: Building a relation *R*.

Based on the mined associative rules on the set of services *X*, construct the relation *R* as follows: include in the relation *R* all pairs of services (X_i, X_j), for which the correctness of the association rule

$$X_i \to X_j$$

is established at the first stage. Extend the relation by applying rule 1.

Third stage: Forming recommendations.

Based on the relation *R*, construct the chains:

$$X_i \to X_j \to \cdots \to X_l ,$$

in which each neighboring pair of services is related by a certain association rule.

Formulate recommendations on the structure of the SP based on the obtained chains.

To visualise the relation *R*, using the graphs is recommended.

Research results

The proposed approach can be illustrated on the example of searching for relationships on a set of services that can be discovered by analyzing a dataset containing transaction data for a set of users.

The dataset from Kaggle [11] is analyzed. The dataset contains 23169 transactions and 3768 services. A part of the dataset is presented in Table 2.

Each dataset transaction (Table 2) contains a set of services that were requested by one user at the same time.

Using the Apriori algorithm at the first stage, the following set of association rules, such as $X_i \to X_j$ was obtained:

In Table 3, the first column is the number of the association rule, the columns "lhs" and "rhs" are the clause and the consequence of the association rule, respectively.

The column «support» is a value of the support of an association rule.

The column «confidence» is a value of the confidence of an association rule.

The column «lift» is a value of the lift of an association rule.

The results are obtained for the thresholds of rule support at the level of 0,005, and confidence at the level of 0,8.

Finally, at the third stage, the following unique chains were built from the obtained association rules:

1) Service 1184 – Service 1179 – Service 1180 – Service 1183;

2) Service 1184 – Service 1181 – Service 1180 – Service 1182;

3) Service 1184 – Service 1181 – Service 1182;

4) Service 1184 – Service 1182 – Service 1183;

5) Service 744 – Service 743 – Service 931.

It is worth noting that each single association rule can be interpreted as a random event. All random events that match the association rules will be pairwise independent.

Then, according to the theorem of multiplication of probabilities of independent random events, we get the value of confidence for the chains 1-5 as shown in Table 4.

Table 2 – **Part of dataset**

Transaction number	Services
1	Service 1940, Service 1302, Service 933, Service 115, Service 118, Service 2170, Service 509
2	Service 634, Service 116
3	Service 888, Service 1024, Service 3734, Service 1765, Service 917, Service 1889, Service 653, Service 654, Service 102, Service 954, Service 260, Service 2082
4	Service 2527, Service 2126, Service 686, Service 696
5	Service 1595
6	Service 3485, Service 285, Service 344, Service 350, Service 1964, Service 1368, Service 478, Service 1450, Service 595, Service 632, Service 43, Service 640, Service 1747, Service 37, Service 36, Service 35, Service 683

Table 3 – **The resulting association rules**

№	Lhs		Rhs	Support	Confidence	Lift
1	{Service 1860}	=>	{Service 1859}	0.006992102	0.8901099	93.74071
2	{Service 1179}	=>	{Service 1180}	0.009581769	0.9327731	89.67394
3	{Service 1180}	=>	{Service 1179}	0.009581769	0.9211618	89.67394
4	{Service 1184}	=>	{Service 1181}	0.008373257	0.9238095	89.55541
5	{Service 1181}	=>	{Service 1184}	0.008373257	0.8117155	89.55541
6	{Service 1184}	=>	{Service 1179}	0.008200613	0.9047619	88.07743
7	{Service 1184}	=>	{Service 1182}	0.008243774	0.9095238	87.43883
8	{Service 1181}	=>	{Service 1179}	0.009236480	0.8953975	87.16582
9	{Service 1179}	=>	{Service 1181}	0.009236480	0.8991597	87.16582
10	{Service 1184}	=>	{Service 1180}	0.008200613	0.9047619	86.98103
11	{Service 1181}	=>	{Service 1180}	0.009322802	0.9037657	86.88526
12	{Service 1180}	=>	{Service 1181}	0.009322802	0.8962656	86.88526
13	{Service 1181}	=>	{Service 1182}	0.009193319	0.8912134	85.67852
14	{Service 1182}	=>	{Service 1181}	0.009193319	0.8838174	85.67852
15	{Service 1182}	=>	{Service 1180}	0.009193319	0.8838174	84.96749
16	{Service 1180}	=>	{Service 1182}	0.009193319	0.8838174	84.96749
17	{Service 1180}	=>	{Service 1183}	0.009279641	0.8921162	84.71082
18	{Service 1183}	=>	{Service 1180}	0.009279641	0.8811475	84.71082
19	{Service 1181}	=>	{Service 1183}	0.009193319	0.8912134	84.62509
20	{Service 1183}	=>	{Service 1181}	0.009193319	0.8729508	84.62509
21	{Service 1179}	=>	{Service 1183}	0.009106996	0.8865546	84.18272
22	{Service 1183}	=>	{Service 1179}	0.009106996	0.8647541	84.18272
23	{Service 1179}	=>	{Service 1182}	0.008934352	0.8697479	83.61489
24	{Service 1182}	=>	{Service 1179}	0.008934352	0.8589212	83.61489
25	{Service 1184}	=>	{Service 1183}	0.007898485	0.8714286	82.74643
26	{Service 771}	=>	{Service 1210}	0.006646813	0.8020833	82.59319
27	{Service 1182}	=>	{Service 1183}	0.009020674	0.8672199	82.34680
28	{Service 1183}	=>	{Service 1182}	0.009020674	0.8565574	82.34680
29	{Service 105}	=>	{Service 1512}	0.008071130	0.8201754	57.58377
30	{Service 744}	=>	{Service 743}	0.012171436	0.8980892	53.90629
31	{Service 144}	=>	{Service 143}	0.007466874	0.8122066	52.71152
32	{Service 105}	=>	{Service 653}	0.008157452	0.8289474	45.08423
33	{Service 105}	=>	{Service 654}	0.008330096	0.8464912	44.57353
34	{Service 744}	=>	{Service 931}	0.011739825	0.8662420	43.91677
35	{Service 743}	=>	{Service 931}	0.013897881	0.8341969	42.29214
36	{Service 1411}	=>	{Service 69}	0.008675385	0.8170732	42.25618
37	{Service 3118}	=>	{Service 69}	0.005999396	0.8034682	41.55258
38	{Service 3118}	=>	{Service 70}	0.006733135	0.9017341	39.56871
39	{Service 2533}	=>	{Service 610}	0.012171436	0.8318584	37.42394
40	{Service 1411}	=>	{Service 70}	0.008804869	0.8292683	36.38886
41	{Service 2052}	=>	{Service 1091}	0.015106392	0.8177570	35.95183
42	{Service 840}	=>	{Service 642}	0.027795762	0.8039950	17.62324

Table 4 – **The resulting confidence values**

Chains	Confidence
Service 1184 – Service 1179 – Service 1180 – Service 1183	0,7528904
Service 1184 – Service 1181 – Service 1180 – Service 1182	0,7379056
Service 1184 – Service 1181 – Service 1182	0,8209517
Service 1184 – Service 1182 – Service 1183	0,7887571
Service 744 – Service 743 – Service 931	0,7491832

Table 4 shows that the obtained chains have a high level of confidence (more than 73%), which indicates their high quality.

The obtained chains are the basis for the formation of recommendations for the forming of service portfolios.

The development of an algorithm for the formation of recommendations based on the discovered chains is the subject of further research.

Conclusions

The solving of the problem will allow to choose the sets of services (or to form the service portfolios) that will be useful to consumers with regard to the integrated use of these services.

The advantage of the proposed approach to the formation of the service portfolios is the ability to integrate services, the relationship between which is hidden and can be detected by the data mining techniques.

It is also worth noting that to implement this approach, it is very important to use statistical information about the services requested by a particular user and to be able to present the data in the form as shown in Table 1.

Thus, it is recommended to focus on receiving such information from service providers. It will allow to use the formulated recommendations in the process of reengineering public services and improve the quality of service providing.

Furthermore, the proposed approach allows to include the time component, that is, the time moments when the user requested certain services. Such problems are solved by sequence analysis and can be the subject of further research.

In addition, there are many metrics that allow to control the quality of the resulting association rules (support, confidence, lift, etc.).

Analysis of the impact of these parameters on the quality of results can also be the subject of further research.

The proposed approach is quite universal and can be proposed not only for the forming of SP, but also in other areas of activities [12].

References

1. (2012*), Pro administratyvni posluhy*: Zakon Ukrainy vid 06.09.2012 No. 5203-VI. [*On Administrative Services*: Law of Ukraine on September 6, 2012 № 5203-VI]", Verkhovna Rada Ukrainy, available at: https://zakon. rada.gov.ua/go/5203-17.
2. Nazarenko, M.O. and Todorova, O.L. (2022), "Reinzhyniryngh v publichnomu upravlinni u sferi nadannja publichnykh poslugh. [Reengineering in public administration in the area of public services]",/ *Naukovi perspektyvy*, No 5(23), pp. 117-126, doi: https://doi.org/10.52058/2708-7530-2022-5(23)-117-126.
3. (2021), *Pro osoblyvosti nadannia publichnykh (elektronnykh publichnykh) posluh*: Zakon Ukrainy vid 15.07.2021 No. 1689-IX. [*On the specifics of the providing of Public Services (digital Public Services)* : Law of Ukraine on July 15, 2021 № 1689-IX], Verkhovna Rada Ukrainy, available at: https://zakon.rada.gov.ua/laws/show/1689-20.
4. Gavrilenko, O., Zhurakovska, O., Kohan, A., Matviichuk, R., Piskun, A., Khavikova, Y. and Khalus, O. (2022), "The principle for forming a portfolio of public services based on the analysis of statistical information",/ *Eastern-European Journal of Enterprise Technologies*, No 3 (3 (117)), pp. 57–64, doi: https://doi.org/10.15587/1729-4061.2022.260136.
5. Leskovec, J., Rajaraman, A. and Ullman, J. (2020). *Mining of Massive Datasets* (3rd ed.), Cambridge: Cambridge University Press, 565 p., doi: https://doi.org/10.1017/9781108684163.
6. Sytnyk, V.F. and Krasnjuk, M.T. (2007), *Intelektualjnyj analiz danykh (dejtamajningh):* Navch. Posib. [*Intelligence data analysis* (datamining)], K., KNEU, 376 p.
7. Aggarwal, C.C., and Yu, P.S. (2001), "Mining Associations with the Collective Strength Approach", *IEEE Transactions on Knowledge and Data Engineering*, No 13(6), pp. 863-873, doi: https://doi.org/10.1109/69.971183.
8. Kumarb, O.S. and Joshi, N. (2016), "Rule Power Factor: A New Interest Measure in Associative Classification", *Procedia Computer Science*, No 93, pp. 12–18, doi: https://doi.org/10.1016/j.procs.2016.07.175.
9. Tjioe, H.C. and Taniar, D. (2005), "Mining Association Rules in Data Warehouses", *International Journal of Data Warehousing & Mining*, No 1(3), pp. 28–62, doi: https://doi.org/10.4018/jdwm.2005070103.

10. Le, Tuong and Vo, Bay (2016), "The lattice-based approaches for mining association rules: a review", *Wiley Interdisciplinary Reviews, Data Mining and Knowledge Discovery*, Vol. 6, Iss. 4, pp. 140–151, doi: https://doi.org/10.1002/widm.1181.
11. *E-commerce Business Transaction (Sales transaction of a UK-based e-commerce (online retail) for one year),* available at: https://www.kaggle.com/datasets/gabrielramos87/an-online-shop-business.
12. Gavrylenko, O. and Protsiuk, Y. (2022), "Approaches to the solution to the problem of news-based events forecasting", *Adaptyvni systemy avtomatychnogho upravlinnja*, Vol. 1, No 40, pp. 65-70, doi: https://doi.org/10.20535/1560-8956.40.2022.261651.

Received (Надійшла).14.09.2022
Accepted for publication (Прийнята до друку).16.11.2022

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Гавриленко Олена Валеріївна – к.ф.-м.н., доцент, доцент кафедри інформаційних систем та технологій, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Olena Gavrilenko – PhD, Docent, Associate Professor at the Department of Information Systems and Technologies, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;
e-mail: gelena1980@gmail.com; ORCID ID: orcid.org/0000-0003-0413-6274.

Хоменко Олександр Миколайович – аспірант кафедри інформаційних систем та технологій, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Oleksandr Khomenko – PhD Student at the Department of Information Systems and Technologies, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;
e-mail: khomenkosasha99@gmail.com; ORCID ID: orcid.org/0000-0003-1964-1097.

Жураковська Оксана Сергіївна – **к.т.н.,** доцент кафедри інформаційних систем та технологій, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Oksana Zhurakovska – PhD, Associate Professor at the Department of Information Systems and Technologies, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;
e-mail: o.zhurakovska@kpi.ua; ORCID ID: orcid.org/0000-0002-2804-5556.

Коган Алла Вікторівна – **к.т.н.,** доцент, доцент кафедри інформаційних систем та технологій, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Alla Kohan – PhD, Docent, Associate Professor at the Department of Information Systems and Technologies, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;
e-mail: a.v.kohan433@gmail.com; ORCID ID: http://orcid.org/0000-0003-2342-8475.

Піскун Андрій Михайлович – керівник компонента 4 «EU4DigitalUA» Академії електронного управління (EGA), Київ, Україна;
Andrii Piskun – Component 4 Leader, "EU4DigitalUA" Electronic Governance Academy (EGA), Kyiv, Ukraine;
e-mail: piskun.an@gmail.com; ORCID ID: https://orcid.org/0000-0001-8662-3257.

Халус Олена Андріївна – **ст. викл.** кафедри інформатики та програмної інженерії, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна;
Olena Khalus – Senior Lecturer at the Department of Informatics and Software Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine;
e-mail: selena.ua@gmail.com; ORCID ID: orcid.org/0000-0002-5485-7881.

Застосування асоціативних правил для формування портфеля публічних (адміністративних) послуг

О. В. Гавриленко, О. М. Хоменко, О. С. Жураковська,
А. В. Коган, А. М. Піскун, О. А. Халус

Анотація. Формування комплексу послуг здійснюється виключно за принципом об'єднання на основі життєвої (ділової) події чи ситуації. Але, як показує практичний досвід, користувачі часто запитують кілька послуг одночасно, і ці послуги не пов'язані однією життєвою подією. Це можна побачити, якщо проаналізувати статистичні дані щодо надання послуг центрами надання послуг. Застосування методів інтелектуального аналізу дозволяє виявити прихований зв'язок між послугами. На основі цього був сформульований новий принцип об'єднання послуг. Об'єктом дослідження є процес формування портфеля послуг на основі асоціативних правил. Розроблено алгоритм формування портфоліо та надання рекомендацій для особи, яка приймає рішення. Рекомендації представлені у вигляді наборів послуг, які можна включити в портфоліо. Застосування цього принципу дозволить розширити набір комплексних послуг за рахунок нових портфелів. Це спростить і покращить процес надання послуг як для органів влади, так і для користувачів.

Ключові слова: комплексна державна послуга; принцип асоціації за життєвою подією; портфель послуг; асоціативні правила.