Networks clustering-based approach for search of reservoirs-analogues

Andrey Bezborodov[1] and Irina Deeva[1] *

ITMO University

Abstract. This article presents a new look at the problem of finding reservoirs-analogues, representing reservoirs as a network and solving the problem of finding reservoirs-analogues as a problem of finding communities in the network. The proposed network approach allows us to effectively search for a cluster of reservoirs-analogues and restore missing parameters in the target reservoir based on the found clusters of reservoirs-analogues. Also, the network approach was compared with the baseline approach and showed greater efficiency. Three approaches were also compared to restore gaps in the target reservoir using clusters of reservoirs-analogues.

Keywords: Networks · Community detection algorithms · Distance metrics · Oil and gas reservoirs · Reservoirs-analogues

1 Introduction

In the early stages of exploration and development of reservoirs, two problems usually arise - high uncertainty in the data and the lack of these data. The decisions that are made at these stages can have a severe impact on the economic development of the entire project. A common practice for solving the above problems is to use information about already investigated reservoirs similar to our target reservoir. Such reservoirs are called reservoirs-analogues. Expert knowledge-based approaches are the simplest method for finding reservoirs-analogues in time and effort. Voskresenskiy et al. proposed searching for reservoirs-analogues using several filters - parameters [8]. The most critical parameter is the distance to the current area with a target reservoir. It is also possible to determine reservoirs-analogues using various machine learning technologies. Various clustering algorithms can be used to obtain groups of reservoirs-analogues with similar properties [3,6].

The methods listed above have their drawbacks, for example, expert knowledge may not always be available, and clustering algorithms are susceptible to dimensionality growth. Therefore, it was decided to look at the problem of finding reservoirs-analogues as the problem of finding communities in a graph structure since all reservoirs can be represented as a network. A wide range of existing applications is in social network analysis [1,9].

* - correspondence author (iriny.deeva@gmail.com)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_30

Suppose we assume that each reservoir is an object in a multidimensional space of features. All objects are interconnected by edges, which are proportional to some distance metric. In that case, the analogous reservoirs are some clusters in this network that are tightly connected. This formulation of the problem requires solving several problems: (1) how exactly to build networks of reservoirs, (2) how to cluster the network to search for reservoirs-analogues, (3) how to use the resulting markup to search for a cluster of reservoirs-analogues for the target reservoir.

The purpose of this article is to solve the above problems and demonstrate the advantage of a network approach for searching for reservoirs-analogues, as well as to present a network approach for recovering gaps in reservoirs using the found reservoirs-analogues clusters Fig. 1.

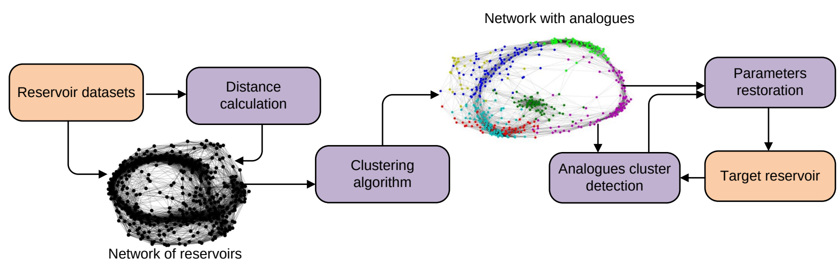


Fig. 1: The pipeline of the proposed network approach to the oil and gas reservoirs-analogues searching.

2 Algorithms and methods

2.1 Networks clustering algorithms

For building a network of reservoirs, an approach was proposed in which a particular distance metric was measured between the parameters of the reservoirs. An edge is drawn between the reservoirs if the metric value is greater than a specific threshold value. As possible distance metrics, we propose cosine distance, Hamming distance, Gower distance and several modified Gower distances - HEOM and HVDM [10]. It is possible to use various community detection approaches in a formed graph. In our study, three algorithms will be compared - Louvain algorithm [2], Leiden algorithm, Newman algorithm [4].

2.2 Search for a cluster of reservoirs-analogues for the target reservoir

When reservoirs are already labelled with clusters of reservoirs-analogues, how to use the resulting labelling for a new target reservoir that has gaps in the

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_30

data, that is, we are simulating a situation where a specialist has a certain reservoir with missing parameters and wants to find a suitable cluster of reservoirs-analogues for this reservoir and use it to restore the gaps. Three approaches are possible to solve this problem (Fig. 2). The first approach is that the missing parameters are excluded from the parameters for building the network. The target reservoir is completely excluded from the network construction in the second approach. Its membership in any cluster is made based on the prediction of FEDOT pipeline [5]. In the third approach, the missing values are first pre-filled based on the values of the entire database with FEDOT pipelines.

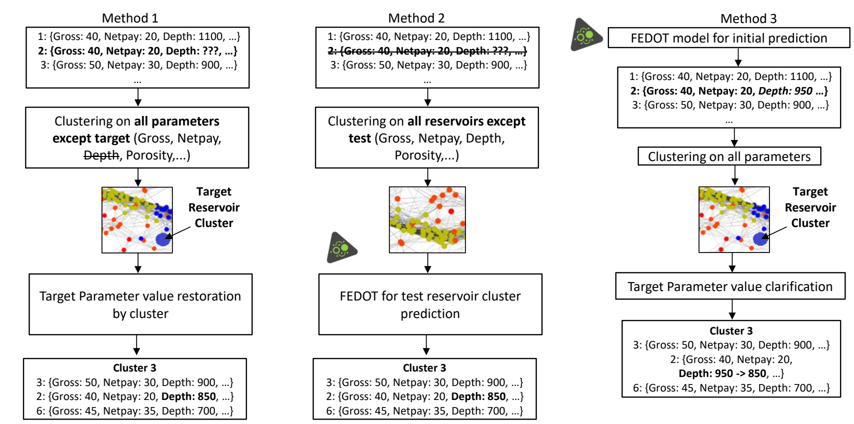


Fig. 2: Three methods of using a network of reservoirs-analogues to restore gaps in the target reservoir.

3 Experiments and results

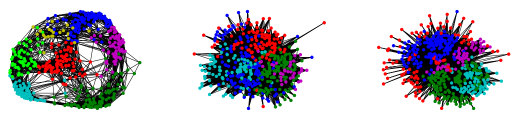
3.1 Using different distance metrics

The analysis and testing of the proposed approaches were carried out on a production database containing 442 reservoirs without missing parameters and 72 reservoirs with gaps in various data. The following experiments were carried out to find out which distance metric is the best for building a network of reservoirs:

1. A network is built based on the distance metric;
2. Clusters are searched in the network by the Leiden algorithm;
3. A reservoir is selected from the database, and it is determined to which cluster it belongs;
4. The values in the reservoir are sequentially deleted and restored based on the values in the cluster. The average restores continuous parameters; the most frequent category restores discrete parameters;

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_30

After filling the gaps in the parameters for all reservoirs, we propose to use $1 - the\_average\_accuracy$ over the cluster for categorical parameters. The Root Mean Squared Error (RMSE) divided by the parameter's range is used for continuous parameters. Fig. 3 shows visualizations of graphs constructed using various metrics with nodes-reservoirs in the color of clusters for Leiden algorithm. The results obtained for these metrics with the further restoration of gaps in the parameters are presented in Table 1. It can be seen from the table that for most parameters, the cosine distance metric shows the best results.



(a) (b) (c)

Fig. 3: Clustered reservoir networks built on different distance metrics. (a) - cosine distance, (b) - Gower distance, (c) - Hamming distance.

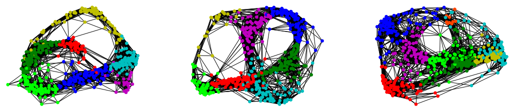
Table 1: The results of restoring parameter values based on the found clusters of analogs for different distance metrics (less is better).

Parameter	**Cosine**	**Gower**	**Hamming**	**HEOM**	**HVDM**
Depth	**0.098**	0.184	0.182	0.174	0.184
Gross	**0.061**	0.118	0.119	0.117	0.118
Netpay	**0.076**	0.092	0.098	0.094	0.093
Permeability	**0.093**	0.122	0.119	0.126	0.123
Porosity	**0.124**	0.139	0.135	0.141	0.138
Tectonic Regime	**0.5**	**0.5**	0.45	0.52	0.52
Hydrocarbon Type	**0.21**	0.24	0.25	0.29	0.25
Structural Setting	0.67	0.65	**0.63**	0.72	0.68
Period	**0.69**	0.79	0.73	0.72	0.7
Lithology	0.46	0.44	**0.42**	0.43	0.36
Depositional System	0.65	0.62	0.61	0.59	**0.58**
Trapping Mechanism	**0.51**	**0.51**	0.54	0.52	0.54

3.2 Using different clusters detection algorithms

The structure of experiments is the same as in the previous section (Section 3.1). Fig. 4 shows the results of clustering the reservoirs network by three different algorithms.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_30



(a) (b) (c)

Fig. 4: The results of clustering networks of reservoirs by three algorithms - (a) - Louvain algorithm, (b) - Leiden algorithm, (c) - Newman algorithm.

An additional comparison was made with a more simplified approach (baseline) for finding reservoir-analogues, which was presented in [7]. The gap recovery results in parameters using described three community detection algorithms with a baseline approach are presented in Table 2. We can conclude that for most parameters, the formation of clusters using the Leiden algorithm makes it possible to obtain a more accurate recovery of gaps in the parameters.

Table 2: The results of restoring parameter values based on the found clusters of analogs for different clusters detection algorithms(less is better).

Parameter	**Leiden Algorithm**	**Louvain Algorithm**	**Newman Eigenvector Algorithm**	**Baseline Approach**
Depth	**0.098**	0.103	0.099	0.155
Gross	**0.061**	0.065	0.068	0.076
Netpay	**0.076**	0.078	0.081	0.081
Permeability	**0.093**	0.101	0.097	0.118
Porosity	**0.124**	**0.124**	0.125	0.127
TectonicRegime	**0.500**	0.580	0.560	0.570
HydrocarbonType	**0.210**	0.260	0.240	0.250
StructuralSetting	**0.670**	0.790	0.690	0.810
Period	**0.690**	0.730	0.740	0.790
Lithology	**0.460**	0.490	**0.460**	0.650
DepositionalSystem	**0.650**	**0.650**	0.670	0.720
TrappingMechanism	**0.510**	0.520	0.550	0.520

3.3 Methods for parameters recovery

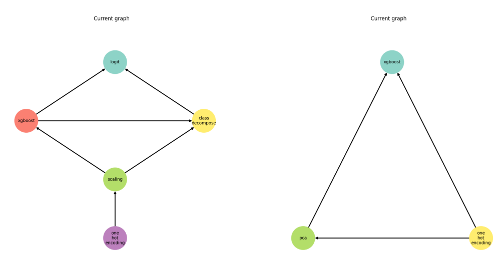
We explored three approaches(Section 2.2) using the Leiden algorithm to find clusters with reservoirs-analogues. For the experiment, ten target reservoirs were selected with four random missing parameters; in total, 30 runs for each method were performed. The results obtained are presented in Table 3. For the prediction task, FEDOT [5] was used, since it allows us to find the best model for classification and regression problems. The Fig. 5 shows the most common

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_30

pipelines for classification (cluster prediction) and regression (gap prefill). For most parameters, the third method, showed the best results.

Table 3: The results of restoring parameter values based on cluster values using three approaches of working with the network and the target reservoir (less is better).

Parameter	**Method 1**	**Method 2**	**Method 3**
Depth	0.238	0.278	**0.223**
Gross	0.228	0.216	**0.205**
Netpay	0.161	0.145	**0.132**
Permeability	0.219	0.266	**0.197**
Porosity	0.238	0.229	**0.212**
Tectonic Regime	**0.47**	0.54	0.53
Hydrocarbon Type	**0.17**	0.25	0.19
Structural Setting	0.74	0.73	**0.64**
Period	0.71	0.76	**0.67**
Lithology	**0.44**	0.49	0.51
Depositional System	0.67	0.64	**0.62**
Trapping Mechanism	0.59	0.56	**0.41**



(a) (b)

Fig. 5: The most frequent FEDOT pipelines for classification and regression problems.

4 Conclusion

This article presented an approach for searching for analogous reservoirs based on a network representation of reservoirs and searching for clusters in such a network. The reservoirs-analogues obtained by using the Leiden algorithm turned

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_30

out to be the most interconnected compared to other studied approaches. Furthermore, all proposed clustering algorithms turned out to be more efficient than the baseline approach. Also, the most efficient method was obtained for processing gaps in the target parameters for further clustering. In the future, it would be interesting to create a kind of hybrid approach that considers expert knowledge and the results of the community search algorithm.

5 Acknowledgement

This research is financially supported by the Ministry of Science and Higher Education, agreement FSER-2021-0012.

References

1. Atay, Y., Koc, I., Babaoglu, I., Kodaz, H.: Community detection from biological and social networks: A comparative analysis of metaheuristic algorithms. Applied Soft Computing **50**, 194–211 (2017)
2. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10), P10008 (2008)
3. Dumont, M., Reninger, P.A., Pryet, A., Martelet, G., Aunay, B., Join, J.L.: Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies. Journal of Applied Geophysics **157**, 1–9 (2018)
4. Newman, M.E.: Finding community structure in networks using the eigenvectors of matrices. Physical review E **74**(3), 036104 (2006)
5. Nikitin, N.O., Vychuzhanin, P., Sarafanov, M., Polonskaia, I.S., Revin, I., Barabanova, I.V., Maximov, G., Kalyuzhnaya, A.V., Boukhanovsky, A.: Automated evolutionary approach for the design of composite machine learning pipelines. Future Generation Computer Systems **127**, 109–125 (2022)
6. Ren, Z., Sun, L., Zhai, Q.: Improved k-means and spectral matching for hyperspectral mineral mapping. International Journal of Applied Earth Observation and Geoinformation **91**, 102154 (2020)
7. Rodríguez, H., Escobar, E., Embid, S., Rodriguez, N., Hegazy, M., Lake, L.: New approach to identify analogous reservoirs. SPE Economics and Management **6** (09 2013). https://doi.org/10.2118/166449-MS
8. Voskresenskiy, A., Butorina, M., Popova, O., Bukhanov, N., Filippova, Z., Brandao, R., Segura, V., Brazil, E.V.: Variations in ranked list of reservoir analogs as an effect of search preferences. In: Saint Petersburg 2020. vol. 2020, pp. 1–5. European Association of Geoscientists & Engineers (2020)
9. Wang, C., Tang, W., Sun, B., Fang, J., Wang, Y.: Review on community detection algorithms in social networks. In: 2015 IEEE international conference on progress in informatics and computing (PIC). pp. 551–555. IEEE (2015)
10. Wilson, D.R., Martinez, T.R.: Improved heterogeneous distance functions. Journal of artificial intelligence research **6**, 1–34 (1997)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_30