

Hyperplane Clustering of the Data in the Vector Space of Features Based on Pseudo Inversion Tools

Iurii Krak^{a,b}, Hrygorii Kudin^b, Veda Kasianiuk^a and Mykola Efremov^{a,b}

^a Taras Shevchenko National University of Kyiv, Kyiv, 64/13, Volodymyrska str., 01601, Ukraine

^b Glushkov Cybernetics Institute, Kyiv, 40, Glushkov ave., 03187, Ukraine

Abstract

Method of the hyperplane data clustering in the vector space characteristics features based on the results of the theory of perturbation of pseudo-inverse and projective matrices and solutions of systems of linear algebraic equations (SLAE) is proposed. The algorithm of a hyperplane clustering with the verification of a given criterion for the effectiveness of the proposed method of the clustering is developed. An example of using the method of scaling characteristic features for recognizing the fingerspelling alphabet of the sign language is given.

Keywords

clustering, pseudo-inverse matrix, SLAE, optimization

1. Introduction

One of the important problems of classification and clustering of information is the problem of minimizing the dimension of the space of features, the choice of criteria for optimal solutions in the processes of practical use. Such problems are effectively solved by the method of multidimensional scaling of empirical data on the proximity of objects, with the help of which the dimension of the measured objects essential characteristics space is determined and the configuration of points - objects in this space - is constructed. This space is a multidimensional scale, similar to the commonly used scales in various applications in the sense that the values of the specially generated essential characteristics of the measured objects correspond to certain positions on the axes of the new space [1]-[5].

The purpose of this work is the development of mathematical methods for the synthesis of systems for solving problems of classification and clustering based on information about the characteristic features of objects [6]-[10]. These problems are proposed to be solved by constructing hyperplanes in a space derived from the original space of features using theory of perturbation of pseudo-inverse and projection matrices and solving systems of linear algebraic equations. The paper proposes a method for synthesizing a piecewise hyperplane cluster to isolate the most effective characteristic features and an algorithm for constructing piecewise hyperplane clusters that allow you to find an effective solution to the problems. The performance and efficiency of the proposed approach is shown on the example of the scaling of characteristic features for recognizing the letters of the fingerprint alphabet of the sign language [8], [9], [11].

2. Method of synthesis of the hyperplane cluster concepts

The problem of synthesis of a piecewise hyperplane cluster for a training sample of vectors $\Omega^0 = \{x : x(j) \in E^m, j = \overline{1, n}\}$, where $x(1), \dots, x(n)$ - vectors from the Euclidean space of features E^m

International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2021), September 20–21, 2021, Kharkiv, Ukraine

EMAIL: krak@univ.kiev.ua (I.Krak); gkudin@ukr.net (H.Kudin); veda.kasianiuk@gmail.com (V.Kasianiuk); nick.yefremov.in@gmail.com (M.Efremov)

ORCID: 0000-0002-8043-0785 (I. Krak); 0000-0002-7310-2126 (H.Kudin); 0000-0003-3268-303X (V.Kasianiuk); 0000-0001-8698-3957 (M.Efremov)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

is to build a cluster so that the training sample points in this space are located quite close, in the sense of a given distance criterion, to some set of hyperplanes that are formed according to this sample.

Note that in the formulation of this problem for clustering of information, the components of the set of hyperplanes are not known in advance. Therefore, for the correct construction of piecewise hyperplane clustering procedures, it is assumed that vectors $x^{(1)}, \dots, x^{(n)}$ from the space of features E^m can belong to one of several hyperplanes $L(A(k), b(k))$, where $A(k) \in E^{s \times m}$, $b(k) \in E^s$, $k = 1, 2, \dots$ some given dimension s , ($s < m$). Here $A(k)$ and $b(k)$ are matrix and vector parameters, respectively, for a fixed hyperplane $L(A(k), b(k))$ $k = 1, 2, \dots$.

The proposed method of a piecewise-hyperplane cluster synthesis is based on using the representation of hyperplanes by means of a set of solutions (pseudo solutions) of systems of algebraic equations.

$$A(k)x = b(k), \quad (1)$$

$$\begin{aligned} &L(A(k), b(k)) = \\ &= \{x \in E^m : x = A^+(k)b(k) + Z(A(k))z, z \in E^m\}. \end{aligned} \quad (2)$$

Here A^+ - pseudo inverse matrix, Z - projection matrix.

Let us give some mathematical results on the inversion (pseudo inversion) of matrices and the construction of projection matrices, which are important for solving the problem of synthesizing a piecewise-hyperplane cluster.

Let the matrix be given $A = (a_{ij}), i = \overline{1, m}, j = \overline{1, n}$ and we write down, which are important in further studies, the representations of this matrix in columns or rows, respectively:

$$\begin{aligned} A &= (a(1) : \dots : a(n)), a(j) \in E^m, j = \overline{1, n}, \\ A &= \begin{pmatrix} a_{(1)}^T & \dots & a_{(m)}^T \end{pmatrix}^T, a_{(i)}^T \in E^n, i = \overline{1, m}, \end{aligned}$$

where T - transpose symbol.

We consider the singular decomposition of an arbitrary matrix A of dimension $m \times n$ of rank $r \leq \min(m, n)$ in the form $A = \sum_{i=1}^r \lambda_i u_i v_i^T$, where: $\lambda_1^2 \geq \dots \geq \lambda_r^2 > 0$ - non-zero eigenvalues of matrices $AA^T, A^T A$; $v_i \in E^n, i = \overline{1, r}$ - orthonormal set of eigenvectors of the matrix $A^T A$, which correspond to non-zero eigenvalues $\lambda_i^2, i = \overline{1, r}$, $A^T A v_i = \lambda_i^2 v_i$, $i = \overline{1, r}, v_i^T v_j = \delta_{ij}$ $u_i \in E^m, i = \overline{1, r}$ - orthonormal set of eigenvectors of the matrix AA^T , which also correspond to non-zero eigenvalues $\lambda_i^2, i = \overline{1, r}, AA^T u_i = \lambda_i^2 u_i$, $i = \overline{1, r}, u_i^T u_j = \delta_{ij}$, δ_{ij} - Kronecker's symbol.

Let us give the definition of a pseudo inverse matrix in the Penrose optimization form [12]. For a matrix $A \in E^{m \times n}$, a pseudo inverse matrix $A^+ \in E^{n \times m}$ is defined by the relation:

$$\forall b \in E^m \quad A^+ b = \arg \min_{x \in \Omega_A(b)} \|x\|^2.$$

$$\text{Here } \Omega_A(b) = \text{Arg} \min_{x \in E^n} \|Ax - b\|^2.$$

Also, using the singular representation of the matrix $A \in E^{m \times n}$, the pseudo inverse matrix $A^+ \in E^{n \times m}$ can be represented as [13]: $A^+ = \sum_{j=1}^r v_j u_j^T \lambda_j^{-1}$.

Will consider important for practical application, matrices that are defined and calculated using matrices A и A^+ :

1) a projection matrix $P(A) = A^+ A \equiv \sum_{i=1}^r v_i v_i^T$ that is an orthogonal projector onto the subspace L_{A^T} generated by the row vectors of the matrix A ;

2) projection matrix $Z(A) = I_n - P(A)$ - orthogonal projection onto a subspace orthogonal to a subspace L_{A^T} , I_n - unit matrix;

$$3) \text{ matrix } R(A) = A^+(A^+)^T \equiv \sum_{j=1}^r v_j v_j^T \lambda_j^{-2}.$$

Note the important properties of projection matrices P and Z :

$$P(A) + Z(A) = I_n, \quad P(A) = A^+ A,$$

$$P(A) = \sum_{i=1}^r v_i v_i^T, \quad Z(A) = \sum_{i=r+1}^n v_i v_i^T.$$

Note that the calculation of the pseudo inverse matrix for an arbitrary matrix can be reduced to the calculation of the corresponding to it some inverse matrix using the following relations:

$$A^+ = (A^T A)^+ A^T = A^T (A A^T)^+, \quad A^+ = R(A) A^T.$$

Using the above mathematical ratios for the inversion and pseudo inversion of matrices, we write the necessary formulas for the synthesis of a piecewise hyperplane cluster.

The distance $\rho(x(j), L(A, b))$ from the point $x(j)$ to the hyperplane $L(A, b)$ will be found from the relation

$$\rho^2(x(j), L(A, b)) = \|A^+(b - Ax(j))\|^2, \quad \|x\|^2 = x^T x.$$

To calculate the sum of the squares of the distances of the set of points $x(j), j = \overline{1, n}$ to the hyperplane $L(A, b)$, we use the following formula:

$$\rho^2(\{x: x(j), j = \overline{1, n}\}, L(A, b)) = \sum_{j=1}^n (b - Ax(j))^T R(A^T) (b - Ax(j)) = \text{tr} R(A^T) \sum_{j=1}^n (b - Ax(j))(b - Ax(j))^T.$$

Here $\text{tr}()$ - matrix trace.

Then, for given values $A, x(j), j = \overline{1, n}$, the optimal value of the vector of the right-hand side of the system of equations determining the hyperplane is determined from the conditions

$$b_{opt} = A \hat{x} =$$

$$= \arg \min_{b \in R^s} \rho^2(\{x: x(j), j = \overline{1, n}\}, L(A, b)), \quad \hat{x} = \frac{1}{n} \sum_{j=1}^n x(j).$$

From here, the distance $\rho(\{x: x(j), j = \overline{1, n}\}, L(A, b))$ of the set of points $x(j), j = \overline{1, n}$ to the hyperplane, with an optimal vector b_{opt} , is calculated by the following formula:

$$\rho(\{x: x(j), j = \overline{1, n}\}, L(A, b_{opt}(A))) = (\text{tr} A^+ A \tilde{X} \tilde{X}^T)^{\frac{1}{2}},$$

where $\tilde{X} = (\tilde{x}(1) : \dots : \tilde{x}(n))$, $\tilde{x}(j) = x(j) - \hat{x}, j = \overline{1, n}$.

The optimal matrix $A \in E^{s \times m}$ is defined as a solution of the problem

$$A_{opt} = \arg \min_{A A^T = E_s, A \in E^{s \times m}} \rho^2(\{x: x(j), j = \overline{1, n}\}, L(A, b_{opt}(A))) = \left(u_{m-s+1}^T \quad \dots \quad u_m^T \right)^T.$$

Wherein $\text{tr} A_{opt}^+ A_{opt} \tilde{X} \tilde{X}^T = \sum_{j=m-s+1}^r \lambda_j^2, (u_1, \dots, u_m)^T (u_1, \dots, u_m) = I_m$.

Using the above results on pseudo inversion of matrices and calculation of distances, a piecewise hyperplane clustering method is proposed. The idea of the method is to perform a sequence of steps, at each of which the parameters of the hyperplanes $L(A(k), b(k)), k = 1, 2, \dots$ are found. These hyperplanes are constructed within the framework of performance the requirements of the implemented hyperplane clustering efficiency criterion. As initial parameters of piecewise hyperplane clustering it is assumed that all the training set vectors $x(1), \dots, x(n)$ from the space of features E^m are optimally approximated by the hyperplane $L(A_{opt}(1), b_{opt}(1))$. At these values $A_{opt}(1), b_{opt}(1)$, the performance efficiency criterion of hyperplane clustering is checked. If the efficiency criterion is satisfied, this means that the result of the construction of the piecewise hyperplane cluster is achieved by building the cluster with one hyperplane and those characteristic features are the optimal set. If the conditions of the efficiency criterion in the framework of this (first) hyperplane are not met, then the

transition to the construction of the second hyperplane of the cluster is carried out. For this purpose, the vectors that determine the non-fulfillment of the efficiency criterion are excluded from the training sample, that is, a subset is formed $\Omega^1 = \{x: x(j_1) \in E^m, j_1 = \overline{1, n_1}\}$. Then the actions on the optimal approximation of the subset Ω^1 by the hyperplane $L(A_{opt}(2), b_{opt}(2))$ are repeated, and at the new optimal values $A_{opt}(2), b_{opt}(2)$ the fulfillment of the clustering efficiency criterion is checked. When the criterion is met, the result of the piecewise hyperplane cluster construction is reached, then the resulting vectors of characteristic features are added to the optimal set of features. When not performing - it is necessary to exclude from the training sample those vectors that cause to the non-fulfillment of the efficiency criterion and repeat the procedure. Note that this finite recurrent process will always ensure the construction of an optimal piecewise hyperplane cluster.

3. Algorithm of synthesis of the piecewise hyperplane cluster

Using the proposed method of synthesizing piecewise hyperplane clusters, the algorithm of such a synthesis can be represented as the following sequence of actions:

1. Formation of a single-link cluster (the number of a link in a cluster is an index in brackets):

1) All vectors of the training sample $\Omega(0) = \{x: x(j) \in E^m, j = \overline{1, n}\}$ from the space of features are to be optimally approximated by a hyperplane $L(A_{opt}(1), b_{opt}(1))$, which is defined as the set of solutions (pseudo-solutions) of systems of algebraic equations (1), (2), where $A(k)$ and $b(k)$ are respectively the matrix and vector parameters of a certain hyperplane (basic formulas for constructing a hyperplane cluster are implemented).

2) To form a set $\Omega(1) = \{x: x(j_1) \in E^m, j_1 = \overline{1, n_1}\}$ of points $\Omega(0)$ for which the condition is satisfied, namely:

$$\begin{aligned} & (b_{1opt}(1) - A_{1opt}(1)x(j_1))^T R(A_{1opt}^T(1)) \times \\ & \times (b_{1opt}(1) - A_{1opt}(1)x(j_1)) > h_{min}, \end{aligned}$$

where h_{min} - valid distance of vectors from components of the cluster of hyperplanes. The linear dependence or independence of vectors that can be removed from the set makes it possible to simplify the form of the formulas for the distances of these vectors from the corresponding hyperplanes.

3) The stop of the algorithm at the stage of the first cluster link can be completed after the distance of each of the vectors of the training sample $\Omega(0) = \{x: x(j) \in E^m, j = \overline{1, n}\}$ to the hyperplane $L(A_{opt}(1), b_{opt}(1))$ does not exceed the allowable distance.

4) If the set $\Omega(1) = \{x: x(j_1) \in E^m, j_1 = \overline{1, n_1}\}$ includes at least two vectors, then go on to form the second link of the cluster.

2. Formation of the second cluster link. Consists of the following:

1) From the set obtained in the process of building the first link $\Omega(1) = \{x: x(j_1) \in E^m, j_1 = \overline{1, n_1}\}$, a hyperplane $L(A_{opt}(2), b_{opt}(2))$ defined as the set of solutions (pseudo - solutions) of a system of algebraic equations

$$A_k(2)x = b_k(2), \quad k = 1, 2, \dots \quad (3)$$

is constructed.

2) To calculate optimal $A_{kopt}(2), b_{kopt}(2)$ for $L(A_k(2), b_k(2))$, $k = 1, 2, \dots$

3) To form a set $\Omega_k^0(2)$, $k = 1, 2, \dots$ by the distance of the vector $x(j) \in \Omega(2)$ to each of the hyperplanes (3):

$$\rho^2(x: x(j), L(A_{kopt}(2), b_{kopt}(2))) = (b_{kopt}(2) - A_{kopt}(2)x(j))^T R^T(A_{kopt}(2)) \times (b_{kopt}(2) - A_{kopt}(2)x(j)),$$

i.e., perform actions similar to paragraph 3 of the construction of the first link.

4) To go to step 2 with new subsets $\Omega_1^j(1), \Omega_2^j(1)$, $j = 1, 2, \dots$. The superscript j denotes the number of iterations at the second link stage.

The algorithm is stopped at the stage of the second cluster link after the distances of each of the vectors of the corresponding partitioning element to the corresponding hyperplane are not improved.

The efficiency criterion of the carried out hyperplane clustering is checked (for example, the requirement of the level of compactness of the cluster links). When the criterion is fulfilled, the cluster is completed, and if it is unfulfilled, the transition to the formation of the third cluster link is carried out. Then the process repeats and, since it is finite, we will always find a solution to the clustering problem.

4. Experimental studies

To test the effectiveness of the proposed method of scaling information, we used characteristic features for recognizing the dactyl letters of the Ukrainian sign language alphabet [6], [7]. As characteristic features, 52 features were taken, which were divided into 6 groups, depending on the method of their getting. The construction of piecewise hyperplane clusters was carried out with groups of features that characterize the geometric - topological parameters of the human hand when showing the letters of the dactyl alphabet and for which an acceptable recognition quality was obtained [6]. Using the example of the classification of nine dactylemes (А, Б, В, Г, Ж, І, Є, И, Ї) according to three and five characteristic features, the separation of these dactylemes on the scaling plane was obtained. Wherein three characteristic features were taken: compactness, directionality, elongation, and in experiments with five features: the ratio of width to height, four angles values between vectors drawn from the center of the hand to the most distant points of the hand.

Experimental results using these three characteristic features are given in Fig. 1, where image of the dactylemes location on the scaling plane is normalized on the interval 0,1. Without reducing the generality of research, dactylem А was placed at the origin.

Dactylemes	X	Y	D
A	0.000 0	0.00 00	0.0354
Б	0.719 6	0.95 17	0.0073
В	0.457 9	0.66 44	0.0519
Г	0.449 5	0.60 88	0.1107
Ж	0.454 7	0.70 62	0.0536
І	0.325 7	0.44 15	0.0306
Є	0.299 8	0.52 37	0.1274
И	0.408 3	0.59 08	0.0643
Ї	0.406 7	0.62 52	0.0707

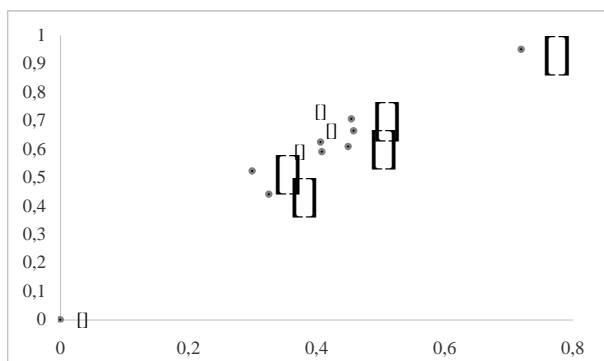


Figure 1: Dactylemes location on hyperplane of scaling and distances of dactylemes from hyperplane of scaling for three characteristic features.

The results of clustering with using of five characteristic features are shown in Fig. 2., where the dactylem A was also placed at the origin.

Dactylemes	X	Y	D
A	0.0000	0.0000	0.2719
Б	1.0345	0.6393	0.0177
В	0.5065	0.4320	0.3419
Г	0.5271	0.3670	0.3828
Ж	0.3878	0.3289	0.2511
И	0.5353	0.4159	0.1825
Є	0.3718	0.2753	0.1976
Й	0.5851	0.4470	0.1580
Ў	0.4585	0.2876	0.1647

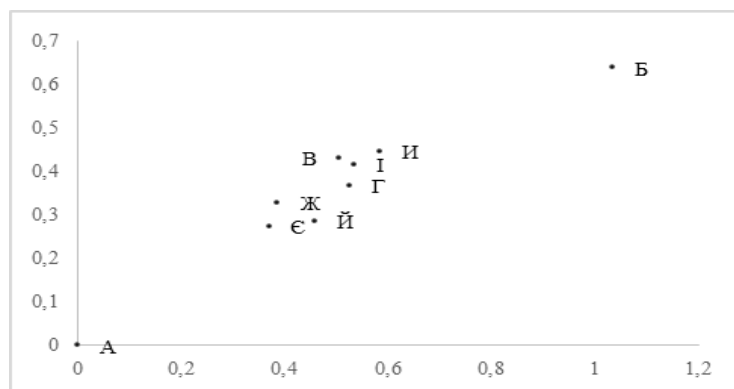


Figure 2: Dactylemes location on hyperplane of scaling and distances of dactylemes from hyperplane of scaling for five characteristic features.

The results of the experiments showed that using of five characteristic features makes it possible to get a clearer separability (the distance of the dactylemes from the scaling plane ranged from 0.1580 to 0.3828), while using three features the distances from the scaling plane were significantly less (ranging from 0.0306 to 0.1274) that is, three to five times less. The exception was dactylem Б, in the first case (0.0177) and in the second case (0.0073), the separability from the scaling plane was not insignificant in both cases.

5. Discussion and conclusions

The paper proposes method and algorithm for multidimensional scaling of recognition objects characteristic features using the means of matrix pseudo inversion, building piecewise hyperplane clusters that provide a solution of the problem [14]-[18]. Proposed method makes it possible to analyze information about the set of characteristic features and to identify those that are essential for solving the recognition problem that is important due to their significant number and for difficult separable classes [19]-[24]. The effectiveness of the proposed approach is shown by the example of obtaining clusters for dactylemes sign language in order to obtain the optimal number of characteristic features for effective recognition.

6. References

- [1] M. Devison. Multidimensional scaling. Moscow: Finansy I Statistika, 1988. p. 254.
- [2] P. Perera, P. Oza, V. M. Patel. One-Class Classification: A Survey. (2021) arXiv preprint arXiv:2101.03064.
- [3] M. Z. Zaheer, J. H. Lee, M. Astrid, A. Mahmood, & S. I. Lee. Cleaning label noise with clusters for minimally supervised anomaly detection, 2021. arXiv preprint arXiv:2104.14770.
- [4] W. Sultani, C. Chen, & M. Shah. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE conference on computer vision and pattern recognition, (2018) 6479-6488.
- [5] I.V.Krak, G.I.Kudin & A.I.Kulias. Multidimensional Scaling by Means of Pseudoinverse Operations, *Cybernetics and Systems Analysis*, 55(1) (2019) 22-29. doi: 10.1007/s10559-019-00108-9.
- [6] L. Cheng, Y. Wang, X. Liu, & B. Li. Outlier Detection Ensemble with Embedded Feature Selection. In Proceedings of the AAAI Conference on Artificial Intelligence 34(04), (2020) 3503-3512. <https://doi.org/10.1609/aaai.v34i04.5755>.
- [7] S. A. N. Alexandropoulos, S. B. Kotsiantis, V. E. Piperigou, & M. N. Vrahatis. A new ensemble method for outlier identification. In 2020 10th International Conference on Cloud Computing, Data Science & Engineering, IEEE, (2020) 769-774. doi: 10.1109/Confluence47617.2020.9058219.
- [8] Iu.G. Kryvonos, Iu.V. Krak, O.V. Barmak, D.V. Shkilniuk. Construction and identification of elements of sign communication. *Cybernetics and Systems Analysis*. 49 (2) (2013). p. 163-172.
- [9] Yu.V. Krak, A.A.Golik, V.S.Kasianiuk. Recognition of dactylemes of Ukrainian sign language based on the geometric characteristics of hand contours defects. *Journal of Automation and Information Sciences*. 48(4) (2016). p. 90-98.
- [10] J. T. O'Brien, C. Nelson. Assessing the Risks Posed by the Convergence of Artificial Intelligence and Biotechnology. *Health security*, 18(3), (2020) 219-227. <https://doi.org/10.1089/hs.2019.0122>.
- [11] I.V. Krak, O.V. Barmak & S.O. Romanyshyn. The method of generalized grammar structures for text to gestures computer-aided translation. *Cybernetics and Systems Analysis*, 50(1) (2014) 116-123. doi: 10.1007/s10559-014-9598-4.
- [12] R. Penrose. A generalized inverse for matrices. *Proceeding of the Cambridge Philosophical Society*. 51. 1955. p. 406-413.
- [13] A. Ben-Israel, T.N.E. Greville. *Generalized inverse: Theory and Applications*. (2-nd Ed.). – Springer-Verlag, New York, 2003. p. 420.
- [14] G. Markowsky, O. Savenko, S. Lysenko, A. Nicheporuk. The Technique for Metamorphic Viruses' Detection Based on Its Obfuscation Features Analysis. In ICTERI Workshops. CEUR Workshop Proceedings, Vol. 2104, (2018), 680-687
- [15] I. Krak, O. Barmak, & E. Manziuk. Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology, *Computational Intelligence* (2020) 1-26. <https://doi.org/10.1111/coin.12289>.

- [16] A.V. Barmak, Y.V. Krak, E.A. Manziuk & V.S. Kasianiuk. Information technology of separating hyperplanes synthesis for linear classifiers. *Journal of Automation and Information Sciences*, 51(5) (2019) 54-64. Doi: 10.1615/JAutomatInfScien.v51.i5.50.
- [17] D. Oosterlinck, D. F. Benoit & P. Baecke. From one-class to two-class classification by incorporating expert knowledge: Novelty detection in human behaviour. *European Journal of Operational Research*, 282(3), (2020) 1011-1024. <https://doi.org/10.1016/j.ejor.2019.10.015>.
- [18] D. Abati, A. Porrello, S. Calderara, R. Cucchiara. Latent space autoregression for novelty detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2019) 481-490.
- [19] D. Gong, L. Liu, V. Le, B. Saha, M.R. Mansour, S. Venkatesh, A.v.d. Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, (2019) 1705 - 1714
- [20] C. You, D. P. Robinson, R. Vidal. Provable self-representation based outlier detection in a union of subspaces. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, (2017) 4323 – 4332*.
- [21] C.-H. Lai, D. Zou, G. Lerman. Robust subspace recovery layer for unsupervised anomaly detection. in *Proc. Int. Conf. Learn. Represent.*, (2020) 1 - 28. arXiv:1904.00152.
- [22] Z. Cheng, E. Zhu, S. Wang, P. Zhang & W. Li. Unsupervised Outlier Detection via Transformation Invariant Autoencoder. *IEEE Access*, 9, (2021) 43991-44002. doi: 10.1109/ACCESS.2021.3065838.
- [23] H. Wang, M. J. Bah, & M. Hammad. Progress in outlier detection techniques: A survey. *IEEE Access*, 7, (2019) 107964-108000. doi: 10.1109/ACCESS.2019.2932769.
- [24] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, ... & K. R. Müller. A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*. 2021. doi: 10.1109/JPROC.2021.3052449.