IEEE Turkey Section

UBMK'2021

Bildiriler Kitabı
Proceedings

6. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı

6th International Conference on Computer Science and Engineering

15-16-17 Eylül (September) 2021 Ankara- Turkey

UBMK'2021'e Hoşgeldiniz
Welcome to UBMK'2021

Sevgili Katılımcılar:

UBMK uluslararası nitelikli konferans serisi, 1990 yılından beri düzenli olarak yapılmakta olan Bilgisayar Mühendisliği Bölüm Başkanları toplantılarında alınan bir kararla altı yıl önce başlamıştır. Konferansın 6.sı UBMK-2021 bu yıl 15-17 Eylül, 2021 tarihlerinde Gazi Üniversitesinin ev sahipliğinde düzenlemiştir. COVID-19 nedeniyle konferans bildirileri çevrimiçi sunulmuş, bazı sunumlar ve toplantılar yüz yüze yapılmıştır.

UBMK-2021 konferansına bu yıl Amerika Birleşik Devletleri, Filistin, Hindistan, Hollanda, İngiltere, Kazakistan, Kırım, KKTC, Novosibirsk, Rusya, Özbekistan, Sırbistan, Tataristan ve Türkiye'den yaklaşık 250 bildiri yollanmış ve bu bildiriler Türk ve yabancı 250 hakem tarafından değerlendirilmiştir.

Her bildiri iki hakem tarafından incelenmiş ve uzlaşma olmadığı durumlarda üçüncü bir hakemin değerlendirmesine başvurulmuştur. Bu değerlendirmelerin sonunda 155 bildirinin sözlü olarak sunulması uygun bulunmuştur. Kabul edilen ve sunulan bildiriler içerik ve kalite ölçünlerini sağlaması durumunda IEEE Xplore Digital Library'de yayımlanacaktır.
Konferans çalışmalarında, Bilgisayar Mühendisliği Bölüm Başkanları Danışma Kurulu olarak görev almışlardır. Bildirilerin değerlendirilmesi Bilim Kurulu üyeleri tarafından yapılmıştır. Konferansın düzenlenmesi ise Yürütme Kurulunun önerileri doğrultusunda, Düzenleme Kurulu tarafından yapılmıştır.

Son olarak, konferansın başarılı bir şekilde yürütülmesi için tüm olanaklarını sunan Gazi Üniversitesi Rektörü Sayın Prof. Dr. Musa Yıldız'a teşekkür ediyoruz.
Ayrıca Düzenleme Kuruluna, bildirileri titizlikle değerlendiren Bilim Kurulu Üyelerine ve değerli araştırmalarının sonuçlarını bilişim camiası ile paylaşan bildiri sahiplerine teşekkürlerimizi iletiriz.

Prof. Dr. Eşref ADALI
UBMK-2021 Konferans Başkanı ve Bildiri Kitabı Editörü

Dear Participants:

The UBMK international conference series started six years ago with a decision taken at the Computer Engineering Department Heads (BMBB) meetings, which have been held regularly since 1990. The 6th edition of the conference, UBMK'21, was held this year on September 15-17, 2021, hosted by Gazi University. Due to COVID-19, conference papers were presented online, and some presentations and meetings were made face-to-face.

Approximately 250 papers from the United States of America, Crimea, India, England, Holland, Kazakhstan, Novosibirsk, Palestine, Russia, Sırbia, TRNC, Uzbekistan, Tatarstan, and Turkey were presented to the UBMK'21 conference this year, and these papers were evaluated by 250 Turkish and foreign referees.

Each paper was evaluated by two referees, and in cases where there was no consensus, a third referee was consulted. At the end of these evaluations, 155 papers were accepted for oral presentation. Accepted and presented papers will be submitted for inclusion into IEEE Xplore Digital Library subject to meeting the scope and quality requirements.

During the conference, Heads of Computer Engineering Departments took part in the Advisory Board. The evaluation of the papers was made by the members of the Scientific Committee. The conference was organized by the Organizing Committee in line with the recommendations of the Executive Committee Members.

Finally, we would like to thank Gazi University Rector Prof. Dr. Musa Yıldız for his continued support for the success of the conference. In addition, we would like to thank the members of Organizing Committee and the Scientific Committee, who carefully evaluated the papers, and the owners of the papers who shared the results of their valuable research with the informatics community.

Prof. Dr. Esref ADALI
UBMK'21 Conference Chair and Proceedings Editor

Danışma Kurulu Üyeleri (Advisory Board Members)

Abant İzzet Baysal	Prof. Dr. Seda Postalcıoğlu
Abdullah Gül	Prof. Dr. Çağrı Güngör
Adana Bilim ve Teknoloji	Doç. Dr. Serdar Yıldırım
Adnan Menderes	Prof.Dr. Yılmaz Kılıçaslan
Ahmet Yesevi	Prof. Dr. Halil İbrahim Bülbül
Akdeniz	Prof. Dr. Melih Günay
Alanya Alaaddin Keykubat	Dr. Öğr. Üy. Özge Öztimur Karadağ
Altınbaş	Dr. Öğr. Üy. Sefer Kurnaz
Amasya	Prof. Dr. Özgen Korkmaz
Ankara	Prof. Dr. Refik Samet
Ankara Bilim	Doç. DR. Hakan Çağlar
Ankara Bilim	Doç. Dr. Eder Sevinç
Arel	Dr. Öğr. Üy. Ferdi Sönmez
Artvin Çoruh	Dr. Öğr. Üy. Şuayip Toprakseven
Atatürk	Prof. Dr. Abdulsamet Haşıloğlu
Atılım	Doç. Dr. Gökhan Şengül
Atılım	Doç. Dr. Korhan Levent Ertürk
Atılım	Prof. Dr. Ali Yazıcı
Avrasya	Dr. Öğr. Üy. Ferudun Keskin
Bahçeşehir	Doç. Dr. Tevfik Aytekin
Balıkesir	Doç. Dr. Ayhan İstanbullu
Bartın	Dr. Öğr. Üy. Evrim Güler
Başkent	Prof. Dr. Nizami Gasilov
Batman	Prof. Dr. Necmettin Sezgin
Beykent	Dr. Öğr. Üy. Ediz Şaykol
Beykent	Dr. Öğr. Üy. Zeynep Altan
Bilecik Şeyh Edebali	Prof. Dr. Cihan Karakuzu
Bingöl	Dr. Öğr. Üy. Kazım Hanbay
Bitlis Eren	Dr. Öğr. Üy. Musa Çıbuk
Boğaziçi	Prof. Dr. Tuna Tuğcu
Bursa Teknik	Doç.Dr. Turgay Tugay Bilgin
Bursa Uludağ	Doç. Dr. Ahmet Emir Dirik
Çanakkale Onsekiz Mart	Doç.Dr. İsmail Kadayıf
Çankaya	Prof. Dr. Hasan Oğul
Çukurova	Prof. Dr. Ramazan Çoban
Dicle	Dr. Öğr. Üy. Mehmet Nergiz
Dicle	Prof. Dr. Siraç Özerdem
Doğuş	Prof. Dr. Yasemin Karagül
Dokuz Eylül	Prof. Dr. Efendi Nasiboğlu
Dokuz Eylül	Prof. Dr. Yalçın Çebi
Düzce	Prof. Dr. Pakize Erdoğmuş
Ege	Prof. Dr. Yasemin Topaloğlu
Erciyes	Prof. Dr. Veysel Aslantaş
Erzincan	Dr. Öğr. Üy. Funda Akar
Erzurum Teknik	Dr. Öğr. Üy. Yusuf Sinan Hanay
Eskişehir Osmangazi	Doç. Dr. Kemal Özkan
Eskişehir Teknik	Dr. Öğr. Üy. Ahmet Arslan
Fatih Sultan Mehmet Vakıf	Dr. Öğr. Üy. Berna Kiraz
Fırat	Prof. Dr. Erhan Akın
Fırat	Prof. Dr. Bilal Alataş
Galatasaray	Prof. Dr. Tankut Acarman
Gazi	Prof. Dr. Şeref Sağıroğlu
Gazi	Prof. Dr. M. Ali Akcayol
Gazi Tek. Fak	Prof. Dr. Ayhan Erdem
Gaziantep	Prof. Dr. Cesur Baransel
Gebze Teknik	Prof. Dr. Erkan Zergeroğlu
Gebze Teknik	Prof. Dr. İbrahim Soğukpınar
Gebze Teknik Bilişim	Dr. Öğr. Üy. Yakup Genç
Gediz	Prof. Dr. Hami Öz
Giresun	Dr. Öğr. Üy. Ali Yılmaz
Gümüşhane	Dr. Öğr. Üy. Özkan Bingöl
Hacettepe	Prof. Dr. Ebru Akçapınar Sezer
Hacettepe	Prof. Dr. İlyas Çiçekli
Hakkari	Dr. Öğr. Üy. Şule Yücebaş
Haliç	Prof.Dr. Zeynep Turgut
Harran	Doç.Dr. İbrahim Berkan Aydilek
Harran	Dr. Öğr. Üy. Serdar Çiftçi
Hasan Kalyoncu	Doç. Dr. M. Fatih Hasoğlu
Hitit	Dr. Öğr. Üy. Mustafa Coşar
Isparta Uygulamalı Bilimler	Prof. Dr. Tuncay Aydoğan
Işık	Dr. Öğr. Üy. Emine Ekin
İhsan Doğramacı Bilkent	Prof. Dr. İbrahim Körpeoğlu
İhsan Doğramacı Bilkent	Prof. Dr. H. Altay Güvenir
İnönü	Prof. Dr. Celaldin Yeroğlu
İstanbul Cerrahpaşa	Prof. Dr. Ahmet Sertbaş
İstanbul Arel	Doç. Dr. Metin Zontul
İstanbul Arel	Dr. Öğr. Üy. Ferdi Sönmez
İstanbul Aydın	Prof. Dr. Ali Güneş
İstanbul Aydın	Doç. Dr. Taner Çevik
İstanbul Bilgi	Prof. Dr. Sedat Ölçer
İstanbul Kültür	Prof. Dr. Murat Taylı
İstanbul Kültür	Dr. Öğr. Üy. Hikmet Çağlar
İstanbul Medeniyet	Dr. Öğr. Üy. Ammar Daşkın
İstanbul Medipol	Doç. Dr. Bahadır Kürşat Güntürk
İstanbul Okan	Prof. Dr. Bekir Tevfik Akgün
İstanbul Sabahattin Zaim	Prof. Dr. Yahya Kemal Yoğurtçu
İTÜ	Prof. Dr. Mustafa Kamaşak
İTÜ	Prof. Dr. Şule Gündüz Öğüdücü
İstanbul Ticaret	Prof. Dr. Rıfat Yazıcı
İzmir Bakırçay	Doç. Dr. Deniz Kılınç
İzmir Ekonomi	Doç. Dr. Senem Kumova Metin
İzmir Ekonomi	Dr. Öğr. Üy.Kaya Oğuz
İzmir Katip Çelebi	Prof. Dr. Ayşegül Alaybeyoğlu
İzmir Yük. Tek. Ens.	Doç. Dr. Tolga Ayav
Kadir Has	Prof. Dr. Feza Kerestecioğlu
Kadir Has	Prof. Dr. Hasan Dağ
Kahramanmaraş Sütçü İmam	Dr. Öğr. Üy. Zeynep Banu Özger
Kahramanmaraş Sütçü İmam	Dr. Öğr. Üy. Mücahid Günay
Karabük	Dr. Öğr. Üy. Hakan Kutucu
Karadeniz Teknik	Prof. Dr. Mustafa Ulutaş
Karadeniz. Of. Tek. Fak.	Doç. Dr. Hamdi Tolga Kahraman
Kastamonu	Dr. Öğr. Üy. Yasemin Gültepe
Kırıkkale	Prof. Dr. Hasan Erbay
Kırklareli	Dr. Öğr. Üy. Bora Aslan
Kocaeli	Prof. Dr. Yaşar Becerikli
Kocaeli	Prof. Dr. Nevcihan Duru
Koç	Prof. Dr. Attila Gürsoy
Kütahya Dumlupınar	Doç. Dr Doğan Aydın
Maltepe	Dr. Öğr. Üy. Ceren Çubukçu
Maltepe	Prof. Dr. Asaf Vural
Manisa Celâl Bayar	Dr. Öğr. Üy. Tuğba Özacar Öztürk
Marmara	Prof. Dr. Haluk Rahmi Topçuoğlu
Mef	Doç. Dr. Şuayb S. Arslan
Mehmet Akif Ersoy	Dr. Öğr. Üy. Ali Hakan Işık
Mersin	Prof. Dr. Hamza Erol
Muğla Sıtkı Koçman	Doç.Dr. Bekir Taner Dinçer
Muğla Sıtkı Koçman	Doç. Dr. Hüseyin Gürüler
Munzur	Dr. Ögr. Üy. Ümit Can
Namık Kemal	Doç. Dr. Ercan Bulus
Necmettin Erbakan	Dr. Öğr. Üy. Mehmet Hacıbeyoğlu
Niğde Ömer Halisdemir	Dr. Öğr. Üy. Ahmet Şakir Dokuz
Nişantaşı	Dr. Öğr. Üy. Roy Küçükateş
Ondokuz Mayıs	Prof. Dr. Erdal Kılıç
Orta Doğu Teknik	Prof. Dr. Halit Oğuztüzün
Özyeğin	Doç. Dr. Hasan Sözer
Özyeğin	Prof. Dr. Fatih Uğurdağ
Pamukkale	Doç. Dr. Sezai Tokat
Sabancı	Prof. Dr. Albert Levi
Sakarya	Prof. Dr. Cemil Öz
Selçuk	Prof. Dr. Fatih Başçiftçi
Selçuk	Prof.Dr. Derviş Karaboğa
Siirt	Doç. Dr. Musa Ataş
Sivas Cumhuriyet	Doç. Dr. Hidayet Yakçı
Süleyman Demirel	Dr. Öğr. Üy. Arif Koyun
Süleyman Demirel	Prof. Dr. Tuncay Yiğit
Ted	Doç. Dr. Gökçe Nur Yılmaz
Ted	Doç. Dr. Tansel Dökeroğlu
Tobb Ekonomi ve Tek.	Prof. Dr. Oğuz Ergin
Tobb Ekonomi ve Tek.	Doç. Dr. A. Murat Özbayoğlu
Tokat Gaziosmanpaşa	Dr. Öğr. Üy. Mahir Kaya
Toros	Dr. Öğr. Üy.M. Ali Aktaş
Trakya	Doç.Dr. M. Tolga Sakallı
Türk – Alman	Dr. Öğr. Üy. Burcu Yıldız
Türk Hava Kurumu	Dr. Öğr. Üy. Ender Sevinç
Üsküdar	Doç. Dr. Serhat Özekes
Van Yüzüncü Yıl	Doç. Dr. Rıdvan Saraçoğlu
Yalova	Dr. Öğr. Üy. Osman Hilmi Koçal
Yaşar	Prof. Dr. Şaban Eren
Yaşar	Doç. Dr. Ahmet Koltuksuz
Yeditepe	Prof. Dr. Gürhan Küçük
Yıldırım Beyazıt	Prof. Dr. Fatih Vehbi Çelebi
Yıldız Teknik	Prof. Dr. Nizamettin Aydın
Yıldız Teknik	Prof. Dr. Banu Diri
Yozgat Bozok	Doç. Dr. Orhan Er
Yüzüncü Yıl	Dr. Öğr. Üy. Recep Özdağ
Zonguldak Bülent Ecevit	Dr. Öğr. Üy.Okan Erkaymaz

Research of Cluster Analysis Methods for Group Solutions of the Pattern Recognition Problem

Lyailya Cherikbayeva
Al-Farabi Kazakh National University
Department of Computer Science
Almaty, Kazakhstan
cherikbayeva.lyailya@gmail.com

Aigerim Yerimbetova
Institute of Information and Computational Technologies CS MES RK
Satbayev University
Almaty, Kazakhstan
aigerian@mail.ru

Elmira Daiyrbayeva
Institute of Information and Computational Technologies CS MES RK
Almaty, Kazakhstan
nurbekkyzy_e@mail.ru

***Abstract* — This paper proposes the study of cluster analysis methods for solving the problem of pattern recognition, including group solution methods. The study selected methods for solving the problem of cluster analysis based on a group solution with incomplete training information, investigated and developed models of group solutions based on existing known algorithms. The novelty of the work consists in a combination of algorithms for collective cluster analysis and nuclear classification methods. Numerical experiments on test problems and a real hyperspectral image demonstrate the effectiveness of the proposed method, including in the presence of noisy data.**

Keywords* — *cluster analysis, pattern recognition, group solution, hyperspectral image

I. Introduction

In this paper one of the options for setting the pattern recognition problem is considered - the problem of semi-supervised classification. In this problem, only for a part of the objects of the original sample, class labels are known; it is required to classify either existing unlabeled objects, or to form a decision rule for recognizing new objects. There are quite a large number of approaches and algorithms for solving the problem of partially controlled classification [1]. Currently, heuristic algorithms of self-training (Self-training), probabilistic methods, transductive support vector machine (TSVM), as well as graph-theoretic algorithms (Graph Laplacian Regularization) [2-3] are widely used. This replacement has a number of reasons. First, it can be assumed that objects from a dense area (cluster) in the feature space are more likely to have common class labels, even if this area has a complex shape. From this point of view, such objects are more similar to each other than other points located at the same distance from each other, but from different clusters. Secondly, it is known that the averaged co-association matrix determines the semimetric in the observation space [4], which means that the frequencies of assigning pairs of objects to the same clusters can be considered as indicators of similarity between the corresponding points. In this case, the resulting matrix depends on the outputs of the clustering algorithms and is less dependent on random outliers than the usual similarity matrix. Involvement of a cluster ensemble makes it possible to increase the stability of the results of cluster analysis in the case of uncertainty in the data structure. Such uncertainty arises, for example, when the true number of clusters is unknown or uninformative, noisy features, complex data structures (for example, spiral or circular clusters) may exist.

Conceptually, an optimization model for the application of group solutions (cluster ensemble) is proposed in the following scheme. The basis of such a model is proposed in [5].

In Figure 1, through $S_1, S_2, \ldots, S_t$ - the input data is designated, which can be represented both as numerical values and in the form of images. $A=\{A_1, A_2, \ldots, A_t\}$-sets of classification algorithms working on different principles, m-number of algorithms. $R_1, R_2, \ldots, R_t$ are algorithms for group solutions (cluster ensemble). Each decision of these algorithms makes decisions based on the results of a set of algorithms A. Further, $F=\{F_1, F_2, \ldots, F_t\}$ denotes the quality (accuracy) criteria of algorithms of both the first level from A and the second level from z_c. The directions of the arrows in the model show the data stream being processed. As you can see from the figure, the optimization model is presented as a network model. You can get out of this model when you get a good result in the sense of the extreme value of the quality functionals.

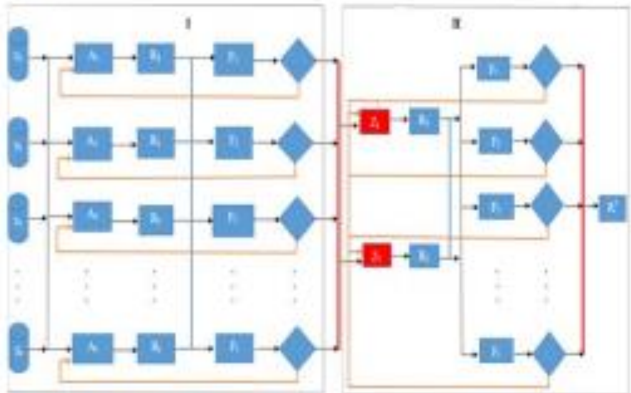

Fig. 1. Hyperspectral image of the "National Academy of Sciences of Kazakhstan" (RGB composite) (a) and tagged data (b)

In Figure 1, through $S_1, S_2, \ldots, S_t$ – the input data is designated, which can be represented both as numerical values and in the form of images. $A=\{A_1, A_2, \ldots, A_t\}$-sets of classification algorithms working on different principles, m-number of algorithms. $R_1, R_2, \ldots, R_t$ are algorithms for group solutions (cluster ensemble). Each decision of these algorithms makes decisions based on the results of a set of algorithms A. Further, $F=\{F_1, F_2, \ldots, F_t\}$ denotes the quality (accuracy) criteria of algorithms of both the first level from A and the second level from z_c. The directions of the arrows in the model show the data stream being processed. As you can see from the figure, the optimization model is presented as a network model. You can get out of this model when you get a good result in the sense of the extreme value of the quality functionals.

In [6], the proposed approach to the construction of a neural network is not based on the traditional approach based on minimizing the functional; rather, it is based on the theory of operators developed by Zhuravlev to solve recognition and classification problems. A distinctive feature of this network is the use of diagonal activation functions in the inner layers, which greatly simplifies intermediate calculations in the inner and outer loops.

This paper proposes a partially supervised classification method. The idea of the proposed method is to use a low-rank representation of the averaged coassociation matrix to reduce the required memory and computational costs.

Further in the work, a mathematical formulation of the problem is given, a brief overview of the methods of partially controlled classification is made. The proposed modification is described and its experimental study is carried out.

II. Mathematical Formulation of the Partially Controlled Classification Problem

Let there be a general set of recognition objects X and a finite set of class labels Y. All objects are described by signs. The feature of an object is understood as the display $f : X \to D_f$, where D_f is a set of feature values

Depending on D_f, signs are divided into types:

* Binary sign: $D_f = \{0,1\}$
* Quantitative sign: $D_f = R$
* Nominal sign: D_f - finite set
* Ordinal sign: D_f - a finite ordered set

With the specified attributes $f_1,...,f_n$, the vector $x = (f_1(\alpha),..,f_n(\alpha))$ is called the feature description of the object $\alpha \in X$. Next, we identify the object and its characteristic description. In the semi-supervised learning problem, a sample of $X_N = \{x_1,..,x_N\}$ objects from X is input. There are two types of objects in this selection:

* $X_c = \{x_1,..,x_k\}$ – marked up objects with the specified classes to which they belong: $Y_c = \{y_1,..,y_k\}$
* $X_u = \{x_{k+1},..,x_N\}$ – unmarked objects

In various versions of the problem statement, it is required either to conduct so-called inductive training – to build a classification algorithm $a : X \to Y$, which will, minimizing the probability of error, match classes to objects of their X_u, as well as new objects X_{test}, which were unavailable at the time of the algorithm construction, or it is required to conduct transduction training - to get class labels only for objects of X_u with minimal error. In this paper, the second variant of the problem statement is considered.

Next, an example will be given that shows how semi-controlled learning differs from the classification with a teacher

III. Matrix of Averaged Pairwise Differences

To construct a matrix of averaged pairwise differences, clustering of all available objects $X = \{x_1,..,x_N\}$ is carried out by a team of various cluster analysis algorithms $\mu_1,...,\mu_M$. Each algorithm gives L_m partitioning options, $m = 1,..,M$. Based on the results of the algorithms, a matrix H of averaged pairwise differences of objects from X is compiled.

$$h(i,j) = \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} h_{lm}(i,j), \tag{1}$$

where $i, j \in \{1,..,N\}$ are object numbers $(i \neq j), \alpha_m \geq 0$ are given weights such that $\sum_{m=1}^{M} \alpha_m = 1; h_{lm}(i,j) = 0$ if pair (i,j) belongs to different clusters in l partitioning variant obtained by algorithm μ_m and 1 if it belongs to the same cluster.

The weights α_m can be the same or, for example, can be adjusted in proportion to the clustering quality index. The optimal choice of weights is studied in.

Nuclear classification methods

To solve the classification problem, nuclear methods are widely used, which are based on the so-called. "Kernel trick". To demonstrate the essence of this "trick", consider the support vector machine (SVM) - the most popular kernel classification method. SVM is a binary classifier, although there are ways to improve it to implement multiclassification.

IV. The Proposed Method

The idea of the method is to construct a similarity matrix (1) of all objects from the input sample X. This matrix will be compiled by applying different clustering algorithms to X. The more often a pair of objects falls into the same cluster, the more similar we will consider them to each other. Two possible options for predicting classes of unlabeled objects X_u using a similarity matrix will be proposed. Further, the idea of the algorithm will be described in more detail. The following is true Theorem 1. Let $\mu_1,..,\mu_M$ be cluster analysis algorithms, each algorithm gives L_m partitioning options, $m = 1,..,M$, $h_{lm}(x,x') = 0$ if a pair of (x,x') objects belongs to different clusters in l partitioning options obtained by algorithm μ_m and 1 if it belongs to the same cluster. $\alpha_m \geq 0$ - given weights such that $\sum_{m=1}^{M} \alpha_m = 1$. Then the function $H(x,x') = \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} h_{lm}(x,x')$ satisfies the conditions of the Mercer theorem

Evidence. Obviously, the $H(x,x')$ function is symmetric. Let C_r^{lm} be the set of indices of objects belonging to the r cluster obtained by the m algorithm in

\

the l partitioning variant. Let us show that $H(x,x')$ is non-negative definite.

Take an arbitrary $z \in R^p$ and prove that $z^T Hz \geq 0$

$$z^T Hz = \sum_{i,j=1}^{p} \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} h_m(i,j) z_i z_j = \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} \sum_{i,j=1}^{p} h_m(i,j) z_i z_j =$$

$$\sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} \Big(\sum_{i,j \in C_1^m} z_i z_j + \ldots + \sum_{i,j \in C_{K_m}^m} z_i z_j \Big) = \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} \Big(\big(\sum_{i \in C_1^m} z_i\big)^2 + \ldots + \big(\sum_{i \in C_{K_m}^m} z_i\big)^2 \Big) \geq 0.$$

Thus, function $H(x,x')$ can be used as a kernel in kernel classification methods, in particular in support vector machine (SVM). Further, we propose two variants of the algorithm that implements the proposed approach: The

The CASVM algorithm:

Thus, the $H(x,x')$ function can be used as a kernel in nuclear classification methods, in particular, in the support vector machine (SVM) method. Next, two variants of the algorithm that implements the proposed approach are proposed:

Input: objects X_c with the specified classes Y_c and objects X_u, the number of clustering algorithms M, the number of clustering L_m by each algorithm $\mu_m, m=1,\ldots,M$.

Output: object classes X_u.

1) Perform clustering of objects $X_c \cup X_u$ by cluster analysis algorithms $\mu_1,\ldots,\mu_M$, obtaining L_m partitioning options from each algorithm $\mu_m, m=1,\ldots,M$.
2) Calculate the matrix H by $X_c \cup X_u$ using the formula (1).
3) Train the SVM on the marked-up data X_c, using the matrix H as the core.
4) Use SVM to predict classes for unmarked objects X_u.

End of Algorithm

Note that in the proposed algorithms it is not required to store the entire matrix H of size $N \times N$ in memory: it is enough to store the clustering matrix of size $N \times L$, where $L = \sum_{l=1}^{M} L_m$, in this case the matrix H can be calculated dynamically. In applied tasks, as a rule, $L \ll N$, for example, when working with image pixels.

V. Experimental research.

Numerical experiments were carried out with the developed algorithm, the purpose of which was to test its effectiveness in conditions of various data volumes and the presence of noise effects. The performance of the algorithm was determined using a real problem of hyperspectral image recognition.

Hyperspectral image analysis:

A typical RGB image contains three channels: saturation values for each of the three colors. In some cases, this is not enough to obtain complete information about the characteristics of the subject being shot. To obtain data on the properties of objects indistinguishable by the human eye, hyperspectral imaging is used.

For the experimental study of the developed algorithm, a hyperspectral image of the "National Academy of Sciences of Kazakhstan" with a size of 145 by 145 pixels, which contains 224 spectral channels in the range of 400-2500 nm, was used. Figure 2(a) shows the RGB composite of the image, and Figure 2 (b) shows the reference division of the image into 5 thematic classes. There are unmarked pixels in the image that are not assigned to any of the classes. These pixels were excluded from consideration during the analysis.

In the experiment, the marked part was 1% of the points selected randomly for each component. In order to reduce the effect of correlation of spectral channels, 5 new features were previously formed from the initial data using the principal component method.

The generated data table was fed to the input of the algorithm. The size of the ensemble was equal to $r = 10$, and different variants of the partition were obtained by varying the number of clusters in the interval $[1000, 1000 + r]$.

a) b)

Fig. 2. Hyperspectral image of the "National Academy of Sciences of Kazakhstan" (RGB composite) (a) and tagged data (b)

A comparison was made with the SVM support vector machine using the RBF kernel (the algorithm parameters recommended by default were selected).

In an experimental study of the algorithm, 1% of the pixels selected at random for each class constituted a labeled sample; the rest were included in the unlabeled part. To study the effect of noise on the performance of the algorithm, randomly selected r% of the spectral brightness values of pixels in different channels were subjected to a distorting effect: the corresponding value was replaced by a value chosen at random from the interval $[x(1-p), x(1+p)]$, where r, p are the specified parameters. A noisy data table containing the values of the spectral brightness of pixels for all channels was fed to the input of the CASVM algorithm, in which the K-means algorithm was chosen as the basic algorithm for constructing the cluster ensemble. Different variants of the partition were obtained by varying the number of clusters in the interval

$[30, 30+L]$, where L was equal to 120. In addition, to construct each variant of the solution, two channels were randomly selected. To speed up the operation of the K-means algorithm and obtain more diverse grouping options, the number of its iterations was limited to 1.

As you can see from the table, the CASVM algorithm is more robust to noise than the SVM algorithm.

TABLE I. ACCURACY OF GROUP DECISION AND SVM ALGORITHMS FOR DIFFERENT VALUES OF NOISE PARAMETERS 39000

Noise parameters r, p	**0%, 0**	**10%, 0.1**	**20%, 0.2**	**30%, 0.3**
CASVM	0.76	0.75	0.73	0.72
SVM	0.68	0.66	0.62	0.60

The paper considers one of the options for setting the pattern recognition problem - the task of semi-supervised learning. CASVM algorithms have been developed to solve this problem. They are based on a combination of collective cluster analysis and nuclear classification methods. An experimental study of the proposed algorithm on a hyperspectral image has been carried out. It is shown that the CASVM algorithm is more robust to noise than the standard support vector machine SVM.

ACKNOWLEDGMENT

This research has is funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP08857179).

REFERENCES

[1] X. Zhu, "Semi-supervised learning literature survey," Tech. Rep. (Department of Computer Science, Univ. of Wisconsin, Madison, 2008), no. 1530.

[2] G. Camps-Valls, T. Marsheva, D. Zhou, "Semi-supervised graph-based hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, 45(10), 2007, pp. 3044-3054.

[3] M. Wu, B. Scholkopf, "Transductive Classification via Local Learning Regularization," Artificial Intelligence and Statistics, 2007, pp. 628-635.

[4] D. Zhou, O. Bousquet, T. Lal, J. Weston, B. Scholkopf, "Learning with local and global consistency," In Advances in Neural Information Processing Systems. 16, 2003, pp. 321-328.

[5] E.N. Amirgaliyev, A.F. Mukhamedgaliev, "An optimization model of classification algorithms," USSR Computational mathematics and mathematical physics. 1985, 6, pp. 95-98

[6] A.E. Dyusembaev, M.V. Grishko, "Construction of a Correct Algorithm and Spatial Neural Network for Recognition Problems with Binary Data," Computational Mathematics and Mathematical Physics. Vol. 58, Issue 10, 2018, pp. 1673-1686