Improving classification of rock types in open pit mines using support vector method

Hristo Nikolov,

Remote sensing systems Dept., Space research and technology institute – BAS, Acad. G. Bonchev bl.1, 1113 Sofia, Bulgaria, hristo@stil.bas.bg

Abstract

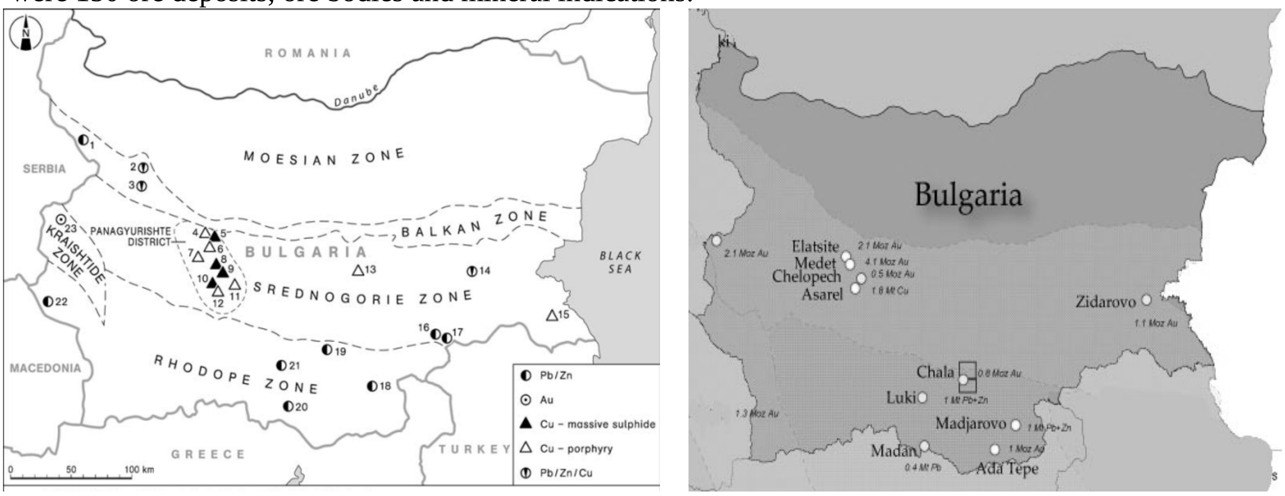
Support vector (SV) method for classification was fully developed and practically implemented in the late 90-ties of the last century. Initially formulated as two class separation procedure latter it evolved in robust multiclass classification technique. In this study SV based classification is applied to discrimination of several rock types found in and around the open pit mines of Asarel-Medet mining complex located in the Srednogorie copper-porphyry mining region. Main source of data used in this study come from instruments delivering multispectral data such as TM/OLI onboard Landsat satellite and will underline the possibilities offered by the MSI of the forthcoming Sentinel 2 mission. Ground truth data are from CORINE EU project which is used widely as reference for investigations at middle scale. Results obtained from this research proved that the classification method selected provides sound base for further investigations on this subject targeted at using more detailed spectral information for the objects for improve the classification accuracy (i.e. more narrow spectral bands) and also offering opportunity to experiment for unmixing the available data in order to overcome the limitations posed by the middle spatial resolution of the source data.

Абстракт

Методът за класификация на данни, използващ опорни вектори (ОВ) бе теоретично обоснован и приложен в практиката в края на 90-те години на миналия век. Този метод първоначално бе формулиран като задача за разделяне на данните на два класа впоследствие бе доразвит като надежден подход при класификация на данни в множество класове. В настоящата работа се изследват възможностите на процедура, използваща ОВ да бъде приложена към разделяне на различни типове скали от района на открит рудник Асарел-Медет от медно порфирния Средногорски руден район. Основен източник на данни са многоканалните апаратури TM/OLI на борда на сателитите Ландсат, но ще се покажат и възможностите на MSI, намиращ се на борда на Сентинел-2, който се очаква до бъде изведен в орбита до края на тази година. Данните за обучение и проверка на точността са от проект КОРИНЕ, който често се използва като референтен при средномащабни изследвания. Получените резултати показват предимствата на избрания метод и доказват, че може да бъде използван при бъдещи изследвания, където ще е налична по-детайлна спектрална информация за изследваните обекти.

Introduction

Srednogorian metallogenic zone (see Figure 1) is located in the central part of Bulgaria and belongs to the metallogenic belt Apuseni-Banat-Timok-Srednogorie, which houses one of the richest porphyry copper and copper-gold epithermal deposits in Europe (Strashimirov et al., 2002). In Srednogorie area, and in particular ore region of Panagyurishte, located 60-90 km east of Sofia, exploited ore deposits contain mainly Cu and Cu-Au-Mo. The total area of the said deposits is 600 square kilometers. In this area found and documented were 150 ore deposits, ore bodies and mineral indications.



a b

Figure 1. Map of the polymetal ores deposits – as surveyed (a) and as economically viable (b)

The main geological features that have controlled the spatial-temporal development of metallogenic processes in Panagyurishte ore region are determined by the characteristics and development of the Upper Cretaceous magmatic complexes (Popov et al., 2003).
The area selected for this research, namely the Asarel-Medet mining complex, has been studied in the early 60s and started production in 1964 of the 20^{th} century, at that time being the largest copper producing mine in Europe and third largest in the world at that time. At this time it the first attempt with starting open pit mining exploitation having copper ore with a percentage lower than 0.4%. The practice for processing such ores later has been widely used around the world. Experience of research and construction of mining complexes accumulated in the development of the MPC “Medet” was used in the construction of similar projects in Bulgaria in the 60s and 70s of the same century. The Assarel deposit is of copper porphyry type and has hydrothermal formation and is located some 12km away from the Medet deposit. The main ore minerals are chalcosine, pyrite, chalcopyrite, bornite and coveline, as the average copper grade is 0,45%. Annually the mining complex comprising the open mine and the ore flotation plant processes 13mln tones of ore delivering copper concentrate and cathode copper. The mining site consists of an open pit mine, main and auxiliary dump sites, a tailings pond, and a processing plant.

Method and data

Support Vector Machines (SVM), also known as Large Margin Separator, are a class of learning algorithms initially defined for qualitative binary discrimination of the values of a single variable. After that they were generalized to the provision of a quantitative discrimination of the variable. In the case of discriminating a binary variable the selection of the SVMs is based on finding the *optimal margin* of the hyperplane who properly separates the two possible classes based solely on the observed data keeping them as far as possible. The principle is to find a classifier, or a discrimination function, whose generalization ability (equal to the quality of prediction) is the highest possible. The central principle of the SVM approach is to integrate precise estimation control of the complexity of the model that is to say the number of parameters that is associated in this case with the number of support vectors. In the general case, the optimal discrimination boundary is nonlinear. Within the SVM, taking account of non-linearities in the model is effected by the introduction of nonlinear kernels whose use does not fundamentally alter the nature of the SVM. This is the kernel function K which maps each pair of observations (x, x) measuring their "reciprocal influence" calculated through their correlation or their distance. Typical examples of the kernels are the polynomial kernels k(x, x’) = (C + <x,x’>).

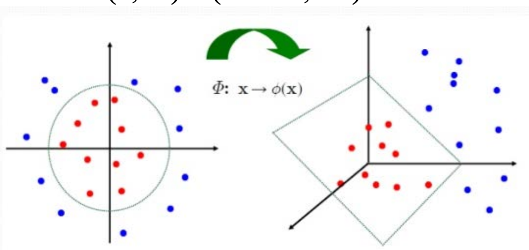


Figure 1 Transformation for linear non-separable case and transformation via kernels in high dimensional space

Most of the data from experiments are as rule noisy and form such areas in the feature space that can not be separated linearly thus prohibiting good generalization. This is the reason nonlinear methods are needed to find dependencies that allow classification on unseen data i.e. prediction. One way to use nonlinear functions in the current feature space are the positive definite kernels that are equal to a dot product in other and as rule high-dimensional feature space. An attractive aspect of the kernels is that there is no need to make computations in this high-dimensional feature space if we use kernel evaluations for all formulations. A clear message should be that by implementing kernels instead of dot products provides suitable method for using the benefits of nonlinear discrimination by means of linear algorithm.
The data used in this research are provided by USGS at no cost in the framework of the Landsat data distribution initiative and have been processed by LEDAPS up to level L1G. They consist of DNs in six multispectral bands, excluding the thermal one, without atmospheric corrections from October scenes from years 1999 and 2006 with dimensions 519 columns and 500 rows. As ground truth data for the classes of interest vector layers from CORINE 2000 project have been used (Stoimenov, 2004). For the purposes set for this study the information classes were formed by considering CLC classes 112, 121, 131 and 132 while

all other classes present in the data were set to non-existing one. Data processing has been carried out by QGIS software.

Results

In this research experimented was with three kernels for the SVM classification procedure – linear, RBF and polynomial. Parameters of the said kernels used during the learning phase are set in Table 1.

year 1999				
kernel type	test samples	verification samples	parameter *C*	kappa index
RBF	500	1000	2	0.653952
	800	1500	2	0.654801
	800	1500	3	0.652561
	800	1500	5	*0.656619*
	500	1000	5	0.655202
linear	500	1000	2	*0.559303*
	500	1000	5	0.560038
	500	1000	10	0.560036
	800	2000	5	0.571523
	1200	2500	5	0.571523
year 2006				
linear	500	1000	2	0.563492
	500	1000	5	0.563489
	800	1500	5	0.650012

Table 1 Parameters for the kernels applied under learning phase

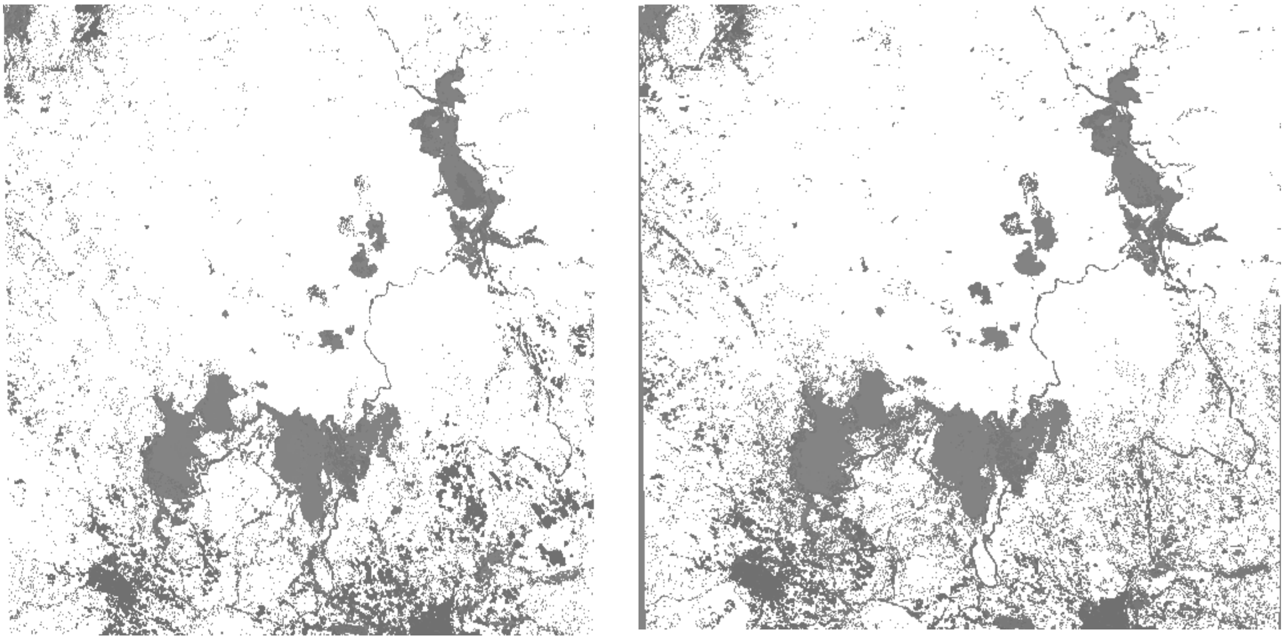
It is clearly visible from the data reported that at the learning stage RBF kernel outperforms the linear regardless of the number of samples used. In both cases the increase of the number of test/verification samples does not results in proportional increase of the overall accuracy (see *kappa* index). The same is valid for the penalty parameter *C* whose increase does not lead to large changes of kappa.

Other experiment we made was to change the type of raster data used. Instead of manipulating the DNs' the reflectance values for the land cover objects was used and the result is reported as follows (see Table 2). From the results it can be drawn the conclusion that for the data used the overall quality index kappa does not change significantly if using reflectance but not DNs.

kernel type	test samples	verification samples	parameter *C*	kappa index
1999				
RBF	800	1500	5	0.562142
2006				
RBF	800	1500	5	0.566007

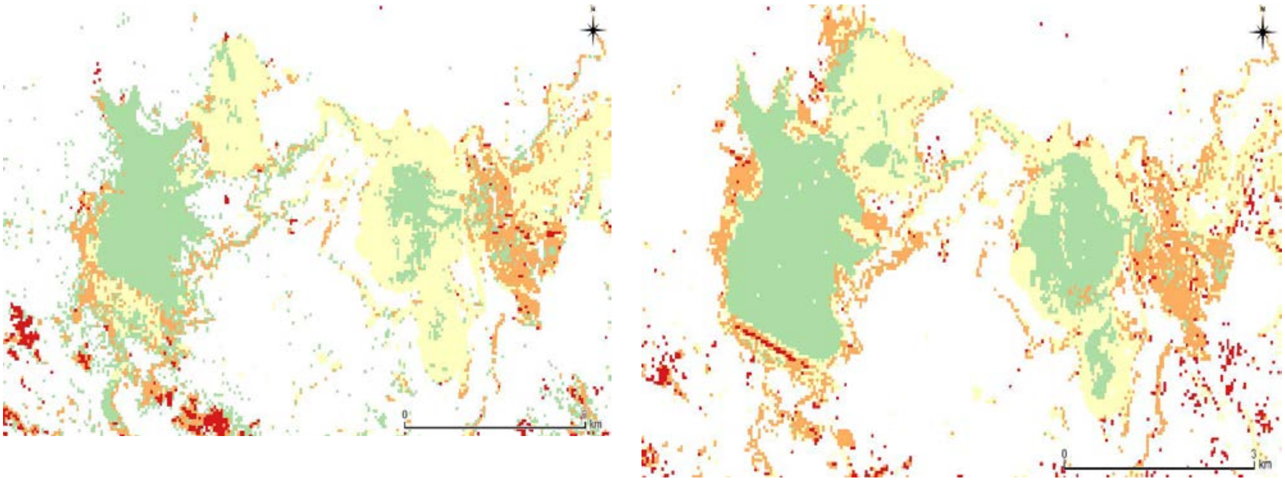
Table 2 Results from the learning phase using reflectance

On the next figure (Figure 2) shown are the resulting files after classification with the models obtained with parameters corresponding to highest and lowest kappa values. Since the classification is pixel-based the low reported accuracy can be due to several reasons just to mention the medium spatial resolution, the effect of terrain, etc. which can not be easily overcome since more data are needed. The second experiment with the reflectance values also did not provide significant change in kappa so one can conclude that the increase of the dynamic range (DNs have 256 values and the reflectance is in the range from 0 to 16000) is not the factor that has direct impact on improvement of the accuracy. On Figure 3 using pseudocoloring shown are the images produced classifying the reflectance values. On those it can be seen the actual increase of the areas occupied by the mine and dump areas (green and light yellow colors).



linear t500v1000c2 RBF t800v1500c5

Figure 2 Classified images in grayscale for the Asarel-Medet region



RBF t800v1500c5 for 1999 (ETM+) RBF t800v1500c5 for 2006 (TM)

Figure 3 Classification of images with reflectance values

Conclusions

From the results reported it can concluded that SVM method is reliable tool in classification of multispectral images. It provides rigid base in obtaining classification maps from remotely sensed data resulting from semiautomatic procedure.

References:

1. OMENTIN project, 2001. Ore mines, processing and waste management in Europe, Technical report, Budapest, http://publikacio.uni-miskolc.hu/data/ME-PUB-20785/One%20mines.pdf
2. Popov, P., S. Strashimirov, K. Popov, R. Petrunov, M. Kanazirski, D. Tzonev, 2003. Main features in geology and metallogeny of the Panagyurishte ore region. *Annual of University of Mining and Geology "St. Ivan Rilski"*, 46 (Part I: Geology and Geophysics), 119-125.
3. Strashimirov, S., R. Petrunov, M. Kanazirski, 2002. Porphyry-copper mineralisation in the Central Srednogorie zone, Bulgaria. Mineralium Deposita, 37, 587–598.
4. Stoimenov A., 2004. CORINE Land Cover 2000 - Bulgaria Project. Final Report, BEEA, Ministry of environment and waters