Publishing House
AKAPIT

COMPUTER METHODS IN MATERIALS SCIENCE
Informatyka w Technologii Materiałów

Vol. 13, 2013, No. 2

SELECTION OF SIGNIFICANT VISUAL FEATURES FOR CLASSIFICATION OF SCALES USING BOOSTING TREES MODEL

SZYMON LECHWAR

ArcelorMittal Poland, Hot Rolling Mill in Kraków, ul. Ujastek 1, 30-969 Kraków
**Corresponding author: szymon.lechwar@arcelormittal.com*

Abstract

The subject of this paper is to design and implement an efficient model for various kinds of scales recognition at the Hot Rolling Mill (HRM) in Kraków. Subsequently, the model and its most important variables can be used to describe and distinguish different kinds of scales. At the moment an extensive knowledge regarding the reasons of scale occurrence is gathered. Nevertheless, the real challenges nowadays seem to be measuring techniques of those phenomena, as well as reliable online classification.

This paper describes the basics of automatic surface inspection system (ASIS) which was used as a source of entry data, as well as the method of interpretation of the data obtained from this system. The ASIS provided numerous features describing single image, which was considered as a defect. The objective of this paper was to supply information regarding the most important visual attributes, which will be subsequently used in building reliable classifier for scale recognition. It was done by use of data mining techniques. The result was a set of measurement data, stored in online production database.

However, some kinds of scales could not be recognized efficiently. The reason behind that was the lack of unique features, which could distinguish them from the other defects. This problem will be solved in following studies by creating offline post processing rules.

Key words: automatic surface inspection system, boosting trees, data mining, hot rolling mil

1. INTRODUCTION

In today's industry, global competition and rising customer’s requirements are becoming increasingly important in production of high quality products. At the same time, each plant puts strong emphasis on the automation of its process and the maximum costs reduction. Combination of these factors often proves to be very difficult or even impossible to achieve with the use of common production methods.

Steel industry is no exception to that rule. Direct customers and subsequent treatment processes (ex. cold rolling process) requires production of higher quality steel while reducing costs. One way to achieve this goal is the application of automatic surface inspection of rolled sheets (ASIS - Automatic Surface Inspection System). The purpose of the system is to take pictures of produced material, to detect local variations in contrast on its surface and to classify individual irregularities. Each picture taken by the system is digital and converted into grey scale pixels. In this way a map can be obtain, which supplies information regarding defective material in terms of various defects and pseudo defects. This type of system brings a significant reduction of visual inspections performed by human inspector.

To make it possible to build reliable classifier of surface defects, ASIS needs to be taught. Person, who is an expert in certain classification, should create sets of defects that will be used to “teach” software provided by the manufacturer. In an ordi-

 ISSN 1641-8581

nary approach, the end user (expert in the field) makes selection of images that he believes belong to specific classes of defects and arrange them in the program supplied by the manufacturer. On this basis, the software creates models, examines the characteristics of images and selects the rules by which it will be possible to classify newly emerging images. This approach cuts user's knowledge about visual characteristics of the images. In this approach it is impossible to distinguish specific types of defects using results given by the ASIS. Furthermore, user cannot build additional rules in third party software to assist classification of similar defect classes.

In this study it was decided to deal with this issue in more detailed manner. The aim of the work was to find visual characteristics of defects that best serve to build a model (Webb, 2002; Bakker et al., 2006). The research was concentrated on scale defects produced at hot rolling mill in Kraków. It was decided to manually select a set of reference defects, analyse visual features of each defect class, which has an influence on construction of a scale classification model and decides which features could be used in future work to build a reliable scale classifier.

2. DISTRIBUTION OF SCALES AT HOT ROLLING MILL IN KRAKÓW

ASIS is monitoring whole coils production at the hot rolling mill, returning as a result map of irregularities in the contrast detected on the surface of hot strip. The number of possible defects that might be produced during production varies depending on steel grade, strip thickness and technological mill settings. At most, about 30 different real defects could occur on the hot strip. Therefore, ASIS is trained to detect and classify all of them. This study focused on scale defects. Based on ArcelorMittal internal defects catalogue (Breitschuh et al., 2007) and expert knowledge it was decided to select and distinguish 10 scale classes.

Different defects occurring at hot rolling mill in Kraków production line were sorted in order to conduct the study. Defects images were taken by ASIS from production line. From the pool of 26000 candidate’s images some of them were isolated as real scale defects. This was followed by manual classification of images based on expert knowledge and reference materials (Melfo et al., 2006; Sun et al., 2003; Sun et al., 2004). As a result, set of 3300 scale defects were gathered and hand-classified. Not all scale classes will be presented in this paper. Classes will be treated as a reference data based on which data analysis will be carried out.

3. SELECTION OF THE MOST RELEVANT VISUAL FEATURES OF THE SCALE DEFECTS USING DATA MINING METHODS

Creation of a rich set of reference data provided opportunity to explain visual features of scales in detail. Images, taken by ASIS, were analysed by the manufacturer software to find local variations in contrast. Such areas, called regions of interest (ROI), received a number of features that describes their visual characteristics in numeric manner. Features, which provided information regarding classification of the currently implemented classifier were removed, as they supply unnecessary data at this stage of study. In the end, raw data consisted of 744 variables, which was passed to the further analyses.

3.1. Preliminary data analysis

The first step in data mining analysis (Hand et al.,2001; Han & Kamber, 2006) focused on data preparation and cleaning, which had been significantly reduced due to correctness of reference data - manual classification. It does not contain any missing fields or repeated observations. Variables, with variance below 10^{-10}, were removed as they did not carry any valuable information. It was assumed that reference data does not contain any unusual values or outliers. Any transitions and transformations were not carried out during data cleaning stage.

In order to seek the most important features describing categorical variable "type of scale" input data was analysed by the two data mining (Statsoft, 2006) modules:

- Decision trees C & RT (Classification and Regression Trees)
- Variables selection and analysis of the causes, which finds the best predictors for each dependent variable. Interactions between predictors are not taken into account.

In total, 145 variables were selected for further analysis.

3.2. Choice of the best set of features describing variable "type of scale"

It was decided to build Boosting Trees model in order to supply information regarding the most important features, that will be used in creation of reliable classifier. Besides classification, model defined the most relevant attributes that are the case of this study. Input data was divided into learning sample and the validation sample with 80% to 20% ratio. In order to find the most efficient model, different parameterisations were analysed.

First step covered testing of the model depending on number of variables. Two variants were tested, i.e. without the use of redundant variables (those with correlation between variables exceeding 0.8) and with redundant variables. First case (figure 1) shows that the best model can be obtained for 21 variables.

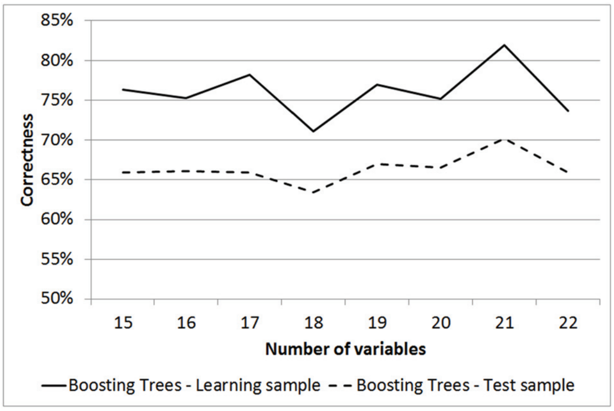

Fig. 1. Correctness of the Boosting Trees model depending on number of variables – without the use of redundant variables.

The second case (figure 2) gave much more promising results at 33 variables, with 94.26% of correctness on the learning sample and 78.65% of correctness on the test sample.

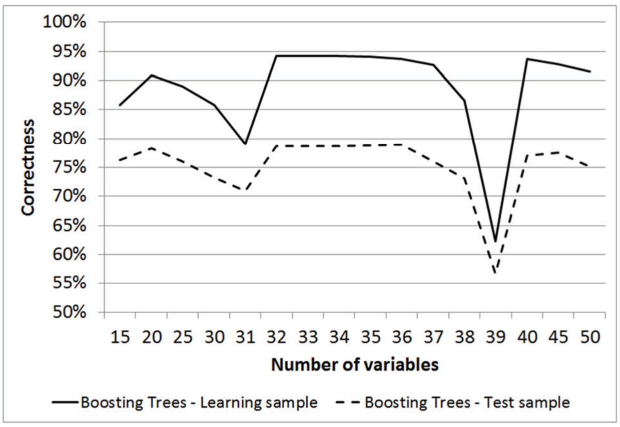

Fig. 2. Correctness of the Boosting Trees model depending on number of variables – with the use of redundant variables.

Parameterisation was continued with the use 33 variables as they gave the best model at this point of study. The change of a priori probability, maximum number of nodes, maximum number of levels and minimum cardinality of descendant did not change efficiency of the model. Only one parameter, minimum cardinality of node at 123, gave better model (figure 3).

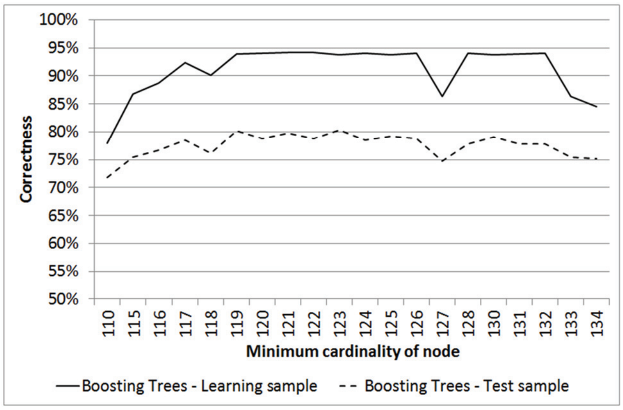

Fig. 3. Correctness of the Boosting Trees model depending on number of minimum cardinality of node.

Finally, the best model reached 93.84% of correctness on the learning sample and 80.34% of correctness on the test sample. The rest of the parameters remained at default values. Table 1 contains gathered parameters of the model.

Table 1. Parameters of best Boosting Trees model.

Number of variables	33
Erase of redundancy data	No
Fast variables selection	No
Minimum cardinality of node	123
Minimum cardinality of node of descendant	1
Maximum number of levels	10
Maximum number of nodes	13
A priori probability	Equal

The benefit from construction of this model was selection of 33 variables, which describe characteristics of the scale defects. Table 2 presents collected variables, along with their significance and short description.

First two items from the table 2 describe decomposition of scale on the strip. It is most often associated with abnormal work of descalers, because they remove only a portion of scale covering the slab. Nevertheless, in this paper, proper work of descalers is only the cause of defects that should be eliminated. The core of the work was to find visual features that could be used in creation of efficient scale classifier. Therefore, position of defects throughout the strip will be ignored.

Table 2. *Significance of visual features.*

Original Name	Unit	Description	Significance [%]
_Pos_CD_mm	mm	Position of defect, cross direction	100,00%
_Pos_RCD	mm	Position of defect, reverse cross direction	84,61%
S_V_MAXSEGD	px	Gray Artefact Scaled [px]	76,13%
AGUV_G_REL		Ratio of H/V Mean of Absolute Gradient	75,84%
AGUV_G_DIF		Relative H/V Diff of Abs Gradient Mean	75,84%
V_MAX_SEG_D	px	Maximum Vertical Dark Segment Length	75,35%
DEV_G_REL		Ratio of H/V Deviation of Gradient	75,32%
DEV_G_DIF		Relative H/V Diff of Gradient Deviation	75,32%
BSEG_FACE	%	Percentage of Real Face Pixels	74,22%
BSEG_SPOT	%	Percentage of Tiny Spot Pixels	73,45%
S_V_MAX_SEG	px	Gray Artefact Scaled V_MAX_SEG	72,25%
R_MAX_SEG_D		Ratio H/V Max Dark Segment Length	71,95%
D_MAX_SEG_D		Difference H/V Max Dark Segment Length	71,95%
BMAP_FACE	%	Percentage of Real Face Pixels	71,73%
V_MAX_SEG	px	Maximum Vertical Segment Length	71,71%
MEAN_AGV_Q		Ratio of Vertical Abs Gradient to Range	71,09%
BMAP_ISOL	%	Percentage of Isolated Pixel Masks	69,74%
D_MAX_SEG		Difference H/V Max Segment Length	66,52%
RC_H_M_D		RowCol X-Dir Mean of Deviation	64,32%
S_RC_H_M_D		Gray Scaled RC_H_M_D	64,03%
RANG_G_REL		Ratio of H/V Range of Gradient	63,24%
RANG_G_DIF		Relative H/V Diff of Gradient Range	63,24%
WLT_L0HTS0YX		Scale L0, Filter HT, Feature S0, Direction YX	62,09%
WLT_L0HTM1XY	gv	Scale L0, Filter HT, Feature M1, Direction XY	61,53%
S_H_AVGSEGD	px	Gray Artefact Scaled H_AVG_SEG_D	60,73%
WLT_L0HTS1YX		Scale L0, Filter HT, Feature S1, Direction YX	60,60%
SYM_GVU		Anti-Diagonal Grayvalue Symmetry	60,31%
SYM_GUV		Diagonal Grayvalue Symmetry	60,27%
H_AVG_SEG	px	Average Horizontal Segment Length	60,05%
H_AVG_SEG_D	px	Average Horizontal Dark Segment Length	59,84%
SYM_GU		Horizontal Grayvalue Symmetry	59,78%
S_TEX_00		Gray Scaled TEX_00	59,58%
S_RC_H_M_M		Gray Scaled RC_H_M_M	59,58%

4. RESULTS - DISTINCTION OF SCALES THROUGHOUT VISUAL FEATURES OF ITS IMAGES

Most relevant features, that had been isolated from a wide variety of attributes given by ASIS, were used to distinguish scale classes. Subsequently, these features along with sufficient logic will support automatic classifier (build by default within supplier software). It is possible to create additional classification rules both in C++ language and T-SQL stored procedures (ASIS database).

One example of feature, that, together with necessary logic, could be implemented as classification support is "horizontal to vertical difference of gradient range". It describes the numerical differences between horizontal and vertical gradient ranges (in grey scale) of the defect. Figure 4 shows decomposition of the feature for line scale, which originates from the first stand of finishing train at the end of rolling campaign.

Figure 5 shows decomposition of the feature for single strip scale - defect formed due to malfunction of the descalers.

The feature could support final classification decision between these two classes. Although, straightforward use of the attribute to classify one of these scales, whenever it lies between -0.3 and 0.5 or -0.3 and 0.1, is not possible.

The other type of visual feature, which - in the opposite to previous one - could be used in direct classification, is "maximum difference between horizontal and vertical dark segment length". The attribute inform the classification system what is the biggest difference between the horizontal and vertical lengths among all dark segments in the defect (segments composed of dark pixels). Figure 6 shows decomposition of the feature for line scale. Figure 7 shows decomposition for "V" scale, which is defect originating from a finishing train. Its shape resembles "V" letter.

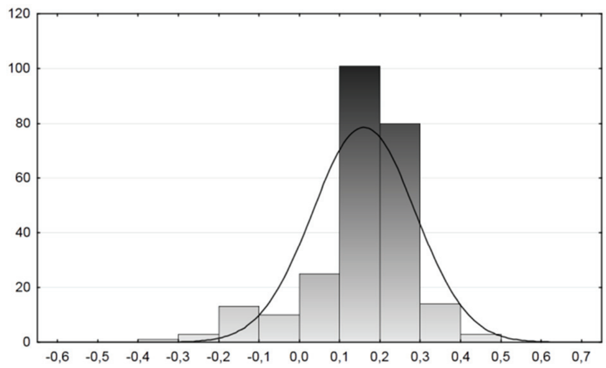

Fig. 4. *Horizontal to vertical difference of gradient range – Line scale.*

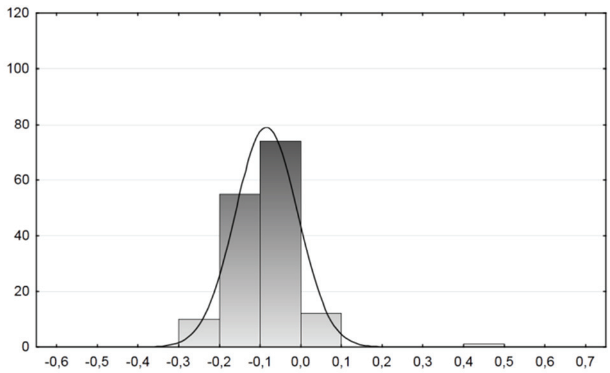

Fig. 5. *Horizontal to vertical difference of gradient range – Single strip scale.*

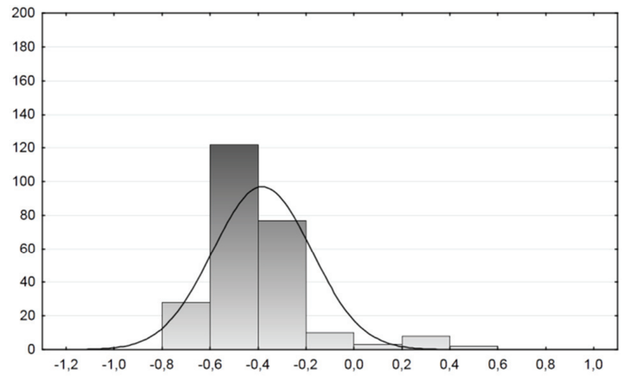

Fig. 6. Maximum difference between horizontal and vertical dark segment length – Line scale.

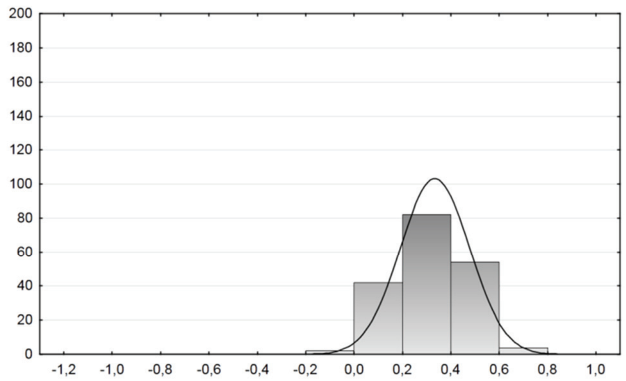

Fig. 7. Difference between horizontal / vertical of maximum dark segment length – "V" scale.

In this case a threshold could be set at 0.2. This kind of rule might be used in supporting logic to distinguish those two scale defects.

5. CONCLUSIONS AND PERSPECTIVES

In the paper scale defects occurring at hot rolling mill in Kraków were divided into unique classes. Second part of the paper describes process of Boosting Trees model creation for the scale classification. Along with the model, 33 most relevant attributes for the model were selected. These numerical visual features were used to describe each scale class by decomposition of its values. Subsequently, the features can be used in building reliable classifiers for scale recognition. Only part of the scale visual features, which could be used in classifier building, were presented in this paper.

The next step in the study will be scale classifier implementation with the use of manufacturer software. It will depend on manual selection of the scale defects and their assignment to proper class. Nevertheless, study assumes that creation of the best possible classifier could be hard to obtain using manufacturer software. To improve its classification decision some additional rules have to be created. The rules can be written in C++ programming language and Transact SQL stored procedures. They will be used in the next classification process, called post-classification. This step will be assisted by knowledge described in this paper.

Acknowledgements. The author would like to express his gratitude to Mr Witold Dymek, who supported this work at Hot Rolling Mill in Kraków and Professor Maciej Pietrzyk and Dr Łukasz Rauch for their enthusiastic encouragement and useful critiques of this research.

REFERENCES

Bakker, A., Hoyles, C., Kent, P., Noss, R., 2006, Improving Work Processes by Making the Invisible Visible, *Journal of Education and Work*, 19, 343-361.

Breitschuh, W., Crowley, G., Dallemagne, L., Deléglise, A., Diaz-Alvarez, J., Valcarcel, J.M., Di Fant, M., Fiori, S., Hemmerlin, M., Koschack, U. Schroyens K., 2007, ArcelorMittal internal defects catalogue.

Han, J., Kamber, M., 2006, *Data mining: concepts and techniques*, University of Illinois, Urbana-Champaign.

Hand, D., Mannila, H., Smyth, P., 2001, *Principles of data mining*, The MIT Press, Cambridge.

Melfo, W., Dippenaar, R., Reid, M., 2006, In-Situ Study of Scale Formation under Finishing-Mill Operating Conditions, *Proc. AISTech 2006*, Association for Iron & Steel Technology, Cleveland, Ohio, II, 25-35.

Sun, W., Tieu, A.K., Jiang, Z., Lu, C., 2004, High temperature oxide scale characteristics of low carbon steel in hot rolling, *Journal of Materials Processing Technology*, 155-156, 1307-1312.

Sun, W., Tieu, A.K., Jiang, Z., Lu, C., Zhu, H., 2003, Surface characteristics of oxide scale in hot strip rolling, *Journal of Materials Processing Technology*, 140, 76-83.

Statsoft, 2006, Elektroniczy Podręcznik Statystyki PL, Kraków (in Polish).

Webb, A., 2002, *Statistical Pattern Recognition* (2nd Edition), John Wiley & Sons, Ltd., Chichester, England.

DOBÓR NAJISTOTNIEJSZYCH ASPEKTÓW WIZYJNYCH ZJAWISKA WYSTĘPOWANIA ZGORZELINY Z UŻYCIEM WZMACNIANYCH DRZEW KLASYFIKACYJNYCH

Streszczenie

Przedmiotem badań jest zaprojektowanie i wdrożenie skutecznego modelu klasyfikującego wszystkie rodzaje zgorzeliny występujące w walcowni gorącej w Krakowie. Model oraz jego kluczowe zmienne mogą opisać i rozróżnić poszczególne typy zgorzeliny. W ramach pracy postanowiono zająć się techniką pomiarową oraz wykorzystaniem danych pomiarowych do budowy optymalnego klasyfikatora wad tego zjawiska. Danych pomiarowych, dotyczących aspektów wizyjnych pojedynczych obszarów pasma, dostarczył automatyczny system kontroli powierzchni (ASIS), którego podstawy działania przedstawiono w pracy. Otrzymane dane pomiarowe zostały przeanalizowane z wykorzystaniem metod selekcji cech, a wybrane cechy posłu-

żyły do budowy klasyfikatora dla wad powierzchni typu zgorzelina. Klasyfikator zaimplementowany został z wykorzystaniem metod eksploracji danych, które, wraz z otrzymanymi wynikami, zostały szczegółowo opisane w niniejszym artykule.

Received: October 27, 2012
Received in a revised form: November 11, 2012
Accepted: November 16, 2012