Proceedings of the Federated Conference on Computer Science and Information Systems pp. 199–202
DOI: 10.15439/2017F382
ISSN 2300-5963 ACSIS, Vol. 11

Feature Selection Methods Applied to Severe Brain Damages Data

Wiesław Paja
University of Rzeszów, Pigonia Str. 1, 35-310
Rzeszów, Poland
Email: wpaja@ur.edu.pl

Krzysztof Pancerz
University of Rzeszów, Pigonia Str. 1, 35-310
Rzeszów, Poland
Email: kpancerz@ur.edu.pl

***Abstract*—Brain injuries seem to be one of the most widespread diseases. Hence, the main goal of our research was to investigate feature importance in the severe brain damages dataset according to the Glasgow Outcome Scale. This scale is recognized as one of several measures used to evaluate patients' functional ability as well as their conditions after applying brain damage therapy. The current approach is focused on an identification of a relevant subset of features with a similar influence on quality of classification models. According to the results gathered, about 12 from 42 descriptive features could be treated as important without the decrease of classification results.**

I. Introduction

ACCORDING to many sources [1-8], brain damages seem to be one of the most widespread civilization illnesses, occurring at different levels of severity, usually described by means of various measures (scales) [9]. It is important to say that there is no single outcome measure which can describe or predict all dimensions of recovery and disability after acute stroke. Several scales have proven reliability and validity in stroke trials [10], including the *National Institutes of Health Stroke Scale* (NIHSS), the *modified Rankin Scale* [8] (mRS, patient's functional agility), the *Barthel Index* (BI), the *Glasgow Outcome Scale* (GOS, assessment of patient's condition after therapy), the *Extended Glasgow Outcome Scale* (GOS-E) [11] and the *Stroke Impact Scale* (SIS). In this domain, several scales have been applied in stroke trials to derive a global statistic for better recognition of the effect of acute interventions, although this composite statistic is not clinically tenable. In practical applications, the NIHSS is efficient for early prognostication and serial assessment. In turn, the BI index is helpful for rehabilitation planning. The mRS and GOS parameters specify cumulative values of outcome and they are appropriate for clinicians and patients considering early intervention, while the SIS scale was created to evaluate the patient's perspective on the effect of stroke. However, the GOS-E extends five original GOS scale categories to eight. It is made to apply wide categories that are insensitive to change and to deal with difficulties with reliability due to lack of a structured interview format. Familiarity with these different scales could support clinicians' interpretation of stroke research and improve their clinical diagnosis.

The Glasgow Outcome Scale (GOS) is a scale in which patients with brain injuries, such as cerebral traumas, can be divided into groups that allow standardized descriptions of the objective degree of recovery. This scale was very often applied before other scales were introduced. After the improvement of disability recognition, the GOS has been replaced by the *Disability Rating Scale* (DRS) [12]. However, it is still cited occasionally in the literature, often in research investigating early acute medical predictors of gross outcome. In these type of approaches, five classes of the original scale are defined: *dead*, *vegetative*, *severely disabled*, *moderately disabled*, and *good recovery*.

II. Methods and Tools

A. Input data

An investigated data set contains the *Glasgow Outcome Scale* characterization for 161 anonymous patients. For a description of each object, 42 features were defined [7]. Objects were assigned into five different categories, according to the *Glasgow Outcome Scale*: **1** means *death*, **2** means *persistent vegetative state*, **3** means *severe disability*, **4** means *moderate disability* and **5** means *good recovery*.

Additionally, features are divided into six groups according to their context:

- ***A1-A9*** – General data about patient.
- ***B1-B14*** – Patient's specific features.
- ***C1-C7*** – Condition of health.
- ***D1-D3*** – Disorders.
- ***E1-E6*** – Treatment.
- ***F1-F3*** – Rehabilitation.

Detailed information about features and their values is presented in Table I.

B. Methods

The main focus during the research is to investigate presented data in the context of finding relevant features inside data that provide similar information after reduction of a feature space [13]. For this purpose, four different approaches for ranking measures and algorithms were

applied. Classification quality was computed before and after an application of a feature selection procedure. Firstly, a simple filter method using a ranking measure in a form of *Information gain* was applied to calculate ranking values for each feature [14]. In this step, the dataset was extended by adding contrast variables to define the threshold between informative and non-informative features [15]. It means that each original feature was duplicated and its values were randomly permuted among all objects. In this way, a set of non-informative, by design shadow, features was added to the set of original features. The features, selected as important rather than random, were treated further as an important feature subset. Then, the classification process using five learning algorithms (*CN2 rules, Classification Tree, kNN, SVM, RandomForest*) was executed. After that, to extract a relevant feature subset, two other algorithms were applied [16]. The first one is based on the frequency of presence of features contained in the rule model that is created on the basis of the original dataset and additionally takes into account the quality of rules in which an analyzed feature occurs. Thus, the importance value of the i^{th} attribute (*DRQualityImp*) could be presented as:

$$DRQualityImp_{A_i} = \sum_{j=1}^{n} Q_{R_j} \cdot Pres(A_i)$$

where n is a number of rules in the learning model, Q_{Rj} is the classification quality of the rule R_j and $Pres(A_i)$ describes the presence of the i^{th} attribute, usually either 1 (feature occurred) or 0 (feature did not occur). In turn, the quality of a given rule R_j is defined as:

$$Q_{R_j} = \frac{E_{corr}}{E_{corr} + E_{incorr}}$$

where E_{corr} is a number of correctly matched learning objects by the j^{th} rule and E_{incorr} is a number of incorrectly classified objects by this rule. In turn, the second algorithm (*DTLevelImp*) is based on the presence of a feature in the decision tree nodes generated from the original dataset and also takes into consideration the product of a weight W_j assigned to a given level j of the tree and the number *Inst(node)* of cases classified in a given node at this level in which the feature A_i occurs. Thus, the *DTLevelImp* of the i^{th} attribute can be presented as:

$$DTLevelImp_{A_i} = \sum_{j=1}^{l} \sum_{node=1}^{x} W_j \cdot Inst(node) \cdot Pres(A_i)$$

where l is a number of levels inside the model, x is a number of nodes inside at a given level and $Pres(A_i)$ denotes the presence of the i^{th} attribute, usually either 1 (feature occurred) or 0 (feature did not occur).
In turn, a weight W of the level j is defined as:

$$W_j = \begin{cases} 1 & \text{if } j = 1, j \in N, \\ \frac{w_{j-1}}{2} & \text{if } 1 \le j \le l. \end{cases}$$

The last approach to feature selection is based on rough set theory. In rough set theory, feature selection refers to finding the so-called decision reducts in a dataset (called a decision table). In general, a decision reduct is an optimal (minimal) subset of attributes preserving the classification ability as the

TABLE I.
FEATURES DEFINED ACCORDING TO THE GLASGOW OUTCOME SCALE

Code	Name	Values
A1	**Gender**	*Male*; *Female*
A2	**Admission_diagnosis** (Acc. to ICD-10 classification)	*Subarachnoid_hemorrhage; Intracerebral_hemorrhage; Cerebral_infarction; Stroke; Other_cerebrovascular_ diseases*
A3	**Final_diagnosis** (Acc. to ICD-10 classification)	*Subarachnoid_hemorrhage; Intracerebral_hemorrhage; Cerebral_infarction; Stroke; Other_cerebrovascular_ diseases*
A4	**Body_temperature** [^{0}C]	Discrete variable
A5	**Age** [years]	Discrete variable
A6	**Abode**	*Town*; *Village*
A7	**Time spent in hospital** [days]	Discrete variable
A8	**Time_elapsed** (from observation of illness occurrence to hospital admission)	*Less_than_1_hour; Less_than_3_hours; 3-6_hours; 6-12_hours; 12-14_hours; 2-3_days; More_than_3_days*
A9	**Patient_cure_location**	*Stroke_ward; Neurology_ward*
B1	**Arterial_hypertension**	*Present*; *Absent*
B2	**Ischemic_heart_disease**	*Present*; *Absent*
B3	**Past_cardiac_infarct**	*Present*; *Absent*
B4	**Atrial_fibrillation**	*Present*; *Absent*
B5	**Organic_heart_disease**	*Present*; *Absent*
B6	**Circulatory_insufficiency**	*Present*; *Absent*
B7	**Diabetes**	*Present*; *Absent*
B8	**Hypercholesterolemia**	*Present*; *Absent*
B9	**Obesity**	*Present*; *Absent*
B10	**Transient_ischemic_attack**	*Present*; *Absent*
B11	**Past_stroke**	*Present*; *Absent*
B12	**Infection_in_a_week_to_stroke**	*Present*; *Absent*
B13	**Alcohol_addiction**	*Present*; *Absent*
B14	**Nicotine_addiction**	*Present*; *Absent*
C1	**Systolic_pressure**	*Present*; *Absent*
C2	**Diastolic_pressure**	*Present*; *Absent*
C3	**Pulse**	Discrete variable
C4	**Heart_action**	*Normal_rythm; Atrial_fibrylation; Other_dysrythmia*
C5	**General_state_at_admission**	*Getting_up_alone; Staying_in_bed_ consciousness; Consciousness_disturbances*
C6	**Consciousness_at_admission**	*Conscious; Coma; Consciousness_disturbances*
C7	**Stroke_type*** (Acc. to Oxford classification, OCSP)	*LACS; PACS; POCS; TACS; Hard_to_class*

TABLE I (CONTINUED).
FEATURES DEFINED ACCORDING TO THE GLASGOW OUTCOME SCALE

Code	Name	Values
D1	**Consciousness_disorders** (during cure)	*Present*; *Absent*
D2	**Speech_disorders** (during cure)	*Present*; *Absent*
D3	**Swallowing_disorders** (during cure)	*Present*; *Absent*
E1	**Aspirine_treatment**	*Present*; *Absent*
E2	**Anticoagulants**	*Present*; *Absent*
E3	**Antibiotics**	*Present*; *Absent*
E4	**Antihypertensives**	*Present*; *Absent*
E5	**Anti-edematous_agents**	*Present*; *Absent*
E6	**Neuroprotective_agents**	*Present*; *Absent*
F1	**Exercise_therapy**	*Present*; *Absent*
F2	**Speech_therapy**	*Present*; *Absent*
F3	**Occupational_therapy**	*Present*; *Absent*
GOS	**Glasgow_Outcome_Scale**	**1** (death) **2** (persistent vegetative state) **3** (severe disability) **4** (moderate disability) **5** (good recovery)

original set of attributes. Various rough set methods were proposed to calculate decision reducts in decision tables, however calculation of all decision reducts is the *NP*-hard problem (see [20]). Therefore, in the experiments, we have used a more efficient method, called the QUICKREDUCT algorithm proposed in [21] and implemented in the *Rough Sets package* for the R environment. It is an example of a method producing the so-called decision superreduct that is not necessarily a decision reduct (i.e., it is a subset of attributes that may be not minimal).

After subset selection, the classification process was applied. All results of classification, before and after feature selection, are presented in Table II. In this table, results were obtained using a dataset divided into five concepts. However, we also provide results gathered using a modified dataset, where five primary concepts were replaced by two more general categories: *healthy* and *sick*. Healthy concept corresponds to the 5th concept, i.e., *good recovery*, in turn, the *sick* concept corresponds to the remaining concepts merged into one.

During the experiments, the Orange data mining tool [17] and the R environment were applied. Our own implementation of algorithms in this environment was also involved. The 10-fold cross validation paradigm was also applied during the classification process.

III. RESULTS AND CONCLUSIONS

The results of feature selection and calculation of quality of classification are acquired in Table II and Table III. Additionally, the average results are presented in a form of a chart, see Figure 1. It could be observed that each method caused decreasing a number of features in comparison to the original dataset. Particularly, in case of the five-class problem, application of contrast features led to selection of 12 relevant features from 42 original features, and at the same time classification accuracy (CA) and area under ROC curve (AUC) [18,19] slightly increased. Other three methods also reduced a feature space, from 42 features to 29, 17 and 9 using *DRQualityImp, DTLevelImp,* and *Rough Set* approaches respectively. However, in these approaches, CA and AUC parameters slightly decreased. In turn, in case of the two-class problem, there could be observed substantial improvement of classification accuracy.

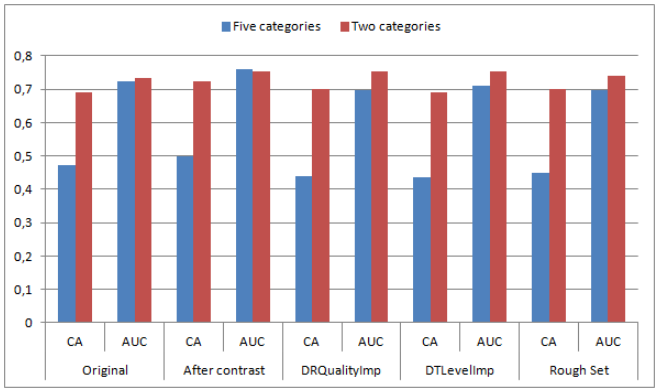

Fig. 1 Average results of classification accuracy (CA) and area under ROC curve (AUC) using five learning models.

During the experiments, some of the features achieved significant values of ranking measures. In turn, other features were estimated as much less important. In this way, it could be stressed that the most important features should be diagnosed very carefully.

The future research should be focused on simplification of the descriptive parameters, finding the compromise of a low classification error rate according to high efficiency of the Glasgow Outcome Scale. Some constructive induction methods could be applied to find general measures that may simplify diagnosis support for medical specialists.

REFERENCES

[1] J. Wright. (2000) The disability rating scale. The Center for Outcome Measurement in Brain Injury. [Online]. Available: http:// www.tbims.org/ combi/drs

[2] P.B. Gorelick, M. Atler (2002) *The prevention of stroke*. CRC Press.

[3] K. M. Hall, T. Bushnik, B. Lakisic-Kazazic, J. Wright, and A. Cantagallo, "Assessing traumatic brain injury outcome measures for longterm follow-up of community-based individuals," *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 3, pp. 367–374, 2001. doi: 10.1053/apmr.2001.21525

[4] L. Wilson, L. E. Pettigrew, and G. M. Teasdale, "Structured interviews for the Glasgow outcome scale and the extended Glasgow outcome scale: Guidelines for their use," *Journal of Neurotrauma*, vol. 15, no. 8, pp. 573–585, 1998.

[5] J. Wilson, L. Pettigrew, and G. Teasdale, "Emotional and cognitive consequences of head injury in relation to the Glasgow outcome scale," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 69, no. 2, pp. 204–209, 2000. doi: 10.1136/jnnp.69.2.204

TABLE II.
CLASSIFICATION RESULTS USING THE ORIGINAL SET AND THE SELECTED SUBSET OF IMPORTANT FEATURES, APPLYING **FIVE** CLASSES.

Dataset	**Original**		**After contrast features**		**DRQualityImp**		**DTLevelImp**		**Rough Set**	
# of features	42		12		29		17		8	
Classification quality	**CA**	**AUC**	**CA**	**AUC**	**CA**	**AUC**	**CA**	**AUC**	**CA**	**AUC**
CN2	0.4279	0.6541	0.5081	0.7512	0.4482	0.6320	0.3662	0.5618	0.4471	0.6228
CT	0.4338	0.6734	0.4165	0.6848	0.3912	0.6453	0.4592	0.6820	0.4904	0.7221
kNN	0.4904	0.6963	0.5397	0.7482	0.4103	0.6718	0.4210	0.7576	0.4217	0.7012
SVM	0.4772	0.7846	0.5151	0.8051	0.4901	0.7393	0.4397	0.7358	0.4401	0.6705
RF	0.5279	0.8145	0.5213	0.8132	0.4529	0.7913	0.4960	0.8117	0.4526	0.7755
AVG	0.4714	0.7246	0.5001	0.7605	0.4385	0.6959	0.4364	0.7098	0.4504	0.6984

TABLE III.
CLASSIFICATION RESULTS USING THE ORIGINAL SET AND THE SELECTED SUBSET OF IMPORTANT FEATURES, APPLYING ONLY **TWO** CLASSES.

Dataset	**Original**		**After contrast features**		**DRQualityImp**		**DTLevelImp**		**Rough Set**	
# of features	42		12		24		17		9	
Classification quality	**CA**	**AUC**	**CA**	**AUC**	**CA**	**AUC**	**CA**	**AUC**	**CA**	**AUC**
CN2	0.6640	0.6858	0.7518	0.7585	0.6702	0.7452	0.6835	0.7471	0.6890	0.7519
CT	0.6452	0.6532	0.6768	0.7074	0.6765	0.7004	0.6640	0.7123	0.6890	0.6778
kNN	0.6640	0.7581	0.6963	0.7301	0.6827	0.7242	0.6893	0.7516	0.6574	0.7082
SVM	0.7511	0.7756	0.7577	0.7842	0.7452	0.7860	0.7151	0.7687	0.7199	0.7610
RF	0.7261	0.7975	0.7386	0.7848	0.7257	0.8067	0.7077	0.7862	0.7449	0.8076
AVG	0,6901	0,7340	0,7242	0,7530	0,7001	0,7525	0,6919	0,7532	0,7000	0,7413

[6] J.T. King, Jr., P.M. Carlier, and D.W. Marion, "Early Glasgow Outcome Scale Scores Predict Long-Term Functional Outcome in Patients with Severe Traumatic Brain Injury", *Journal of Neurotrauma*. September, vol. 22, no. 9, pp. 947-954, 2005.

[7] J. W. Grzymała-Busse, Z. S. Hippe, T. Mroczek, W. Paja, and A. Bucinski, "A preliminary attempt to validation of Glasgow outcome scale for describing severe brain damages," in *Human-Computer Systems Interaction: Backgrounds and Applications*, Z. S. Hippe and J.L. Kulikowski, Eds. Berlin Heidelberg: Springer, 2009, pp. 161–170.

[8] A. Bruno, N. Shah, C. Lin, B. Close, D. C. Hess, K. Davis, V. Baute, J. A. Switzer, J. L. Waller, and F. T. Nichols, "Improving modified Rankin Scale assessment with a simplified questionnaire," *Stroke*, vol. 41, no. 5, pp. 1048–1050, 2010. doi: 10.1161/STROKEAHA.109.571562

[9] B. Jennet and M. Bond, "Assessment of outcome after severe brain damage: A practical scale," *Lancet*, vol. 1, pp. 480–484, 1975.

[10] S.E. Kasner, "Clinical interpretation and use of stroke scales," *The Lancet Neurology*, vol. 5, no. 7, pp. 603–612, 2006.

[11] J. Lu et al., "A method for reducing misclassification in the extended Glasgow Outcome Score," *Journal of Neurotrauma*, vol. 27, no. 5, pp. 843-852, 2010.

[12] A.D. Nichol et al. "Measuring Functional and Quality of Life Outcomes Following Major Head Injury: Common Scales and Checklists," *Injury*, vol. 42, pp. 281-287, 2011. doi: 10.1016/j.injury.2010.11.047.

[13] W. R. Rudnicki, M. Wrzesien, and W. Paja, "All relevant feature selection methods and applications," in *Feature Selection for Data and Pattern Recognition*, U. Stanczyk and L. C. Jain, Eds. Berlin Heidelberg: Springer, 2015, pp. 11–28. doi: 10.1007/978-3-662-45620-0_2

[14] H. Stoppiglia, G. Dreyfus, R. Dubois, and Y. Oussar, "Ranking a random feature for variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1399–1414, 2003.

[15] E. Tuv, A. Borisov, and K. Torkkola, "Feature selection using ensemble based ranking against artificial contrasts," in *Proceedings of the 2006 IEEE International Joint Conference on Neural Network*, 2006. doi: 10.1109/IJCNN.2006.246991 pp. 2181–2186.

[16] W. Paja, "Feature selection methods based on decision rule and tree models," in *Intelligent Decision Technologies 2016: Proceedings of the 8th KES International Conference on Intelligent Decision Technologies (KES-IDT 2016) - Part II*, I. Czarnowski, A. M. Caballero, R. J. Howlett, and L. C. Jain, Eds. Cham: Springer International Publishing, 2016, pp. 63–70. doi: 10.1007/978-3-319-39627-9_6

[17] J. Demsar et al., "Orange: Data mining toolbox in Python," *Journal of Machine Learning Research*, vol. 14, pp. 2349–2353, 2013.

[18] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. doi: 10.1016/j.patrec.2005.10.010

[19] J. Hernández-Orallo, P. Flach, and C. Ferri, "A unified view of performance metrics: Translating threshold choice into expected classification loss," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 2813–2869, 2012.

[20] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," in *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, R. Słowinski, Ed. Dordrecht: Kluwer Academic Publishers, 1992, pp. 331–362. doi: 10.1007/978-94-015-7975-9_21

[21] Q. Shen and A. Chouchoulas, "A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems," *Engineering Applications of Artificial Intelligence*, vol. 13, no. 3, pp. 263–278, 2000. doi: 10.1016/S0952-1976(00)00010-5