© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

Stroke risk factors classification modelling

A. Lourenço[1], A.C. Braga[2] & O. Belo[1]
[1]*Department of Informatics, University of Minho, Portugal.*
[2]*Department of Production and Systems, University of Minho, Portugal.*

Abstract

Medicine like many other business and scientific areas has began to realise the advantages of Knowledge Discovery and Data Mining applications. As data volumes are increasing almost exponentially and physicians do not have to perform data analysis, new methodologies bringing light to the subject are more than welcome. This paper presents a very actual medical problem, studying the behaviour of distinct types of strokes according to several well-known risk factors. The aim of this work is to perform risk factors evaluation, figuring out what are the best data mining techniques to be taken and dealing with the conditionings of the problem and data. This can be viewed as a feature selection problem, where the most prevalent features to each type of stroke are to be selected. There were built several classification models relating stroke types, comparing the discriminative strength of each one and the acquired knowledge. Additionally, an association rules approach was also taken, confronting results and enriching the obtained knowledge.

Introduction

Knowledge Discovery in Databases (KDD) and data mining are by far one of the most prominent research areas of the moment. Knowledge acquisition is the key of any organisation's success, in spite of its particular domain, interests and resources. Adequate, timely data analysis has a preponderant role within in organisational processes.

During the last decades, KDD has been enlarging its "domains" as it began to prove its real value in areas like Finances, Stock Market Analysis, Business Analysis and Medicine. The idea of seeking knowledge with automatic, or semi-automatic, processing seemed quite appealing especially because data volumes have been increasing almost exponentially. Today, there are an enormous number

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

of data mining algorithms ready to dig into almost all sorts of data, tracking different kinds of patterns and answering a broad number of questions. Moreover, there have been implemented several pre-processing and pos-processing techniques, improving both data mining and output readability and understanding.

This paper has the clear purpose of applying KDD methodologies to a real-world problem in the area of Medicine. Although, it was established a perfectly delimited problem, the main idea is to analyse the advantages these methodologies can bring to the area, rather then actually extracting something novel or particularly interesting from the data. Confronting KDD approaches with real data and its conditionings and implications allows the researchers to evaluate and eventually modify approaches towards problems' compliance and incites organisations' administrators and analysts to put them in practice [1] [2] [3]. The case study that was selected describes a problem of risk factors evaluation in the area of strokes. The aim is to analyse the influence of each risk factor or particular subsets of factors, over each type of stroke. This knowledge will help physicians in stroke diagnosis and prognosis, granting a more suitable and opportune treatment. Patients clinical history concerning diabetes, cardiac diseases, previous occurrence of strokes, high cholesterol levels, high blood pressure and renal failure data, among others, were taken into account. The intracerebral hemorrhage, the subarachnoid hemorrhage and the cerebral ischemia were the stroke types considered. A fourth type, the clinical stroke, was also included, reporting cases of this seizure, where it was not possible to determine the specific type.

In the next section, it is given a brief introduction to strokes domain, in order to help the reader understanding the problem's conditionings and the analysts' limitations. Besides explaining what is a stroke and exposing a major taxonomy, it is put special interest in justifying the necessity of risk factors evaluation. The principal stages in strokes treatment are described, pointing out its main implications. By far, prevention raises itself as the key to minimize the occurrence of strokes or, at least, to detect the individual's predisposition to "brain attack", reducing its effects and, therefore, eventual disabilities. Intentionally, there are listed the most common symptoms and consequences (temporal loss of faculties or disabilities) of strokes in order to highlight how crucial and urgent risk factors analysis really is to the problem. Later, the dataset is introduced and data mining is described. Due to several dataset's specificities there were made some "adjustments" to the set and two different approaches were taken, combining classification and the generation of association rules, trying to extract relevant and valuable knowledge [4]. The results of each approach are exposed, comparing each other outputs and some final considerations are made.

The genesis of strokes

A stroke is a sudden injury to the brain caused by an abnormality in a brain blood vessel that causes the interruption of blood flow to the brain. As a result, there is

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

an inadequate delivery of oxygen and glucose to brain tissue, which might results in death of brain cells (neurons), if blood flow is not restored as quickly as possible.

There are two kinds of situations that may cause strokes: the blockage of a blood vessel that all of a sudden interrupts the blood supply to a certain part of the brain and the burst of a blood vessel in the brain, spilling blood into the spaces surrounding the brain cells. Due to these, strokes can be classified in two major categories: hemorrhagic strokes, which are caused by bleeding into the brain, and ischemic strokes, originated by blockage of blood flow to the brain, ultimately leading to infarction.

Hemorrhagic strokes occur when blood from a ruptured blood vessel compresses and damages normal functioning brain tissue. This rupture may occur because of underlying damage to the blood vessel from years of high blood pressure or from an underlying abnormality in a blood vessel, such as an aneurysm. High blood pressure is pointed out as the most important risk factor for a hemorrhagic stroke, but other possible causes include alcohol or drug abuse (especially cocaine), and cigarette smoking. On the other hand, ischemic strokes are the most frequent cause of stroke, although not usually fatal, and are responsible for about 80 percent of strokes. In an ischemic stroke, a blocked artery prevents blood carrying oxygen and other nutrients from reaching a certain portion of the brain, leading to dysfunction and death of that brain tissue. These blockages stem from three conditions: the formation of a clot within a blood vessel, called thrombosis; the movement of a clot from another part of the body such as the heart, named as embolism; or a severe narrowing of an artery in or leading to the brain due to the build-up of plaque (a mixture of fatty substances), known as stenosis. What happens during an ischemic stroke resembles in many ways what happens in a heart attack in which there is a blockage of blood flow to the heart. So, it is natural that strokes and heart diseases share many risk factors like high blood pressure, diabetes, increased cholesterol levels, cigarette smoking, obesity and physical inactivity, among others.

Despite of its origin, the effects of a stroke vary considerably, depending on the particular part of the brain affected, and the size of the area involved. A stroke can leave a person with one or more of the following problems: paralysis in one limb or on one side of the body or face, loss of sensation, loss of balance, loss of bladder control, decreased level of consciousness or alertness, complete or partial blindness, swallowing difficulties, speech difficulties or inability to speak or thought and memory difficulties. Therefore, it is in the best interest of all of us to prevent these situations from actually happening, attacking the problem from the beginning. As Figure 1 shows, there are three possibilities of treatment stages for strokes: prevention, therapy and rehabilitation. First, it is vital to establish adequate prevention methodologies, capable of treating the underlying risk factors and reducing the probability of a stroke occurring.

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

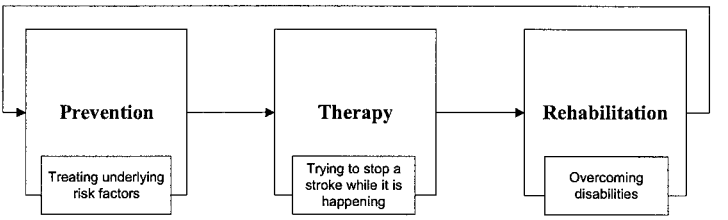


Figure 1: Treatment stages for stroke.

Acute stroke therapies try to stop a stroke while it is happening. It is often said that when it comes to stroke, "time is brain". This means that any delay in getting a stroke victim to a hospital and receiving proper assistance may result in greater damage to the brain tissue. The first step in treating a stroke is to identify whether the symptoms are caused by an ischemic stroke or a hemorrhagic stroke, since the therapeutic interventions and prognosis are quite different. This can be achieved by performing a CT scan (a special x-ray test that produces images of thin cross sections of the body) of the head. In the case of a hemorrhagic stroke, blood may be seen in the brain tissue itself (intracerebral hemorrhage) or surrounding the brain (subarachnoid hemorrhage). In the case of subarachnoid hemorrhage, further evaluation is required to identify the source of bleeding, which is usually due to rupture of an aneurysm. Finally, post-stroke rehabilitation deals with the disabilities coming from brain tissue injuring, minimizing the lifetime sequels. These techniques can be also understood as a first step towards the prevention of eventual future strokes.

Although all stages play key roles in the process, it is undeniable that prevention must be privileged. The emotional, physical and inevitable monetary costs involved in the recovery of a "brain attack" are too expensive. Individuals experience traumatic experiences and suffer from severe disabilities that in most cases take too long to heal or even are never overcome. So, it is important to target the problem as soon as possible, trying to avoid strokes.

Stroke prevention

In order to prevent strokes occurrence or, at least, to minimize its consequences, there are two major steps to be taken: know stroke's warning signs and control stroke's risk factors. Persons must be aware of the signs and symptoms of a stroke so they can identify it on the spot and look for immediate medical attention. The symptoms of a stroke can vary, depending upon which part of the brain is affected, but it is important to realize that a stroke is not painful and generally does not cause a headache (the one exception is subarachnoid hemorrhage). It manifests itself as a combination of neurological symptoms or loss of functions, such as: numbness or weakness on one side of the body; difficulty with eyesight or vision; complete or partial loss of the visual field (vision to one side), which involves both eyes; complete or partial loss of vision

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

in one eye; double vision; the sensation of spinning (vertigo); slurred speech; disturbance of language functions; inability to express oneself, through both speech and writing; inability to understand what is being said or read.

On the other hand, individuals' underlying risk factors must be controlled. There are a large number of factors that might be pointed out as cause of stroke [5] [6]. Many of these come as genetic inheritance, like race or some hereditary diseases (for instance, heart malfunctioning problems and degenerative diseases), or are simply non modifiable, like older age and gender. However, there are also modifiable risk factors that can be detected, treated or controlled, such as: high blood pressure, atrial fibrillation and other heart diseases, high cholesterol levels, cigarette smoking, diabetes mellitus, physical inactivity, carotid artery stenosis or transient ischemic attack and migraine. Even when the suppression of the risk is not possible, knowing opportunely that an individual is predisposed to stroke is already a major step towards the success of his treatment.

A risk factors evaluation case study

As referred earlier, the case study presented here relates the influence of certain risk factors over four different types of strokes. The data source was collected from the population of the Internal Medicine service of the Hospital of Guimarães, in Portugal, during the last quarter of 1991 and over the year 1992. Data was firstly collected retrospectively and then during 1992, the process became prospective, holding an inquiry among the patients of this medical service [7].

In this particular case study the emphasis is put in the identification of the most prevalent risk factors for each type of stroke, i. e., it is intended to study a certain number of well-known risk factors, evaluating their relation (individually or in small conjunctions) with each stroke type. There were considered the following risk factors: age; gender; high blood pressure (hbp); increased cholesterol levels or hyperlipidemia (hc); diabetes (diab); cigarette smoking (cigar); three cardiac diseases: atrial fibrillation (af); hypertrophy of the left ventricle (hlv) and infarct of the myocardium (im); previous stroke occurrence (pstr); renal failures (rf). All of these variables are binary with the exception of hypertension for which there were considered three levels, and age that is numeric. Each of the binary variables expresses the absence or presence of certain risk factor in the individual's clinical history (for gender the value 0 signs males, while value 1 stands for females). The four different types of strokes in question are: intracerebral hemorrhage (type=1); subarachnoid hemorrhage (type=2); cerebral infarction or cerebral ischemia (type=3); clinical stroke, used to classify individuals that present a clinical diagnosis proving the occurrence of stroke, but which was not confirmed neither with TC scan nor with any other "specialized" mean and therefore, it was not possible to determine its real type (type=0).

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

A data mining approach to strokes' risk factor analysis

At a first glance, the case study can be viewed as a classification problem, where the built decision tree will present at the leaves level the different types of stroke and the various branches will be risk factors discrimination towards them [8]. In this sense, it might be possible to detect what paths (conjunction of variables and values) lead to a certain type of stroke.

To perform classification, the c4.5 algorithm [9] was chosen for two main reasons: this is a well-known mining algorithm, which does not require further description; it selects relevant features by itself in tree branching, so it can be used as a benchmark, as in Liu & Setiono [10], to verify the effects of the feature selection attributes. The output is presented in Figure 2 in a rule- based form, in order to improve comprehensibility. After giving a look at the rules, one notices that the subarachnoid hemorrhage was overlooked, having cerebral ischemia certain predominance. Moreover, some of the rules seem trivial and unproductive.

If gender=0 and age>82 and af=1 and rf=0 and pstr=0 **then** class 0 (63.0%)
If hbp=1 and hlv=0 and af=1 and cigar=0 and diab=0 and rf=1 and pstr=1 **then** class 1 (75.8%)
If gender=0 and age>60 and age<=69 and hbp=0 and hlv=0 and af=0 and diab=0 and hc=0 and rf=0 and pstr=0 **then** class 1 (57.9%)
If gender=0 and age>50 and age<=60 and im=0 and hc=1 and pstr=0 **then** class 1 (49.6%)
If gender=1 and age<=81 and age>50 and hbp=1 and hlv=0 and af=0 and im=0 and hc=0 and pstr=0 **then** class 1 (44.3%)
If pstr=1 **then** class 3 (84.7%)
If hc=1 **then** class 3 (81.0%)
If af=1 and im=0 **then** class 3 (80.7%)
If gender=0 and hbp=1 and hlv=0 **then** class 3 (74.7%)

Figure 2: The decision rules obtained with c4.5 algorithm.

In order to understand what was happening, there was made a stroke types representativeness analysis, which is presented in Figure 3. Now, it becomes clear that cerebral ischemia prevails over the rest of types, detaining more than 70 percent of the population (in a total of 656 cases). The other types are underrepresented, especially subarachoid hemorrhage, and this situation implies an extra care during data mining and results pos-processing. In particular, it is quite probable that decision tree generation algorithms "ignore" certain types, tending to the majority, i. e., misclassifying these cases as cerebral ischemia cases, as the overall precision does not suffer almost any impact.

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

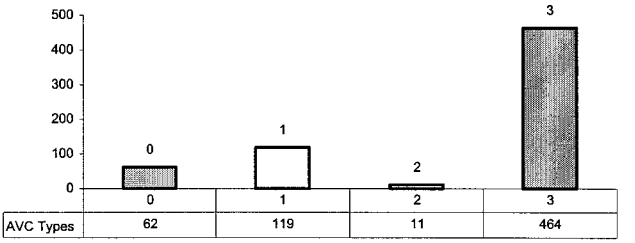


Figure 3: The stroke types representativeness within the dataset.

Even when a different data mining approach was taken, proceeding to rule association generation, the situation did not get any better. The Apriori algorithm [11], also a well-known mining algorithm, was not able to retrieve any rule concerning subarachoid hemorrhage and the rules related to the hemorrhagic types present very low support and confidence levels, making it difficult to realise their true "potential". In this sense, to overtake this population problem, there was made a first attempt of sampling by substitution, trying to increase the presence of "weaker" types in the dataset. However, this attempt did not bring any profit. As information is so scarce, the algorithms continues to tend to majority. As a second approach, the original dataset was fragmented into four new datasets, isolating one stroke type per dataset. Now, the class is binary, determining the occurrence of a specific stroke instead of trying to discriminate between several types. In Figure 4 there are the rule sets generated for each one of these datasets.

Cerebral Ischemia

If af=0 and im=1 and cigar=0 **then** class 1 (88.6%)
If pstr=1 **then** class 1 (84.7%)
If diab=0 and hc=1 **then** class=1 (82.5%)
If af=1 **then** class=1 (79.7%)
If gender=0 and hbp=1 **then** class 1 (73.5%)
If af=0 and im=0 and hc=0 and pstr=0 **then** class 0 (44.9%)

Intracerebral Hemorrhage

If hbp=1 and hlv=0 and af=0 and cigar=0 and diab=0 and rf=1 and pstr=1 **then** class 1 (75.8%)
If cigar=1 and diab=1 and hc=1 **then** class 1(50.0%)
If af=1 **then** class 0 (95.1%)
If pstr=1 **then** class 0 (89.0%)
If im=1 **then** class 0 (88.7%)
If diab=0 and hc=1 **then** class 0 (87.7%)
If cigar=0 and hc=1 **then** class 0 (87.6%)
If hbp=0 **then** class 0 (83.1%)

Clinical Stroke

If gender=0 and age>82 and af=1 and rf=0 and pstr=0 **then** class 1 (63.0%)
If hlv=1 and diab=1 and pstr=0 and age>72 **then** class 1 (63%)
If gender=1 and diab=1 and rf=1 age>72 **then** class 1 (45.3%)
If hc=1 **then** class 0 (95.0%)
If pstr=1 **then** class 0 (94.7%)
If cigar=1 **then** class 0 (94.6%)
If age<=72 **then** class 0 (93.8%)
If af=0 and diab=0 **then** class 0 (91.1%)

Figure 4: Decision rules brought by modified datasets.

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

Cerebral Ischemia

If pstr=1 **then** tavc=3 [Coverage=0.341 (224); Support=0.296 (194); Strength=0.866; Lift=1.22; Leverage=0.0542 (35)]
If hbp=0 & pstr=1 **then** tavc=3 [Coverage=0.142 (93); Support=0.134 (88); Strength=0.946; Lift=1.34; Leverage=0.0339 (22)]
If hc=1 **then** tavc=3 [Coverage=0.223 (146); Support=0.186 (122); Strength=0.836; Lift=1.18; Leverage=0.0286 (18)]
If af=1 **then** tavc=3 [Coverage=0.192 (126); Support=0.159 (104); Strength=0.825; Lift=1.17; Leverage=0.0227 (14)]
If diab=1 & pstr=1 **then** tavc=3 [Coverage=0.081 (53); Support=0.078 (51); Strength=0.962; Lift=1.36; Leverage=0.0206 (13)]
If age < 67 & pstr=1 **then** tavc=3 [Coverage=0.104 (68); Support=0.091 (60); Strength=0.882; Lift=1.25; Leverage=0.0181 (11)]
If gender=0 & diab=1 **then** tavc=3 [Coverage=0.111 (73); Support=0.091 (60); Strength=0.822; Lift=1.16; Leverage=0.0128 (8)]
If im=1 **then** tavc=3 [Coverage=0.067 (44); Support=0.058 (38); Strength=0.864; Lift=1.22; Leverage=0.0105 (6)]
If age < 67 & diab=1 & pstr=1 **then** tavc=3 [Coverage=0.029 (19); Support=0.029 (19); Strength=1.000; Lift=1.41; Leverage=0.0085 (5)]
If age > 76 & diab=1 & pstr=1 **then** tavc=3 [Coverage=0.023 (15); Support=0.023 (15); Strength=1.000; Lift=1.41; Leverage=0.0067 (4)]

Subarachnoid Hemorrhage

If gender=1 **then** tavc=2 [Coverage=0.480 (315); Support=0.017 (11); Strength=0.035; Lift=2.08; Leverage=0.0087 (5)]
If gender=1 & age < 67 **then** tavc=2 [Coverage=0.149 (98); Support=0.011 (7); Strength=0.071; Lift=4.26; Leverage=0.0082 (5)]

Intracerebral Hemorrhage

If af=0 & pstr=0 **then** tavc=1 [Coverage=0.538 (353); Support=0.143 (94); Strength=0.266; Lift=1.47; Leverage=0.0457 (29)]
If af=0 & diab=0 & hc=0 **then** tavc=1 [Coverage=0.468 (307); Support=0.125 (82); Strength=0.267; Lift=1.47; Leverage=0.0401 (26)]
If hbp=1 & af=0 & hc=0 **then** tavc=1 [Coverage=0.337 (221); Support=0.101 (66); Strength=0.299; Lift=1.65; Leverage=0.0395 (25)]
If pstr=0 **then** tavc=1 [Coverage=0.659 (432); Support=0.149 (98); Strength=0.227; Lift=1.25; Leverage=0.0299 (19)]
If af=0 **then** tavc=1 [Coverage=0.808 (530); Support=0.175 (115); Strength=0.217; Lift=1.20; Leverage=0.0287 (18)]
If gender=0 & pstr=0 **then** tavc=1 [Coverage=0.329 (216); Support=0.076 (50); Strength=0.231; Lift=1.28; Leverage=0.0165 (10)]
If age < 67 **then** tavc=1 [Coverage=0.329 (216); Support=0.075 (49); Strength=0.227; Lift=1.25; Leverage=0.0150 (9)]
If age > 76 & hbp=1 & af=0 & hc=0 **then** tavc=1 [Coverage=0.091 (60); Support=0.029 (19); Strength=0.317; Lift=1.75; Leverage=0.0124 (8)]
If cigar=1 & diab=1 & hc=1 **then** tavc=1 [Coverage=0.003 (2); Support=0.003 (2); Strength=1.000; Lift=5.51; Leverage=0.0025 (1)]

Figure 5: Association rules obtained for each dataset.

Additionally, these datasets were mined with Magnum Opus™ association rules tool [12], trying to take advantage of certain concepts brought by this software. In particular, it was used a leverage search, seeking strength and coverage, but also weighting these metrics in terms of types proportion and, insignificant rules were filtered, in order to have a better perception about each "good" rule's real value. (Figure 5).

After comparing the different results, some considerations can be made. In terms of cerebral ischemia it can be pointed the following considerations: a previous

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

stroke attack, high levels of cholesterol, atrial fibrillation or an infarct of the myocardium will be likely to predispose individuals to this stroke; also, clinical scenes of diabetes conjugated with previous seizures and/or certain age ranges might influence. Concerning the subaracnoid hemorrhage, the scarce data available only points to a female predominance, although this can be deceiving. The intracerebral hemorrhage is affected by high blood pressure, diabetes and high levels of cholesterol, appearing to have no remarkable relation with cardiac diseases or with the occurrence of previous strokes. Having high blood pressure is a major risk for most of strokes. About the clinical stroke type further considerations need to be made. As this is a default classification, when there is not enough information about the specific type, it is probable that most of its cases actually relate one of the other types. In this sense, these elements only introduce noise to the process, weakening even more the presence of hemorrhagic types. Moreover, as it is not a "true" type, the obtained conclusions become pointless.

Conclusions

Throughout this paper strokes had been "dissected". As the aim here was to perform a data mining task towards the evaluation of risk factors affecting strokes, it was indispensable to understand the domain. Before data analysis, strokes were explained in terms of nature, types, symptoms and risk factors. Data analysis can only be profitable if the problem is completely understood. Data is always unpredictable and their conditionings must be taken into account to ensure knowledge discovery. In this particular case, the dataset offered some "resistance" due to a badly distributed population of stroke types. This situation raised some obstacles as both decision rules and association rules tended to overlook the minorities, privileging the predominant type. The type in cause, cerebral ischemia, represents one of the most frequent strokes in the world, which explains why there is plenty of information about it. However, this situation prevented or at least made it difficult to extract interesting rules about the other types. Moreover, the clinical stroke type stands for "unsolved" cases, where it was not possible to identify the specific type. In this sense, it is likely that these cases belong indeed to one of the other studied types, creating noise within the process. Generally, the aim of exposing the benefits brought by KDD and data mining to Medicine domains were accomplished. The evaluation of risk factors affecting strokes is a very relevant matter, as strokes are one of the main death causes of nowadays. This study proves that, not only it is possible to use KDD methodologies, but also the combination of different data mining approaches actually allows the acquaintance of important conclusions, supporting physicians in timely and adequate strokes diagnosis and prognosis.

References

[1] Magoulas, G. D. & Prentza, A., Machine Learning in Medical Applications in Machine Learning and Its Applications, Springer-Verlag, pp. 300-307, 2001.

© 2002 WIT Press, Ashurst Lodge, Southampton, SO40 7AA, UK. All rights reserved.
Web: www.witpress.com Email witpress@witpress.com
Paper from: *Data Mining III*, A Zanasi, CA Brebbia, NFF Ebecken & P Melli (Editors).
ISBN 1-85312-925-9

[2] Pazzani, M. J., Mani, S. & Shankle, M. J., Acceptance of Rules Generated by Machine Learning among Medical Experts, Methods of Information in Medicine; 40: 380-385, 2001.

[3] Mani, S., Shankle, W.R., Pazzani, M. J., Smyth, P. & Dick, M.B., Differential Diagnosis of Dementia: A Knowledge Discovery and Data Mining (KDD) Approach. American Medical Informatics Association (AMIA) Annual Fall Symposium, Nashville, 1997.

[4] Liu, B., Hsu, W. & Ma, Y., Integrating Classification and Association Rule Mining. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, New York, USA, 1998.

[5] Karter, A. J., Gazzaniga, J. M., Cohen, R. D., Casper, M. L., Davis, B. D. & Kaplan, G. A., Ischemic heart disease and stroke mortality in African-American, Hispanic, and non-Hispanic white men and women, 1985 to 1991. Western Journal of Medicine, 169, 139-145, 1998.

[6] Reyes-Iglesias, Y., Melendez, R., Hernandez, E. & Perez, M., Stroke in Hispanic veterans: a descriptive study. Southern Medical Journal, 92, 28-32, 1999.

[7] Braga, A. C. & Iglésias Pereira, H. M., Acidente Vascular Cerebral e seus Factores de Risco: Estudo de Ocorrência de 4 tipos de AVC. In Proceedings of II Congresso Anual da Sociedade Portuguesa de Estatística, Lisboa, pp. 445-454, 1994 (in portuguese).

[8] Alexopoulos, E., Dounias, G.D. & Vemmos, K., Medical diagnosis of stroke using inductive machine learning. In Proceedings of Workshop on Machine Learning in Medical Applications, Advance Course in Artificial Intelligence-ACAI99, Chania, Greece, pp. 20-23, 1999.

[9] Quinlan, J.R., C4.5: Programs for Machine Learning, Morgan Kaufmann, 1993.

[10] Liu, H. & Setiono, R., Feature Selection via Discretization. IEEE Trans Knowledge and Data Engineering 9:4, 642-645, 1997.

[11] Agrawal, R. & Srikant, R., Fast Algorithms for Mining Association Rules. Proc. of the 20th Int'l Conference on Very Large Databases, Santiago, Chile, 1994.

[12] http://www.rulequest.com/.