EVALUATING RELIABILITY AND RELEVANCE FOR WOWA AGGREGATION OF SLEEP APNEA CASE DATA

David Nettleton
Computer Languages &
Systems Department
University Polytechnic of Catalunya, Spain
netleton@lsi.upc.es

Lourdes Hernández
Respiratory Disease Institute,
Hospital Clinic
University of Barcelona, Spain
fburgos@medicina.ub.es

Summary

In this article, joint medical and data analysis expertise is brought to bear using contrasting data analysis methods and the WOWA aggregation operator to solve a difficult medical diagnosis problem, that of sleep apnea syndrome screening. We describe a method of calculating the relevance and reliability weights used by the WOWA operator.
Key Words: sleep apnea diagnosis, questionnaire responses, WOWA aggregation, clustering, classification, reliability, relevance.

1. INTRODUCTION

Screening of Apnea cases is a difficult diagnosis problem, at present not satisfactorily resolved by standard statistical modelling techniques. We propose that part of the problem is due to the inherent fuzzy nature of a significant part of the data: questionnaire responses. We use diverse clustering and classification techniques to establish the relevance and reliability of each variable, which is then given to the WOWA aggregation operator to generate an aggregated value for each patient with high correlation to the apnea diagnosis.

The article is structured as follows: in section 2, a brief clinical description of the sleep apnea syndrome is given; in section 3, the method of diagnosis using WOWA operators and the problem of assigning the relevance and reliability weights is detailed; in section 4, unsupervised clustering techniques identify variables used to partition the data; in section 5, supervised modelling techniques are used to generate a ranking of variables in order of significance to the diagnosis output; in section 6, WOWA aggregation results are detailed using the reliability and relevance weights derived in sections 4 and 5; finally, section 7 discusses some conclusions of this method and the applicability to apnea diagnosis.

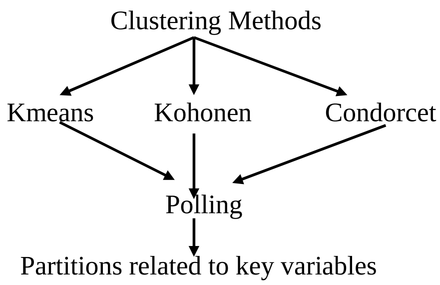

Figure 1a: Clustering Techniques determine relation of key variables to clusters

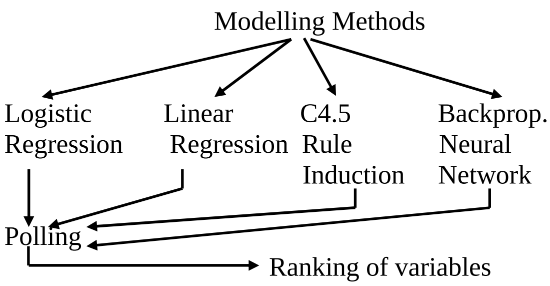

Figure 1b: Contrasting methods are polled to determine a ranking of relevance and reliability of the variables with respect to apnea diagnosis

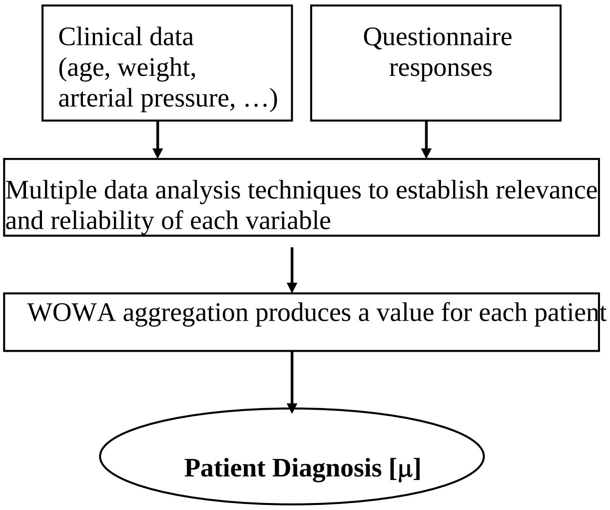

Figure 1c: Data processing of the apnea data input variables to produce a diagnosis

2. THE SLEEP APNEA SYNDROME

The Obstructive Sleep Apnea Syndrome (OSAS) is a set of secondary clinical manifestations relating to the ceasing (apnea) or reduction (hypopnea) of air flow during sleep, caused by a partial or total collapse of the upper air way at the faringe level. The severity of the OSAS is defined by the *apnea hypopnea index* (AHI) or the number of apneas plus the number of hypopneas per hour during sleep. Generally an AHI ∃10-15 is considered pathological.

2.1 DIAGNOSIS

The predictive value of the clinical data in OSAS diagnosis is low. Hoffstein [1] published results that indicated that clinical data explains 36% of the variability of the IAH (apnea hypopnea) and Katz [3] reported a figure of 39%, other authors report lower figures. The subjective clinical evaluation of the interviewer has also been evaluated and tends to have a low sensibility and specificity, in the order of 55%-65% respectively, for correctly classifying the sick. On the other hand, The predictive models for IAH based in clinical data have a higher sensibility of up to 90%. Their specificity, in the best of cases, does not reach 70%. At present, it is not appropriate to define rigid diagnostic criteria in this rapidly developing area. Neither is it possible to identify the ideal equipment for sleep studies.

2.2 APNEA PATIENT DATA

We have the collected data of the standard sleep patient questionnaire, for 154 patients, captured over a 1 year period. The data set contains 68,2% positive outcomes and 31.8% negative outcomes. The questionnaire consists of two main sections: the first records clinical data (age, weight, blood pressure, etc. ..); the second section consists of 41 questions to which the patient responds. The questions are divided in 3 subsections: 15 general sleep questions {s1...s15}, 16 respiratory related questions {r1…r16} and 9 somnolence related questions{s1…s9}. Based on this information, the doctor then gives a clinical evaluation: healthy; simple snorer; doubtful; typical apnea; other illness. We simplify this to: typical apnea; no apnea.

3. ESTABLISHING RELIABILITY AND RELEVANCE FOR WOWA AGGREGATION

We consider applying diverse data analysis techniques to the variables which have been collected for apnea patients. We wish to establish, for each variable, its relevance with respect to the apnea diagnosis, and its reliability. The first value will be the ρ weights in the first WOWA vector, the same vector being used for all variables and cases. The second value is represented by one ω vector per variable. Also we wish to establish the most significant variables with respect to apnea diagnosis. The four classification techniques used are: C4.5 rule induction, backpropagation neural network, logistical and linear regression. The techniques have been chosen so as to contrast the results given by significantly different approaches (Figures 1a, 1b, 1c).

3.1 WOWA AGGREGATION

In [5], Nettleton evaluated different aggregation methods for diagnosing sleep apnea. The following aggregation methods were considered: Ordered Weighted Average (OWA) [7], Weighted Ordered Weighted Average (WOWA) [6] and Principal Components. OWA uses a vector in which, for each variable, a value is assigned which indicates its relevance. In WOWA a second vector is incorporated whose values indicate the reliability of each variable. In this article we extend the previous work, calculating the weights from statistical analysis of the data.

3.2 RELEVANCE AND RELIABILITY

Relevance is a standard data analysis objective for which we can apply diverse algorithms and interpret the results. Relevance is more straightforward to establish in statistical data analysis, than reliability.

Reliability is influenced by different aspects. There are data aspects, such as %missing and %erroneous. Then there are application dependant aspects which, in the case of the questionnaire responses can be the truthfulness with which the patient responds (it may be that if the patient goes to sleep at the wheel of a car, s/he does not wish to make that known, and thus tends to give a higher negative response rate to this question {**s5**}, than it really should have). Each variable has a ω vector which, on ordered values, heightens some values while dampening others. We execute the WOWA aggregation function with an Ai data vector for each case, the $V\rho$ vector remains constant and represents the *relevance* of the variables, while the $V\omega_n$ vector represents the *reliability* of the ordered values of each variable and varies for each variable. For example, we call WOWA with (A_i, $V\rho$, $V\omega_n$), where $A_\mathbf{i}$ is data row **i**. Also, $V\omega_1$=even bias (E); $V\omega_2$=bias on low values (L); $V\omega_3$=bias on high values (H); $V\omega_4$=bias on medium values (M); $V\omega_5$=bias on high & low values (O). For example, $V\omega_1$=even bias (E) could be assigned as {0.2, 0.2, 0.2, 0.2, 0.2} for the five data values belonging to variable 1, while $V\omega_3$=bias on high values (H) could be assigned as{0.1,0.1,0.2,0.3,0.3}, for variable 3.

4. UNSUPERVISED CLUSTERING AND STATISTICAL TECHNIQUES

With reference to Table 1 (below) we can see that in methods 1, 2 and 3, *partner* has influenced partitioning. Kohonen y Kmeans tend to have biased the *'g'* responses while Condorcet has used the *'s'* responses more for partitioning. There is not a clear consensus between the different clustering and statistical techniques. Methods 1, 2 and 4 used numerical representation of all the data, while method 3 used a categorical representation with ChiSquared for the significance tests. **Kohonen net:** Various architectures of net were tried: input layers of 41 neurons (questionnaire responses only), 27 (clinical data only), and 68 (questionnaire responses and clinical data). **Kmeans:** Standard SPSS Kmeans was used for 2 clusters, maximum iterations set to 100, convergence at 0.02. **Condorcet – mixed data type clustering:** A proprietary IBM algorithm based on the *Condorcet* [2] distance criteria was used to generate 9 segments. All data was prepared as categorical and a chisquared measure was used to rank the variables in each segment and between segments. **Cross product covariances (Pearson):** A standard SPSS numeric covariance was used with the pearson product moment option, to generate covariances between all the variables, defined as numeric.

5. SUPERVISED CLASSIFICATION AND STATISTICAL MODELS

We contrast four techniques, each using a different algorithmic basis, with the objective of realising a consensus for the variables being evaluated. With reference to Table 2 (below) we can see that methods 1, 2 and 3 have identified *waist* as a significant attribute, while methods 1, 3 and 4 have identified *g1* as significant. Other identified variables are *r2*, *partner*, *weight* and *s10*. **C4.5 rule induction:** Quinlan's standard C4.5 algorithm was used, with 25% pruning, no external test set, and no grouping. **Backpropagation neural network:** The neural network training phase generates a *sensitivity analysis* which provides a ranking of the variables with respect to the output (in this case, the diagnosis yes/no). **Logistic regression:** Standard SPSS logistic regression was used with 3 test models. Overall precision's were: 89,66%, 88% and 75%. **Linear regression:** One SPSS linear regression was generated. The R^2 value was 0.31309, the standard error was 0.51035.

Table 1. Clustering and statistical techniques applied to the apnea cases and the identification of key variables which distinguish the resulting partitions

	Kohonen (1)	Kmeans (2)	Condorcet (3)	Cross product covariances (Pearson) (4)
Most significant variables	partner, weight, g1, r1, g4, s5 (2 and 6 clusters)	partner, sex, g4, r6, g13, g5, s5 (2 clusters)	s5, s2, s1, s6, r13, partner, g6, g7 (9 clusters)	neck, weight, age, alcohol

Table 2. Significance ranking of input variables for different methods

	Logistic regression (1)	Linear regression (2)	C4.5 rule induction (3)	Back propagation neural network (4)
Most significant variables	neck, g1, partner, s9, s8, s7, s6, s10, waist, r12, r2, r5, r6, g2, g6	g8, partner, waist, hip, weight	r3, r2, waist, age, weight, g1	sex, r15, g10, g1, r9, r1,hta,tabacco, height, alcohol, weight, r3, r8, s7, g5, s10, r2, r5

6. WOWA AGGREGATION

Nettleton, in [4] defined, jointly with the medical expert, a most significant subset of 15 variables for apnea diagnosis. This was: {age(0.5, E), sex(0.7, E), weight(0.7, M), imc(0.7, M), neck circumference(1.0, E), alcohol(0.5, M), hta(0.7, E), r1(0.9, H), r2(0.9, H), r11(0.9, M), r13(0.9, M), s3(1.0, M), s4(1.0, M), s5(1.0, M), s6(1.0, M)} with figures in brackets being the respective relevance and reliability weighting profiles, also assigned by the medical expert. The reliability weights are coded as indicated in Section 3.2. Thus each variable has its own characteristic vector of reliability weights which emphasise the more credible responses to a question and dampen the less credible/certain ones. For example, a response of 'always', (assigned in the data the numeric value '5') to question **s6** may be unusual and therefore we would lower its reliability rating, by giving it a smaller proportion of the ω weight. The above reliability and relevance values have to be prepared for input to WOWA so that they sum to 1 for each case. The values are reduced proportionately to achieve this. The normalised value is converted into the relevance and reliability weights for WOWA which respectively sum to 1 for all variables, that is , $\Sigma\rho = 1$, $\Sigma\omega = 1$, where ρ is relevance and ω is reliability.

With reference to Table 3 (below) we can see a favourable result for diagnosis of positive cases and a good result for negative cases, in comparison with the methods used in the literature[1][3].

Table 3. Correlation of WOWA with Apnea Diagnosis for three different weight assignment methods for reliability and relevance

Weight assignment method	Expert assignment of weights	Data analysis assignment of weights	Expert + data analysis assignment
Diagnosis of positive cases	0,75	0,78	0,81
Diagnosis of negative cases	0,65	0,61	0,67

7. DISCUSSION OF RESULTS AND CONCLUSIONS

The resulting consensus from all data analysis methods (tables 1 and 2) indicated the following 9 most significant variables and their corresponding relevance and reliability weighting profiles: {partner(0.70, E), weight(0.70, M), neck(0.92, E), g1(0.68, M), s5(0.95, M), sex(0.7, E), r15(0.60, M), hta(0.67, E), r5(0.90, M)}. Question **g1** is*: "how many hours do you normally sleep?"*; question **s5** is *"do you fall asleep while driving on the motorway?"*; question **r15** is *"do you have lapses of memory or loss of concentration"* and question **r5** is *"have you noticed an increase in the intensity of your snoring recently?"*. Executing the WOWA aggregation with the above input weighting vectors and the 154 patients case data rows, produced the results of table 3. In table 3 the output aggregated value produced by WOWA has been correlated with the binary value which represents the apnea diagnosis.

In this article we have described a method for establishing the relevance and reliability weights needed by the WOWA aggregation operator, applying them to complex data of a real medical problem. With the WOWA aggregation method to apnea diagnosis we can include relevance and reliability information in a more precise manner, to improve the success rate for correct diagnosis.

The authors are grateful to the Medical Faculty of the University of Barcelona, and to Vicenc Torra of the Institute for Investigation in Artificial Intelligence, Bellaterra, Spain, for their collaboration.

References

[1] Hoffstein, V., Szalai J.P. *"Predictive value of clinical features in diagnosing obstructive sleep apnea"*. Sleep 1993; 16: 118-122.

[2] IBM Data Management Solutions White Paper. IBM Corp., 1996.

[3] Katz I., Stradling J., Slutsky A.S., et al. *"Do patients with sleep apnea have thick necks?"* American Review of Respiratory Diseases, 1990; 141: 1228-1231.

[4] Nettleton, D.F., *"Attribute fusion using a heterogeneous representation of crisp and fuzzy medical data"*. Eighth International Fuzzy Systems Association World Congress. Taipei, Taiwan, 1999.

[5] Nettleton, D.F., L. Hernandez. *"Improving questionnaire screening of sleep apnea cases using fuzzy knowledge representation and aggregation techniques"*. Research Report Ref. LSI-99-29-R, Dept. Computer Languages and Systems, University Polytechnic of Catalunya, Spain (1999).

[6] Torra, V., *"The Weighted OWA Operator"*. International Journal of Intelligent Systems, Vol. 12, 153-166. John Wiley & Sons (1997)

[7] Yager, Ronald R. *"Families of OWA operators"*. Fuzzy Sets and Systems 59 (1993) pp125-148.