33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A rule-based automatic sleep staging method

Sheng-Fu Liang, Chih-En Kuo, Yu-Han Hu, and Yu-Shian Cheng

Abstract—**In this paper, a rule-based automatic sleep staging method was proposed. Twelve features, including temporal and spectrum analyses of the EEG, EOG, and EMG signals, were utilized. Normalization was applied to each feature to reduce the effect of individual variability. A hierarchical decision tree, with fourteen rules, was constructed for sleep stage classification. Finally, a smoothing process considering the temporal contextual information was applied for the continuity. The average accuracy and kappa coefficient of the proposed method applied to the all night polysomnography (PSG) of twenty subjects compared with the manual scorings reached 86.5% and 0.78, respectively. This method can assist the clinical staff reduce the time required for sleep scoring in the future.**

Index Terms—**Automatic sleep staging, decision tree, PSG.**

I. INTRODUCTION

Human beings spend approximately one third of their life sleeping. Sleep diseases, such as insomnia and obstructive sleep apnea, seriously affect patients' quality of life. Without restrictive criteria, the prevalence of insomnia symptoms is approximately 33% in the general population [1].

Obstructive sleep apnea affects over 2% of adult women and 4% of adult men [2]. These sleep issues may cause daytime sleepiness, irritability, depression, anxiety or even death.

For the diagnosis of sleep issues, all night polysomnographic (PSG) recordings, including electroencephalogram (EEG), electrooculogram (EOG) and electromyogram (EMG), are usually taken from the patients and the recordings are scored by a well-trained expert according to the Rechtschaffen & Kales (R&K) rules presented in 1968 [3]. According to the R&K rules, each epoch (i.e., 30 s of data) is classified into one of the sleep stages, including wakefulness (Wake), non-rapid eye movement (stages 1-4, from light to deep sleep) and rapid eye movement (REM). Recently, stages 3 and 4 were combined and are now known as the slow wave sleep stage (SWS) [4].

Because visual sleep scoring is a time consuming and subjective process, automatic sleep staging methods, including rule-based methods [5-6], numerical classification methods [7-8] and the hybrid system, that take advantage of both approaches [9], were developed. In rule-based methods, signal information and human knowledge are combined to deduce a reasonable sleep state. Rule-based methods required the detection of specific patterns, such as K-complexes, sleep spindles in EEG and rapid eye movements in EOG. However, it may be a time consuming task to construct the system, especially with human knowledge. In contrast, numerical classification methods do not require a set of rules or any human knowledge. Spectral analysis was commonly used for feature extraction. However, without human knowledge and microstructure pattern recognition, some situations may not be taken into consideration, such as the transition between S1 and S2. Although hybrid systems take advantages of both methods, they are more complicated to implement and the results may only show only incremental improvement over the rule-based methods. The overall agreement of these methods is in the range of 80% to 85%.

In this paper, a rule-based sleep staging system was proposed. The central EEG (C3-A2), the difference between two EOG and the chin EMG were analyzed, and 12 features were extracted for each 30-s epoch. The features were normalized to reduce the individual variability, so that the system parameters were subject-independent. A distribution distance (DD) measure was proposed to extract the representative features for each stage in order to construct the decision tree and finally, fourteen rules were constructed. For performance evaluation, the proposed method was applied to the all night PSG of twenty subjects for sleep staging, and the results were compared with the manual scorings of the expert. In addition to the accuracy, the kappa coefficient [10] was also analyzed to demonstrate the robustness of the proposed method.

II. MATERIALS AND METHODS

A. Subjects and recordings

All-night polysomnographic sleep recordings were obtained from 20 healthy subjects (12 males and 8 females) ranging from 19 to 23 years in age (mean = 21.2 ± 1.1). The data from three subjects were used to generate the system, and data from all 20 subjects were used for testing. The recordings included six EEG channels (F3-A2, F4-A1, C3-A2, C4-A1, P3-A2, and P4-A1, according to the international 10-20 standard system), two EOG channels (positioned 1 cm lateral to the left and right outer canthi), and a chin EMG channel (Siesta 802 PSG, Compumedics, Inc.). The sampling rate was 1 KHz

S.-F. Liang is with the Department of Computer Science and Information Engineering & the Institute of Medical Informatics, National Cheng Kung University, Tainan 701, Taiwan (phone: 886-6-2757575 Ext: 62549; fax: 886-6-2747076; e-mail: sfliang@mail.ncku.edu.tw).

C.-E Kuo, Y-H Hu, and Y-S Cheng are with the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan (e-mail: chihen.kuo@gmail.com; hangohu@gmail.com; yukimcheng@gmail.com).

All correspondence should be addressed to Professor S.-F. Liang.

This work was supported by the National Science Council of Taiwan under grant NSC 98-2221-E-006 -161- MY3.

with 16-bit resolution. The 20 PSG sleep recordings were visually scored by a sleep specialist using the R&K rules with a 30-s interval (termed the epoch).

B. Feature extraction

The automatic sleep staging system analyzes the data for three channels: the central EEG (C3-A2), the difference of the two EOGs, and the chin EMG. After downsampling the signals to 256 Hz, the EEG and EOG data were filtered with an eighth-order Butterworth band-pass filter with a cutoff frequency of 0.5–30 Hz, and the EMG data were filtered with a 5–100 Hz eighth-order Butterworth band-pass filter. The continuous time signals were segmented with every 30-s epoch.

Before extracting spectral features, the signal was segmented into non-overlapping intervals of 2 seconds for a 512-point fast Fourier transformation (FFT) calculation. The spectrums corresponding to the 15 2-s segments were averaged to represent the spectrum for a 30-s epoch. Table I lists the 12 features used in this paper [7-8, 11-12].

TABLE I
The features for sleep scoring

No.	Type	Feature	Source	Label
1	PS	Total power of 0-30 Hz	EEG	0-30 E
2	PS	Total power of 0-30 Hz	EMG	0-30 M
3	PR	0-4 Hz/0-30 Hz	EEG	0-4 E
4	PR	8-13 Hz/0-30 Hz	EEG	8-13 E
5	PR	22-30 Hz/0-30 Hz	EEG	22-30 E
6	PR	0-4 Hz/0-30 Hz	EOG	0-4 O
7	SF	Mean frequency of 0-30 Hz	EEG	Mean(fre.) E
8	SF	Mean frequency of 0-30 Hz	EMG	Mean(fre.) M
9	DR	Alpha ratio	EEG	Alpha E
10	DR	Spindle ratio	EEG	Spindle E
11	DR	SWS ratio	EEG	SWS E
12	EMG energy	Mean amplitude	EMG	Amp M

* PS(=Power spectrum), PR(=Power ratio), SF(=Spectral frequency), DR(=Duration ratio)

After feature extraction, normalization of features was employed to reduce the effects of individual variability. For each feature, the mean of maximal 10% data was calculated as the maximum value of the feature, and the mean of minimal 10% data was calculated as the minimum value of the feature for the subject. This process can prevent extremely high or low values from influencing any conclusions.

C. The structure of the decision tree

The design concept for the developed decision tree for sleep stage classification was to separate the 30-s epochs into two different clusters for each decision node from top to bottom. Each decision node tended to separate two different stages. There are a total of 13 decision nodes in the proposed decision tree, as shown in Fig. 1.

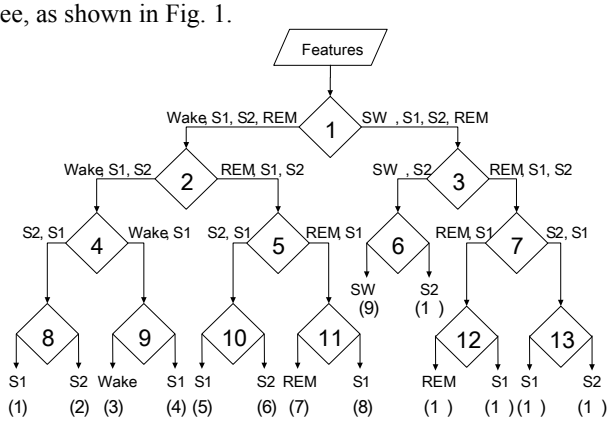

Fig 1. Diagram of the proposed decision tree. There are fourteen decision rules and each epoch was classified as one of five sleep stages, including Wake, S1, S2, SWS, and REM.

The stages that were easier to identify were processed at the nodes in the upper layers, while the stages that were more difficult to distinguish were processed at the nodes in the lower layers. According to the R&K rules, Wake and SWS differ the most, so they are separated in node 1. S1, S2, and REM are more similar, so these stages were classified in the final layer.

Table II presents the operational details of each decision node. The two stages to be separated and the utilized features for each node are given. Taking node 1 as an example, two features, the alpha band of the EEG and the EEG signal from 8-13 Hz, were used to separate the Wake and SWS stages. According to Fig. 1 and Table II, if the feature vectors of an epoch fit the splitting predicate in a decision node, the epoch goes in the direction of the left branch; otherwise, the epoch follows the right branch. After passing through three or four decision nodes, each epoch passes through one of the 14 rules of the decision tree. The final result is the stage that the epoch belongs to.

TABLE II
The details of each decision node

Node No.	Separate stages	Used features and the predicate
1	Wake / SWS	Alpha E > Th & 8-13 E > Th
2	Wake / REM	Alpha E > Th & 0-30 M > Th
3	SWS / REM, S1	0-4 E > Th & 22-30 E < Th
4	S2 / Wake	0-4 E > Th
5	REM / S2	0-30 E < Th & Spindle E < Th & SWS E < Th
6	SWS / S2	0-30 E > Th & Spindle E < Th & & SWS E > Th & 0-4 O > Th
7	REM / S2	0-30 E < Th & Spindle E < Th & SWS E < Th
8	S1 / S2	0-4 E < Th & Spindle E < Th
9	Wake / S1	Mean(fre.) E > Th & Mean(fre.) E > Th
10	REM / S1	Amp M < Th
11	S1 / S2	0-4 E < Th & Spindle E < Th

12	REM / S1	Amp M < Th
13	S1 / S2	0-4 E < Th & Spindle E < Th

D. Predicate in each decision node

After the elementary construction of the decision tree, the two following steps needed to be completed: (1) selecting the proper features for the decision nodes and (2) setting the thresholds of the selected features as the splitting predicates. For the first step, a distribution distance (DD) measure was proposed to select the effective features for each node. At each node, assuming the two stages were separated into stages *A* and *B*, the DD measure was calculated with respect to these two stages for each feature. The means and the standard deviations (SD) of the analyzed feature corresponding to stages *A* and *B* were ($\overline{A}$, $\overline{B}$) and (σ_A, σ_B), respectively. The distribution distance (DD) of the feature with respect to *A* and *B* was calculated using the following equation:

$$DD(A,B)=\begin{cases}1-\dfrac{\sigma_A+\sigma_B}{2\left|\overline{A}-\overline{B}\right|} & \text{if } \sigma_A+\sigma_B \le 2\left|\overline{A}-\overline{B}\right| \\ 0 & \text{else}\end{cases} \quad (1)$$

A feature with a large DD value indicates that the values for this feature in stages A and B differ to a larger extent. The DD analysis can be used for selecting proper features for each node. The results for feature selection for each node by DD analysis are shown in Table II.

For step 2, the threshold for the feature was determined by equation shown below.

$$Th=\frac{1}{2}\cdot[(mean(the\ higher)-\frac{1}{2}\cdot Std(the\ higher)) + (mean(the\ lower)+\frac{1}{2}\cdot Std(the\ lower))] \quad (2)$$

E. Classification

All 20 subjects were used for testing the system. After the same preprocessing procedure as the training data, the 13 features (Table I) described above were extracted for each epoch. The classification process comprises four steps: 1) movement epochs detection, 2) staging with the presented rule-based decision tree, 3) contextual rule smoothing, and 4) movement epochs elimination.

1) Movement epochs detection

The movement (MT) epochs were detected with the following rules: if over 1/3 of the epoch of the absolute EMG amplitude was higher than double of Th1, or if the difference between the two consecutive epochs was higher than Th2, the epoch was defined as MT. Th1 is the mean value of the feature Amp M in Wake epochs of the training data, and Th2 is the mean value of the Amp M difference between every two consecutive epochs of the training data.

2) Staging with the presented rule-based decision tree

The non-MT epochs were classified using the rule-based decision tree, shown in Fig. 1. There were 14 rules in the tree, and each epoch was classified as one of five sleep-wake stages, including Wake, S1, S2, SWS, and REM.

3) Contextual rule smoothing

After classifying the sleep stage using the decision tree, a smoothing process, considering the temporal contextual information, was applied for continuity [4]. These rules refer to the relationship between epochs prior to and posterior to the current epoch.

4) Movement epochs elimination

After smoothing, an elimination procedure was used on those MT epochs with the AASM scoring methods [4]. The final result of staging (hypnogram) was still characterized by five stages (Wake, S1, S2, SWS, and REM).

III. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed sleep staging method, 20 all-night PSG recordings from 20 subjects were used for testing. The performance was evaluated in the following aspects: (1) the average performance of the method without smoothing (Table III) and (2) with smoothing (Table IV) for all subjects. The performance was evaluated by computing the sensitivity (SE) [13] of each sleep stage using the proposed method and the average accuracy. Sensitivity measures the proportion of actual positives which are correctly identified as such. In addition, Cohen's kappa coefficient [14] was also calculated for each subject to assess the robustness of our system. Cohen's kappa coefficient (κ) is a statistical measure of inter-rater agreement among two or more raters. It is generally thought to be a more robust measure than simple percent agreement calculations because κ takes into account agreements that occur by chance.

A. Global performance

Table III and Table IV show the confusion matrices of the five-stage epoch classification by the proposed automatic staging, with or without smoothing, versus manual scoring. The rows and columns are the results staged by the expert and our system, respectively. Unknown epochs corresponding to the unidentified signals and the movement epochs are not taken into account here. The testing dataset for estimation was comprised of 20 PSGs with 16,976 30-second epochs. As shown in Table III, the overall sensitivity was 83.01% for the proposed rule-based method. The sensitivities for all stages, except for S1, were higher than 82%. In addition, the sensitivities for Wake and SWS were close to 90%. Obviously, misclassifications of each stage generally occur between adjacent stages. This may be because the sleep process is continuous; the adjacent stages are more likely with each other than other stages.

As shown in Table IV, the performance of the rule-based method can be further improved by smoothing the results. The overall sensitivity between the expert and our system was 86.58%. The results in S2 and REM were increased by 4.78% and 2.61%, respectively, with the use of the smoothing technique. The sensitivity of SWS was enhanced to over 90%.

B. Cohen's kappa performance

Cohen's kappa coefficient (κ) is a statistical measure of inter-rater agreement among two or more raters. It is generally

thought to be a more robust measure than simple percent agreement calculations because κ takes into account the agreements that occur by chance. The average kappa value was $\kappa = 0.79 \pm 0.04$.It was observed that the average kappa (0.79) of our system showed substantial reliability.

TABLE III
Confusion Table for Comparison between Computer Scoring without Smoothing and Visual Scoring for All 20 Subjects

		Computer						
		Wake	S1	S2	SWS	REM	Total	SE %
Expert	Wake	**436**	28	17	0	7	488	89.34
	S1	98	**192**	153	1	183	627	30.62
	S2	318	327	**7650**	360	597	9252	82.68
	SWS	9	0	317	**2695**	0	3021	89.21
	REM	51	217	200	0	**3120**	3588	86.96
	Overall						**16976**	**83.01**

* SE = sensitivity

TABLE IV
Confusion Table for Comparison between Computer Scoring with Smoothing and Visual Scoring for All 20 Subjects

		Computer						
		Wake	S1	S2	SWS	REM	Total	SE %
Expert	Wake	**436**	27	18	0	7	488	89.34
	S1	98	**208**	167	1	153	627	33.17
	S2	318	246	**8092**	189	407	9252	87.46
	SWS	9	0	264	**2748**	0	3021	90.96
	REM	39	183	152	0	**3214**	3588	89.57
	Overall						**16976**	**86.58**

* SE = sensitivity

IV. Discussion and Conclusion

In this paper, an automated sleep staging system, based on the multi-rule decision tree, was proposed. The main idea of the tree structure is to separate the stages with the largest variation first and the similar stages last. The distribution distance measure was proposed and combined with the R&K criteria to construct the tree. The innovative aspect of our method is that it combines the advantages of numerical methods and traditional rule-based methods, yet eliminates the shortcomings of both methods. The overall sensitivity of the proposed method applied to the PSGs from 20 subjects reached 86.58%. Our method correctly classified the vast majority of epochs. Except for S1, the sensitivities of all the stages were higher than 87%. In addition, the average Cohen's kappa κ for all subjects was 0.79 (S.D. = 0.04) and represented a substantial agreement, as described by [14], of our method compared to the scoring of the expert.

Another advantage of our method is that a simple threshold was set to separate the stages. Compared to other classifiers, such as neural networks [7] or linear discriminator analysis [15], the thresholding is less computationally complex. In the future, these results for young healthy individuals need to be extended to older healthy individuals and patients. This method can be applied to clinical practice to work with experts to reduce the scoring time. Moreover, we will combine this algorithm with hardware to develop a portable polysomnography system for home healthcare.

References

[1] M.M. Ohayon, "Epidemiology of insomnia: what we know and what we still need to learn," *Sleep Medicine Reviews*, vol. 6, pp. 97–111, 2002.

[2] M.W. Mahowald and C.H. Schenck, "Insights from studying human sleep disorders," *Nature*, vol. 437, pp. 1279–1285, 2005.

[3] Rechtschaffen A., A. Kales, eds. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*, CA: BI/BR, Los Angeles, 1968.

[4] C. Iber, S. Ancoli-Israel, A. Chesson, and S.F. Quan, *"The AASM manual for the scoring of sleep and associated events,"* 1st ed. Westchester, IL: American Academy of Sleep Medicine, 2007.

[5] J.R. Smith, M. Negin, and A.H. Nevis, "Automatic analysis of sleep electroencephalograms by hybrid computation," *IEEE Transactions on Systems Science, and Cybernetics*, SSC-5: 278–284, 1969.

[6] G.H. Jansen and B.M. Dawant, "Knowledge-based approach to sleep EEG analysis-a feasibility study," *IEEE Transaction on Biomedical Engineering*, vol. 36, pp. 510–518, 1989.

[7] N. Schaltenbrand, R. Lengelle, M. Toussaint, et al., "Sleep stage scoring using the neural network model: Comparison between visual and automatic analysis in normal subjects and patients," *Sleep*, vol. 19, pp. 26–35, 1996.

[8] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Transaction on Biomedical Engineering*, vol. 48, pp. 1412–1423, 2001.

[9] H.J. Park, K.S. Park and D.U. Jeong, "Hybrid neural-network and rule-based expert system for automatic sleep stage scoring," in Proc. 22th Annual EMBS Int. Conf., Chicago, pp. 1316-1319, 2000.

[10] J.R. Landis and G.G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159–174, 1977.

[11] C. Berthomier, J. Prado, and O. Benoit, "Automatic sleep EEG analysis using filter banks," *Biomedical Sciences Instrumentation*, vol. 35, pp. 241–246, 1999.

[12] F. Duman, A. Erdamar, O. Erogul, Z. Telatar, and S. Yetkin, "Efficient sleep spindle detection algorithm with decision tree," *Expert Systems with Applications*, vol. 36, pp. 9980–9985, 2009.

[13] D.G. Altman, *Practical statistics for medical research*, Chapman & Hall, London, 1991.

[14] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp. 37–46, 1960.

[15] P. Anderer, G. Gruber, S. Parapatics, et al, "An e-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24x7 utilizing the Siesta database," *Neuropsychobiology*, vol. 51, pp. 115-133, 2005.