32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Decision-Making System for Orthodontic Treatment Planning Based on Direct Implementation of Expertise Knowledge

Masakazu Yagi, *Member, IEEE,* Hiroko Ohno, and Kenji Takada

***Abstract*—Development of the decision-making systems has been highly demanded to provide objective evidence for the decisions of experts, especially in medicine, and a variety of systems have been developed by means of the state-of-the-art technology. In orthodontics, there has been no objective criterion for the decisions of whether or not to perform one of the invasive treatments, tooth extraction. Therefore, the prediction system for the extraction-nonextraction decisions was developed by intuitive implementation of expertise knowledge in this study. The system was successfully optimized with respect to knowledge descriptions and an inference algorithm to provide the prediction accuracy of 90.5% and simulations of the decision-making process on the optimized model were performed to obtain the terse representation of the expertise knowledge elements that are assumed to affect the decision-making of experts.**

I. Introduction

Over past few decades, the concept of evidence-based medicine (EBM) has changed the nature of experts' decision-making in the medical/dental domain. It has been widely accepted that it is required to formulate the elaborate decision-making of experts objectively for achievement of EBM. One of the frameworks for realization of the objective formulation is to develop and evaluate a system that emulates the decision-making of experts on a computer. Thus, there have been a number of researches on the development of various kinds of the medical/dental decision-making systems, such as the decision-tree-based [1], [2], fuzzy logic-based [3]-[5] neural networks-based [6], and template-matching-based [7] systems.

In orthodontics, although the procedure of tooth extraction, i.e. removal of a tooth or teeth, is one of the invasive options in a treatment plan, there have not been objective criteria for deciding whether or not to extract teeth. Accordingly, the primary prediction systems for the extraction-nonextraction decisions [8] and the determination of extraction sites in case of extraction [9] were developed and assessed from the clinical viewpoint. Nevertheless, the principal factors to determine the robust nature of the system, that is, the schemes to implement the expertise knowledge in the systems were not investigated in these studies.

Hence, the purposes of this study are 1) to develop a system that predicts the optimum extraction-nonextraction decisions in orthodontic treatment planning and supports the decision-making of orthodontists by optimizing the implementation schemes of expertise knowledge, and 2) to simulate the decision-making process on the optimized system to elucidate the concise representation of the knowledge elements that are sensitive to the decision-making process in the system.

Manuscript received March 31, 2010. This work was supported in part by the Grant-in-Aid for Young Scientists (B) (#19791575) from Japan Society for the Promotion of Science and the Global COE program " in silico medicine" at Osaka University.

M. Yagi is with the Center for Advanced Medical Engineering and Informatics, Osaka University, Osaka, 565-0871, Japan (phone: +81-6-6879-2958, fax: +81-6-6879-2960, e-mail: mgoat@dent.osaka-u.ac.jp).

H. Ohno and K. Takada are with Dept. Orthodontics and Dentofacial Orthopedics, Graduate School of Dentistry, Osaka University, Osaka, 565-0871, Japan (e-mail: {h-ohno, opam}@dent.osaka-u.ac.jp).

II. Material and methods

A. Subjects and material

A hundred and eighty-eight patients (Japanese female; mean age, 17 years 5 months) were randomly selected according to the following selection criteria:

- Completion of orthodontic treatment with 'good treatment outcome' using fixed appliances
- Full permanent dentitions except the third molar teeth
- No congenital and/or acquired dental anomalies of the craniofacial forms or skeletal deformities
- No history of surgical orthodontic treatment.

The condition of 'good treatment outcome' was defined as greater than 70% reduction (improvement) of the peer assessment rating (PAR) index score [10], representing the degree of malocclusion, as a result of orthodontic treatments. The patients were divided into two classes according to the actual treatment, i.e. extraction treatment class (*Ext*, class label=1, n=100) and non-extraction treatment class (*Nonext*, class label=-1, n=88) and their clinical records such as pre-treatment dental casts, lateral and frontal head films were obtained.

B. Basic concept of decision-making process in the system

The decision-making process in the system, i.e. prediction of the optimum decisions, is briefly illustrated in Fig. 1. Some knowledge datasets (KDs), that is, a set of a feature vector representing a case with the factual good treatment outcome and its treatment class (*Ext* or *Nonext*), were stored in the system knowledge database. When an input case is given to the system, a feature vector was generated from the pre-treatment clinical records. The optimum decision-making was predicted by means of the neighboring search in the knowledge database and an inference function, like recalling the similar cases to a new patient in the clinical experience of experts.

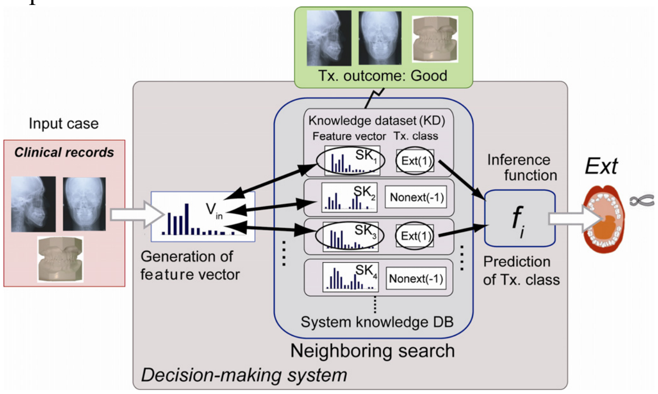

Fig. 1. Conceptual illustration of the prediction algorithm for the optimum decisions.

C. Knowledge descriptions

A total of 27 variables, representing the morphological features of hard and/or soft tissues in the anterior-posterior (SK2, SK3, FMIA, U1 to NA, L1 to NB, L1a-Cli, L1a-Cla, EL-ls, EL-li, OJ, Molar-R, Molar-L), vertical (FMA, ALFH, OB) and lateral (MDL_U and MDL_L) directions, and intra-dental-arch conditions (II_U, II_L, SCD_U, SCD_L, CAL_U/SCD_U, CAL_L/SCD_L, CAW_U/SCD_U, CAW_L/SCD_L, BAL_U*BAW_U/SCD_U, BAL_L*BAW_L/SCD_L), were measured on the clinical records of each patient. Definitions of the variables were described in [5]. Some variables measured on a lateral head film were exemplified in Fig. 2.

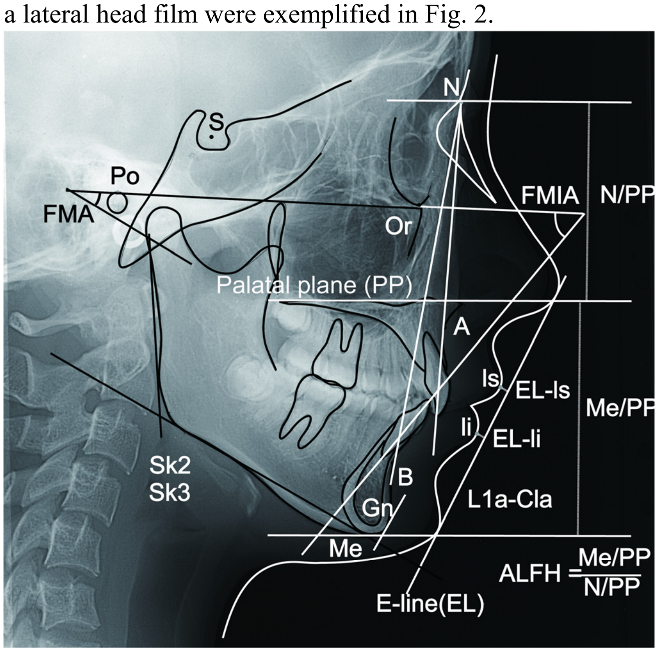

Fig. 2. Examples of the feature variables obtained from a lateral head film.

1) Scaling algorithm for feature variables

The range of each feature variable was scaled by means of three types of transformation functions as shown below:

- Normalization (*N*) using mean and standard deviation of the variables for all the subjects,
- Linear scaling (*LS*) of each variable to the range of [0, 1],
- Nonlinear scaling (*NLS*) by means of a transformation function that was developed using the normative values employed in the decision-making of orthodontists.

The transformation functions for the variables of OJ (Distance from the upper central incisor tip to the lower incisor labial surface), ALFH (Lower anterior face height), $MDL_{U(L)}$ (Upper and Lower dental midline deviations from the facial midline), II_U (Sum of the distances between the neighboring anatomic contact points from the first molar tooth on a side to that on the other side in upper dental arch) were demonstrated in Fig. 3.

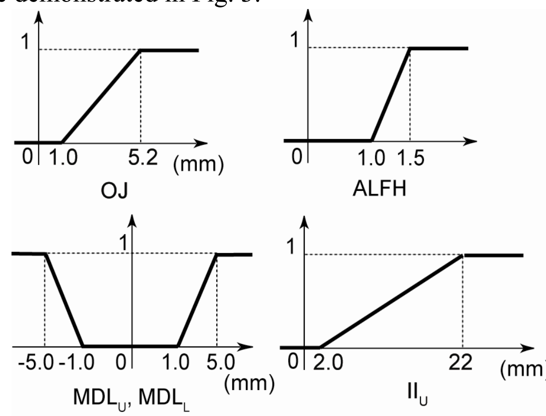

Fig. 3. *NLS*-transformation functions for OJ, ALFH, $MDL_{U(L)}$ and II_U.

2) Generation of knowledge datasets

A total of 1,000 kinds of feature vector spaces that had the vector dimensions ranging from 2 to 27, were generated according to the combinations of the variables that were assumed to affect the decision-making of experts. Sets of a feature vector and its corresponding treatment class label were obtained as the ***raw*** knowledge datasets (KDs).

D. Inference algorithm

In the neighboring search of the decision-making process, the N_m-nearest feature vectors to an input (N_m=1,2,…,15) were selected to obtain the reference KDs in the knowledge database, using the similarity measure as defined in (1).

$$S(j) = \sum_i w_i \left| V_{in}(i) - SK_j(i) \right| \tag{1}$$

where V_{in} and SK_j are feature vectors of an input and j-th KD in the system knowledge database; i is an index of vector element; $W = \{w_i \mid 0 < w_i \leq 5\}$ are the weight coefficients in the calculation of similarity measure. A total of 1,000 patterns of W were determined exploratorily by orthodontists.

The treatment class label $\hat{C}$, i.e. the predicted optimum decision, was determined, using the three kinds of inference functions with the reference KDs:

- Majority voting (*MV*),
- Inference algorithm based on the orders of reference KDs (*IO*) as in (2),
- Inference algorithm based on the aforementioned similarity measures between an input and the reference KDs (*IS*) as in (3).

$$\hat{C} = \mathrm{sgn}(\sum_i (N_m - i + 1) \cdot C_{sk}(i)) \quad (2)$$

$$\hat{C} = \mathrm{sgn}(\sum_i S(N_m - i + 1) \cdot C_{sk}(i)) \quad (3)$$

where C_{sk} is the treatment class label in the reference KD; S is the similarity measure between an input and the reference KD; i is an index of order in the neighboring search process (i=1,2,…,N_m).

E. System optimization

The prediction performances of the systems were evaluated by means of 10-fold cross-validation to avoid the overfitting problem. In the cross-validation process, the ***raw*** KDs were divided into 10 subsets at random. One of the subsets was tested using the other nine subsets in a trial, and each subset was employed as an input. In a test, it was assigned 'correct' when the predicted treatment class label $\hat{C}$ coincided with the factual one of the input KD. The ratio of the number of 'correct' cases to that of the inputs was obtained in each trial and the mean of the ratios (ROC) for all the trials was computed as the prediction accuracy to optimize the system.

F. Simulation of the decision-making process on the basis of the optimized model

1) Resolution of feature vector elements

The resolution of a feature vector elements was down-sampled linearly from 4-bytes (32-bit) to n_b-bit (n_b=1, 2,…, 8) in order to investigate the reasonable resolution of the variables for representing expertise knowledge in a vector form.

2) Application of learning algorithm to generate the knowledge datasets (KDs) in the system

The generalized Lloyd algorithm [11] was applied to the feature vectors that belong to each treatment class (*Ext* or *Nonext*) separately. The codes, namely, the typical patterns of the feature vectors, were generated to form the ***learned*** KDs, i.e. sets of the n_c-codes (n_c=3, 5, 10, 15, 25) per treatment class and its associate class label. In each trial of the cross-validation process, one of the subsets was employed for test with the ***learned*** KDs that were generated from the other nine subsets.

III. Results and Discussions

In the optimized system, the *NLS* scaling and *IS* inference algorithms were employed with N_m=7 to provide the highest ROC of 90.5 %. The feature vector elements and their corresponding weights adopted in the system are shown in Table I. The combinations of the feature variables that composed of a vector form well matched with the perception of orthodontists as the feature elements that are sensitive to the extraction-non-extraction decisions.

The prediction accuracy of the systems with the scaling algorithms of *N*, *LS* and *NLS* were compared in Fig. 4. The system with the scaling algorithm *NLS* demonstrated the highest performance of prediction accuracy. It indicates the successful implementation of expertise knowledge in the algorithm.

TABLE I.
FEATURE VECTOR ELEMENTS AND THEIR WEIGHTS WI EMPLOYED IN THE OPTIMIZED MODEL.

FV	w_i	FV	w_i
AP direction		***Lateral direction***	
SK2	0.25	MDL_U	1.00
SK3	0.50	MDL_L	1.00
FMIA	0.50	***Intra-Dental-Arch Condition***	
U1 to NA	1.00	II_U	1.00
L1 to NB	1.00	II_L	1.00
EL-ls	0.25	SCD_U	0.50
EL-li	1.00	SCD_L	0.30
OJ	2.00	CAL_U/SCD_U	1.00
Molar-R	1.00	CAL_L/SCD_L	1.00
Molar-L	1.00	CAW_U/SCD_U	1.00
Vertical direction		CAW_L/SCD_L	1.00
ALFH	0.50	$BAL_U \cdot BAW_U/SCD_U$	2.00
OB	0.50	$BAL_L \cdot BAW_L/SCD_L$	1.00
FMA	0.25		

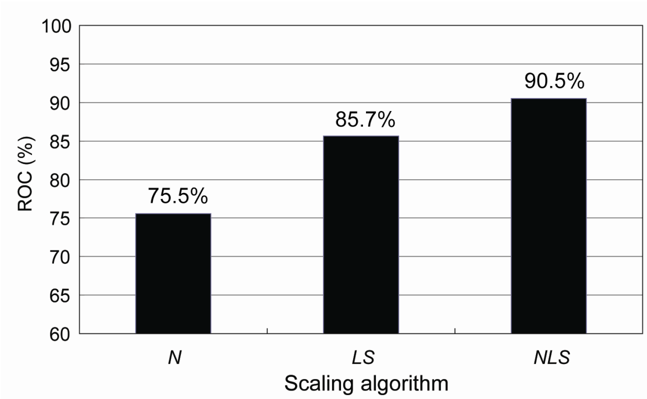

Fig. 4. Prediction results by three kinds of scaling algorithms (*N*, *LS*, and *NLS*).

The inference algorithms of *MV*, *IO*, and *IA* were investigated as shown in Fig. 5. The system that employed the inference algorithm *IS* performed the decision-making process with the highest ROC. Thus, the robust nature of prediction performance in the system was enhanced by *IS* algorithm that determined the degree of contribution for each reference KD to the decision-making based on the similarity measures in the prediction process.

Fig. 6 designates the simulation results of the decision-making process on the optimized model with respect to the resolution of feature vector elements, i.e. the number of bits (n_b) that were employed to represent a feature vector element. The prediction performance that was the most similar to the experts (ROC=90.5%) was achieved with the feature vectors in the 32-bit resolution, whereas there was the local peak (ROC=88.9%) when n_b=4. Despite reduction of amount of information by one-eighth, the 1.6% decrease of ROC was observed. It reveals that each feature vector element, the fragment of knowledge description for prediction of the optimum extraction-nonextraction decisions, could be represented in an approximately 4-bit resolution effectively.

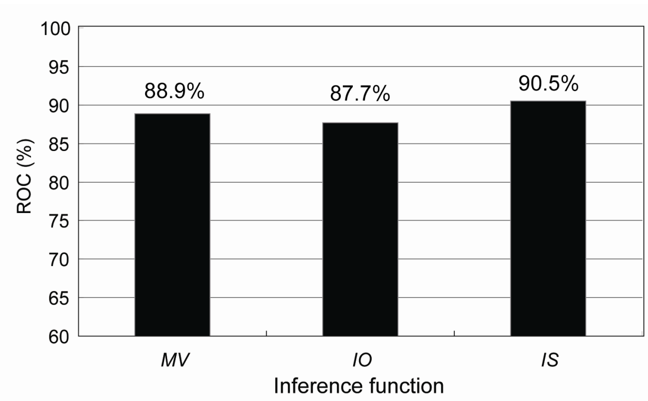

Fig. 5. Prediction accuracy of the decision making systems using the inference algorithms of *MV*, *IO* and *IS*.

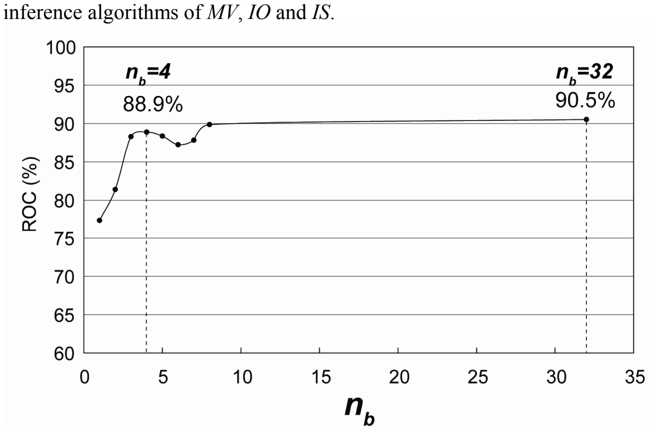

Fig.6. ROCs of the system as a function of the resolution of a feature vector element (***n_b***).

The prediction accuracy of the system as a function of the number of KDs (***learned*** and ***raw)*** in the system knowledge database (2*n_c) was demonstrated in Fig. 7.

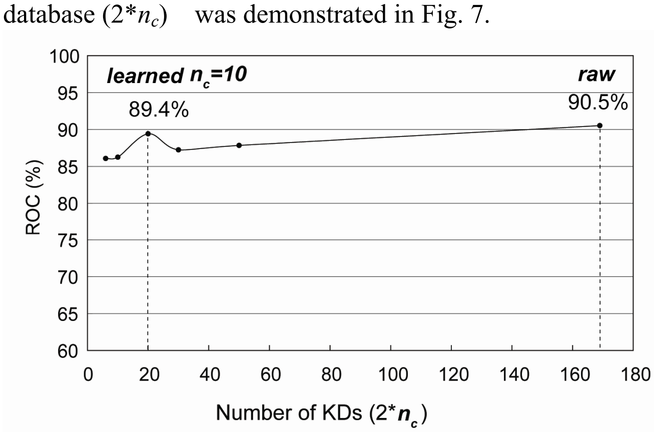

Fig.7. ROCs of the systems with the different number of KDSs (***learned*** and ***raw***).

The system with the more KDs did not necessarily provide the higher ROC. The local peak was observed with slight decrease of ROC (ROC=89.4%) as compared with the ***raw*** KDs when twenty of KDs (n_c=10) were employed. It suggests that the ten KDs per treatment class (*Ext* or *Nonext*) well represented the distribution of all the cases in the feature vector space with less than one-eighth number of KDs and could be the concise prototypal knowledge descriptions for the decision-making system.

IV. Conclusion

The decision-making system for orthodontic treatment planning was successfully developed by implementing the expertise knowledge in the prediction algorithms for the optimum decisions intuitively to demonstrate the prediction accuracy of 90.5%. The simulation of the decision-making process on the optimized system was performed and the resolution of the descriptions and the number of the prototypes for expertise knowledge were formulated for emulation of the orthodontic decision-making process regarding tooth extraction.

References

[1] I.D. Brown., S.R. Adams, C.D. Stephens, S.J. Erritt, and J.H. Sims-Williams, "The initial use of a computer-controlled expert system in the treatment planning of class II division 1 malocclusion," *British J of Orthod*, vol. 18, pp. 1-7, February 1991.

[2] C. Stephens and N. Mackin, "The validation of an orthodontic expert system rule-base for fixed appliance treatment plannning," *Eur. J. Orthod*, vol. 20, pp. 569-578. Oct., 1998.

[3] H. Ying, F. Lin, R.D. MacArthur, J.A. Cohn, D.C. Barth-Jones, Y. Hong, and L.R. Crane, "A Self-Learning Fuzzy Discrete Event System for HIV/AIDS Treatment Regimen Selection," *IEEE Trans. Syst. Man Cybern. Part B Cybern.* , vol.37, no.4, pp.966-979, Aug. 2007

[4] N.H. Phuong, N.B. Tu, L. Ding, and K. Hirota, "Case based reasoning using fuzzy set theory and the importance of features in medicine," *Joint 9th IFSA World Congress and 20th NAFIPS International Conference, 2001*, vol.2, pp.872-876, Jul., 2001.

[5] H. Noroozi, "Introduction of a New Orthodontic Treatment PlanningSoftware; a fuzzy logic expert system," *Int J Orthod Milwaukee*, vol. 17, 2006.

[6] D. Anthony, E. Hines, D. Taylor, and J. Barham, "An investigation into the use of neural networks for an expert system in nuclear medicine image analysis," *Third International Conference on Image Processing and its Applications*, pp.338-342, Jul., 1989

[7] R.M. Hammond and T.J. Freer, "Application of a case-based expert system to orthodontic diagnosis and treatment planning," *Aust J. Orthod*,.vol. 14, pp.229-234, Mar., 1997.

[8] K. Takada, M. Yagi and E. Horiguchi, "Computational formulation of orthodontic tooth-extraction decisions part I: to extract or not to extract," *Angle. Orthod*, vol. 14, pp. 885-891, Sep., 2009.

[9] M. Yagi, H. Ohno and K. Takada, "Computational Formulation of Orthodontic Tooth-Extraction Decisions Part II: Which tooth should be extracted?" *Angle Orthod*, vol.14, pp. 892-898, Sep., 2009.

[10] R.M. Little, "The irregularity index: a quantitative score of mandibular anterior alignment," Am. J. Orthod. vol. 68, pp. 554-563, Nov., 1975.

[11] S.P. Lloyd, "Least Squares Quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129-137, 1982.