Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 47 (2) 215-220 (2015)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Schematic Cycle of Case-Based Reasoning Technique Implements in Clinical Decision Support Systems Used for Diagnosis of Liver Disease

S. S. ZIA[++], P. AKHTAR*, T. J. A. MUGHAL

Faculty of Engineering, Sciences and Technology (FEST), Hamdard University, Karachi- Pakistan.

Received: 10th January, 2015 Revised 27th May 2015

Abstract: In this research, schematic cycle of case-based reasoning technique is used for medical data diagnosis & classification. Similarity measurement is the major weakness in executing the CBR technique accurately and efficiently. With the intention of resolving this difficulty we have proposed a similarity algorithm by using Average Weighted Euclidian distance method which calculates distance between the new case and stored cases and then using similarity function to retrieve closely related cases from the case repository. Once the case is retrieved, reinstantiation strategy of case adaption phase is being used for adapting the suggesting way out of the current problem. Then proposed solution of the current problem is revised by the medical expert in the revise phase of the CBR technique. Finally the revised solution is updated or stored into the case repository as a new case. This research performed data analysis on the Indian Liver Patient dataset from UCI Machine Learning Repository is used for diagnosis & classification.

Keywords: Case-based reasoning, Similarity measurement, Euclidian distance, Similarity function, Diagnosis, Classification.

1. INTRODUCTION

Categorizing medical data sets for the purpose of analyzing the specific situation is a complex task. In medical domain large extent of data is produced which relates to patient's records, resources of hospitals, medical apparatus etc (Holzinger *et al.*, 2014). These huge consignments of medical data play an important role for mining useful knowledge that aid medical practitioners in the phase of diagnosis and treatment. In medical domain it takes years of training and practice for a doctor to make proper and correct decision during patient evaluation (Sox, *et al.*, 2013). Clinicians may start their practices by considering some previous experiences (solved cases) which could be used as a reference to handle the new situation. Thus a new experience (new case) has been created which enriches the clinician's set of experiences.

Since last two decades, number of data mining techniques is considered to be functional or useful for the diagnosis purpose (Begum, *et al.* 2009). DM techniques bring together the data analysis methods with different algorithms to process large and complex sets of data (Kantardzic 2011). For solving the DM tasks, different algorithms and technique are employed namely: Support vector machine, neural network, fuzzy logic, rough set theory, association rule mining, decision tree, genetic algorithms and case based reasoning (Salem 2011).

Case based reasoning is one of the famous cognitive science based procedure for medical decision support systems which guides while evaluating the circumstances (Saleem 2007). Categorizing medical data sets for the purpose of analyzing the specific situation is a complex task. CBR technique that comes under the artificial intelligence area allows medical practitioners to match previous cases stored in the knowledge base (KB) with a new case (Rodríguez-González, *et al.* 2013). (**Fig. 1).**

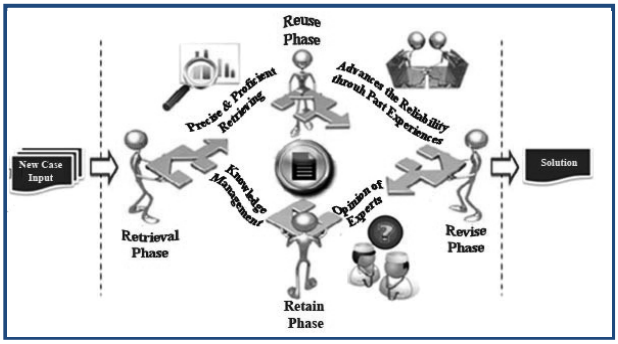

Fig. 1: Schematic cycle of CBR technique

Case based reasoning approach is being widely used during crucial phases of patient-care which include diagnosis and treatment suggestion. CBR is an appropriate method to explore in a medical context where symptoms represent the problem, and diagnosis and treatment represent the solution (Marling, *et al.* 2014).

[++] Correspondence author: Syed Saood Zia, saood_zia@hotmail.com , cell: +92-346-2767788
*Pakistan Navy Engineering College, National University of Science & Technology

The fundamental feature of the CBR is an inter-thread procedure of four phases introduced by Aamodt and Plaza which can be explained by a schematic cycle namely retrieval, reuse, revise and retain (Aamodt *et al.* 1994). Retrieval phase is an initial step which inquires about previous experiences that are similar to the new case. In this phase most similar cases will be retrieved from the case repository on the basis of the new input case. Reuse phase is the second step which is responsible in suggesting a solution for the new case from the available solutions of the cases that were retrieved from the case repository. Solution proposed into the reuse phase will be then revised by an expert (either human or machine). Once the solution is revised by the experts it has to be determined whether to keep this new solution in the case repository in order to facilitate the future diagnosis of new case. Keeping the proposed solution in case repository as a new case is the retain phase of this cycle (Pal *et al.* 2004).

This paper is organized as follows. Section 2 detailing the method and algorithm of case-based reasoning technique used for medical diagnosis & classification problems, while Section 3 presents the experimental analysis and results on Indian Liver Patient Dataset from UCI Machine Learning Repository. Section 4 provides the conclusions and future directions of researches.

2. CASE BASED REASONING

Case based reasoning is an approach to obtain knowledge and make inferences by employing previously stored cases. In Case based reasoning system, knowledge is personified as historical cases. These cases are consisting of detailed description of the problem along with their way out. In order to solve the new case, we compare previously stored case with the new case and then repossess the similar cases. In medical domain we use case based reasoning system for diagnosis the disease as it compare the attributes of current case with the stored cases in knowledge base.

2.1. Case retrieval phase of CBR technique

Retrieval phase is deemed to be fundamental part of the CBR technique. In retrieving phase an efficient way to extract the most similar cases is being used. The case retrieval process identifies to extract out the most suitable cases from the case repository which is closely related to the new given case.

2.1.1. Processing of Retrieval phase

Retrieval phase is the foremost step in the CBR cycle. In this phase retriever extracts out the most similar cases to solve the current problem. Retriever input the new case and compare with the stored cases in case repository. (**Fig. 2).**

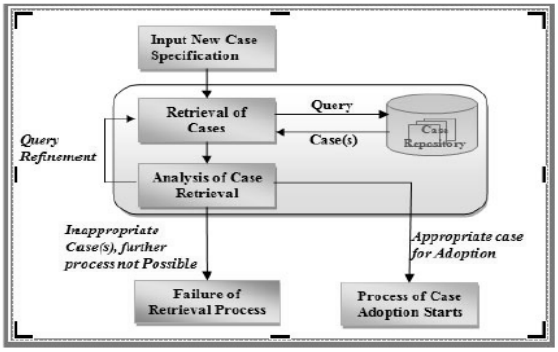

Fig. 2: Case retrieval phase of CBR Cycle

This comparison is carrying out on the basis of similarity. The similarity will allows retrieving one or more cases from the case repository. Once the cases are being retrieved from the case repository, retriever will make some scrutiny to check whether the retrieved case is similar enough to the new input case or there is necessitate making some modifications into the searching parameters to extract again. The case retrieved during this analysis will forward to the case adaptation phase. On the contrary if there is no further appropriate retrieval then concluded as failure of the retrieval process.

2.1.2. Calculate distance between the new case and the old cases

Different algorithms are applied to check similarity between new input case and stored cases in a case repository. Average weighted Euclidian distance is used in our proposed model for computing the distance between the new input case and the stored cases. For easy computation, the range of the distance measures between the new and stored cases can be normalized into 0 to 1. For this purpose, C_{Max} is used for converting the value into normalized form. The weighted Euclidian distance formula for calculating the distance between the input case and the pile up cases that stored in the case library is shown in eq.1 (Zia, *et al.* 2014).

$$d(C_N, C_O) = \frac{\sum_{i=1}^{n} w_i \times \sqrt{\left|\frac{C_{N\,i} - C_{O\,i}}{C_{Max\,i}}\right|^2}}{\sum_{i=1}^{n} w_i} \qquad (1)$$

Where

- $\mathbf{C_N}$ = new input case from clinicians
- $\mathbf{C_O}$ = Stored old cases in a case repository
- $\mathbf{C_{Max}}$ = Maximum value selected from the new input case or previously stored cases for converting it into normalized form.
- **n** = Attributes or Feature value in each case.
- **i** = is an individual or single attribute.
- **W** = Weight of each attribute. These weights determine the importance of each attributes and are assigned by medical experts.

2.1.3. Compute Similarity between the new case and the old cases

Once the distance between the new input case and the stored cases are computed in normalized form then apply similarity measurement function that will shows the most similar cases with the new input case. The calculation performs for computing similarity measurement function shows in eq. 2 and eq. 3 (Zia, *et al.* 2014).

$$\text{Sim}\,(C_N, C_O) = [1 - d(C_N, C_O)] * 100 \quad (2)$$

$$\text{Sim}(C_N, C_O) = \left[1 - \frac{\sum_{i=1}^{n} w_i \times \sqrt{\left|\frac{C_{Ni} - C_{Oi}}{C_{Max\,i}}\right|^2}}{\sum_{i=1}^{n} w_i}\right] * 100 \quad (3)$$

2.1.4. Proposed retrieval algorithm

The proposed retrieval algorithm is processed as follows:

Algorithm: Retrieve Similar Records from the Case Repository	
Input:	Medical Data (Pathological Report of the Patient).
Output:	Retrieve **Similar Records** on the basis of input case.

1. First, Select the major disease of the patient (like Cancer, Cardiac, Nephrology, Orthopedic, Eye etc.).
2. After that, identify the sub-disease of the selected disease for the patient.
3. Assign the threshold value (borderline value) that will help to retrieve the records from the case repository.
4. Retrieve the attributes value of the selected disease.
5. Assign the weights to the given attributes that is based on the importance of each attribute in a selected disease. (Weights are assigned by Medical Experts of the selected disease and that can be changeable from experts to experts).
6. Input the current pathological report data of the patient.
7. Using Average Weighted Euclidean Distance method to calculate the distance (in normalized form) between the input case value and the stored cases. Then compute similarity between the cases.
8. Return Similar Records

2.2. Case Reuse phase of CBR technique

The second step of CBR cycle is case reuse or case adoption phase. This step is reusing one of the retrieved cases from the case repository and returning it as the proposed solution for a current case. Usually, the past solution needs adjustment to fit the new situation. The process of fixing (i.e., adapting) **((Fig. 3)** the old solution is called case adaptation (Pal, *et al.* 2004). An adaption of the obtained solution is required in order to give an accurate result. It calculates the differences between the retrieved case and the current case and then applies algorithms or rules that take the differences onto account to suggest solution. This adaptation could be done by a domain expert. The expert determines if it is reasonable solution to the problem and he/she can modify the solution before approved (Ahmed, *et al.* 2010).

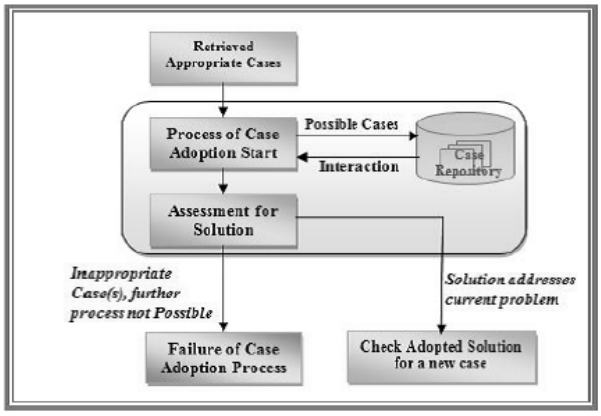

Fig. 3: Case Reuse Phase in CBR Cycle

2.2.1. Case adaption strategies of Case Reuse phase

Subsequently three case adaptation strategies are discussed (Pal *et al.* 2004), (Zia, *et al.* 2014):

(i) **Reinstantiation**: This is a straightforward way of adaptation. In this strategy, simply copied the solution of the most similar extracted case from the case repository and used this solution for the current problem without any modification or adjustment.

(ii) **Substitution:** This strategy restores elements of the previous solution features that unsound with the new input case.

(iii) **Transformation:** In the absence of suitable alternatives, use transformation strategy. In this situation a modified solution is derived.

After retrieving the similar cases from the case repository, the medical expert is now going for the adoption process. In this phase, the authors employ the reinstantiation strategy for adopting the solution of the best similar case and mapped the solution to the new

case. If the similar case not available in the case repository then the medical expert can take their own decision for the new case.

2.3. Process of case revise phase

Once the case is retrieved and adapted, it is proceed to revise stage. In this stage the adapted solution is being analyzed and verified for the accuracy and offered as a validated solution for the new problem case (Pal *et al.* 2004), (Zia, *et al.* 2014).

2.4. Process of case retain phase

This is the final or last stage of CBR cycle and it is responsible for integrating new cases into the case repository or knowledge base for future utilization (Pal and Shiu 2004), (Zia, *et al.* 2014).

3. EXPERIMENTAL ANALYSIS AND RESULTS

The proposed CBR tool is deployed as an effective prototype application in the medical field used to help the medical practitioners for their decision making process during patient evaluation phase. In order to accomplish an effective functioning of this system the data analysis performed on Indian Liver Patient Dataset (ILPD) that is taken from the UCI machine learning repository (Ramana, *et al.* 2012). **(Table 1)**

Table 1: Indian Liver Patient Dataset (ILPD)

Attribute Name	Min. Value	Max. Value	Mean	SD
Age	4	90		
Gender	Male = 441, Female = 142			
Total Bilirubin (TB)	0.4	42.8	3.3	6.21
Direct Bilirubin (DB)	0.1	19.7	1.49	2.81
Alkphos Alkaline Phosphotase (ALKP)	63	850	290.58	242.94
Sgpt Alamine Aminotransferase (SGPT)	10	779	80.71	182.62
Sgot Aspartate Aminotransferase (SGOT)	10	844	109.91	288.92
Total Protiens (TP)	2.7	9.2	6.48	1.09
Albumin (ALB)	0.9	4.9	6.48	1.09
A/G Ratio Albumin and Globulin Ratio (A/G)	0.3	2.8	0.95	0.31
Class (Decision)	1 = Liver Patient; 2 = Non Liver Patient			

The number of instances inside the liver dataset is 583. Each record contains ten attributes plus the class attribute. **Table 1** shows the attributes and their possible values. 70% of the elements belong to the liver patients denoted by '1' and remaining 30% as non-liver patients. The medical practitioners used the proposed system for classification and diagnosis the disease of the patient.

Fig. 4: Diagnosis Information Screen

(Fig. 4) shows the diagnosis process of the liver patient. In this screen the medical practitioners assigned the weights to the attributes of disease that starts from [1 to 10] i.e. 1 is minimum and 10 is the maximum weight of the feature value. The medical experts input the pathological report value of the patient to the proposed system and that will be treated as a new case. After that, the medical expert will assign the weights of each attribute. These weights determine the importance of each attributes and will be different from the opinion of the different medical experts. The percentage of the threshold value is basically help the medical expert for retrieving the similar record that will not be below the identified threshold value. After assigning the weights of each attributes along with the input of the new case value, the proposed system will retrieve the most similar cases from the knowledge base or case repository.

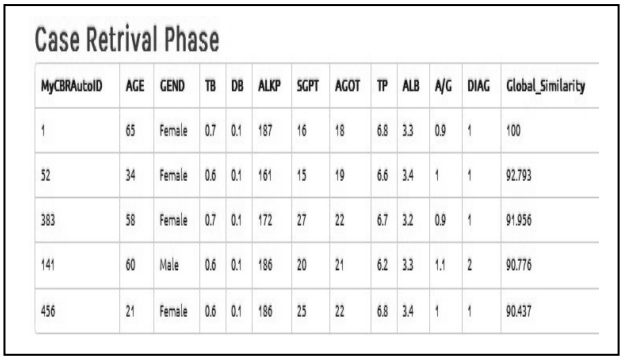

Fig. 5: Retrieve the most similar cases from the case repository

(Fig. 5) shows the most similar cases that are retrieved from the case repository on the basis of the new input case values.

After the retrieval of similar records the medical experts adopt the solution of the most similar retrieved case that is the solution of the new input case. When the similar cases are retrieved from the case repository, the medical practitioners then go to the case reuse phase of the CBR technique where they can adopt the solution of the most similar case with respect to the new input case values. **(Fig. 6)**

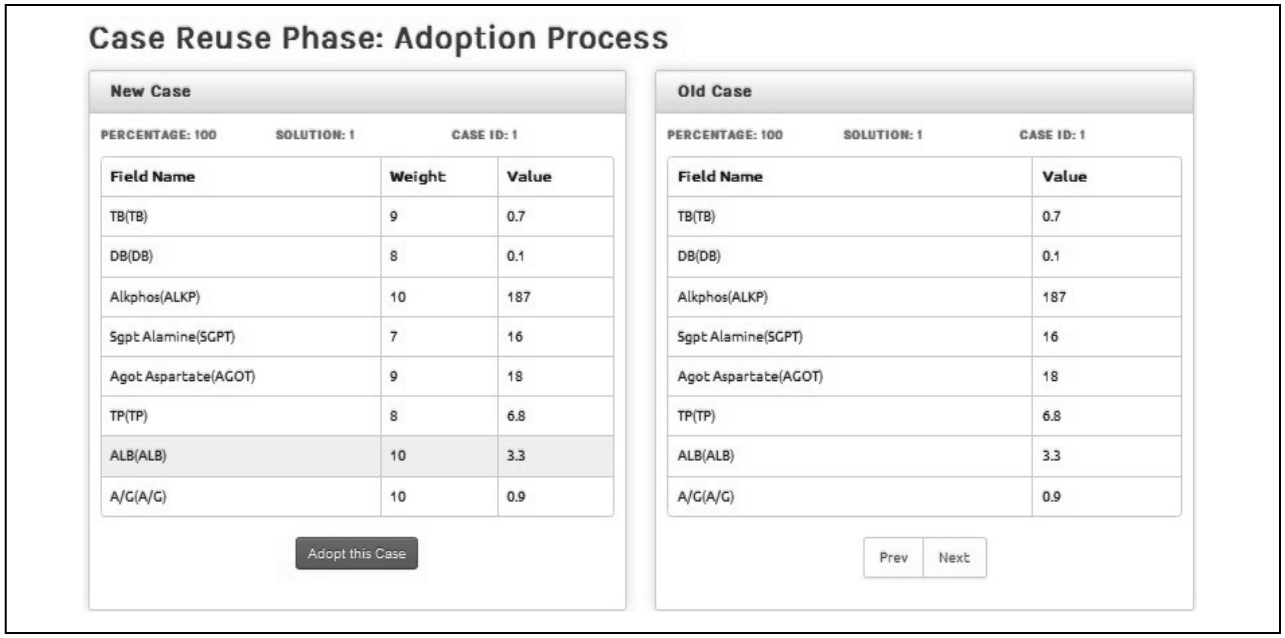

Fig. 6: Adopt the solution of the most similar case that retrieved from the case repository

(Fig. 6) shows the case adoption process in the proposed KBCDSS tool. The process of the case adoption is to show the new input case along with their weights of each attribute and the most similar retrieved case along with the solution and their similarity percentage. The medical practitioners check the similarity percentage of the retrieved cases and then adopt the solution of the most similar retrieved case as the solution of the new input case. After completion of the adoption process, the medical practitioners again check the diagnosis solution of the new input case before storing the new case in a case repository.

(Fig. 7) shows the case revise phase of the CBR technique where the medical practitioners ensured that the diagnosis solution of the new input case is accurate. **Fig.7** shows the solution of the new case is 1, i.e. the patient have a liver disease. When the medical practitioner satisfies the solution of the new case then stores this new case into the case repository or knowledge base for future utilization.

(Fig. 8) shows the confirmation message to the medical practitioners that the new case is stored in the case repository and that will be available for future use.

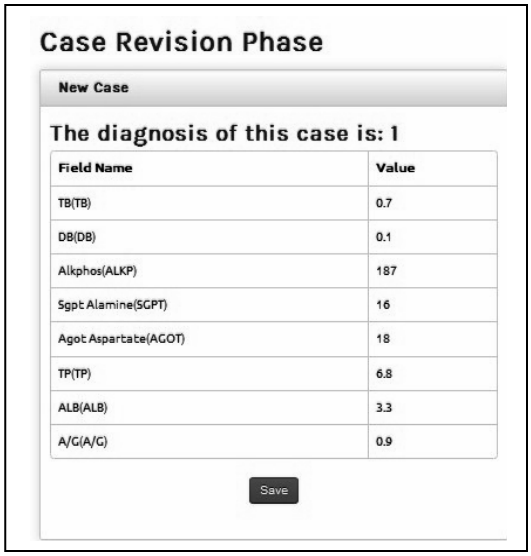

Fig. 7: Verified and validate the accuracy of the new case solution

Fig. 8: New case is stored in the case repository

4. CONCLUSION

This study analyzes the significance of domain knowledge for clinicians during the patient evaluation phase. In this article, we try to implement the schematic cycle of case based reasoning technique as an inference mechanism of clinical decision support system that is used for medical diagnosis & classification. The following features of CBR technique which we covered in this research article are given below:

i. Similarity algorithm is proposed for the accomplishment of an efficient retrieval of cases from the case repository.
ii. Reinstantiation strategy of case adaption is being used for adapting the suggesting way out of the current problem.
iii. Proposed solution of the current problem is revised by a medical expert.
iv. Once the solution is revised, it is then stored into the case repository as a new case.
v. System provides a user friendly GUI environment that solves the human computer interaction problem.

In case revision phase, we employed the reinstantiation strategy for case adaptation of CBR technique which simply copied the solution of the most similar extracted case from the case repository and used this as a solution for the current input case without any modification. Therefore as a future direction if there is an absence of the similar stored cases with respect to a new input case in a case repository, we need an automated rules based reasoning system that will generates the solution for the new input case without the supervision of expert. For this purpose hybrid reasoning approach will be used to solve this type of issues.

REFERENCES:

Aamodt, A., and E. Plaza (1994) "Case-based reasoning: Foundational issues, methodological variations, and system approaches." *AI communications* 7, no. 1, 39-59.

Ahmed, M. U., S. Begum, E. Olsson, N. Xiong, and P. Funk. (2010) "Case-Based Reasoning for Medical and Industrial Decision Support Systems." *In Successful Case-based Reasoning Applications-I.* Springer Berlin Heidelberg,.7-52.

Begum, S., M. U. Ahmed, P. Funk, N. Xiong, and B. Scheele. (2009) "A Case-Based Decision Support System For Individual Stress Diagnosis Using Fuzzy Similarity Matching." *Computational Intelligence* 25, no. 3 180-195.

Holzinger, A., and I. Jurisica. (2014) "Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions." *In Interactive Knowledge Discovery and Data Mining in Biomedical Informatics* (Springer Berlin Heidelberg), 1-18.

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms.* John Wiley and Sons,

Marling, C., S. Montani, I. Bichindaritz, and P. Funk. (2014) "Synergistic case-based reasoning in medical domains." *Expert systems with applications* 41, no. 2 249-259.

Pal, Sankar K., and S. C.K Shiu. (2012) *Foundations of soft case-based reasoning.* Vol. 8. John Wiley & Sons, 2004.

Ramana, B. V., M. Surendra, P. Babu, and N. B. Venkateswarlu. (2012) "A critical comparative study of liver patients from USA and INDIA: an exploratory analysis." *International Journal of Computer Science* 9, no. 3: 506-516.

Rodríguez-González, A., J. Torres-Niño, R. Valencia-Garcia, M. A. Mayer, and G. Alor-Hernandez. (2013): "Using experts feedback in clinical case resolution and arbitration as accuracy diagnosis methodology." *Computers in biology and medicine* 43, no. 8 975-986.

Saleem, A. M. (2007)"Case based reasoning technology for medical diagnosis." *World academy of science, engineering and technology* 31: 9-13.

Salem, A. M. (2011) "Intelligent Technologies and Methodologies for Medical Knowledge Engineering." *ICT in Education, Research and Industrial Applications:Integration, Harmonization and Knowledge Transfer*,: 25.

Sox, H. C., M. C. Higgins, and K..Douglas (2013) "Medical Decision Analysis in Practice: Advanced Methods." *Medical Decision Making, Second Edition:* (Wiley Online Library), 307-335.

Zia, S. S., P. Akhtar, R. Hussain, and I. Mala. (2014): "CBR: Cycle, Framework and Applications." *World Applied Sciences Journal* 32, no. 7 1349-1355.

Zia, S.S., P. Akhtar, T. J. A. Mughal, and I. Mala. (2014): "Case Retrieval Phase of Case-Based Reasoning Technique for Medical Diagnosis." *World Applied Sciences Journal* 32, no. 3 451-458.

.