Improving Weighted *k* Nearest Neighbor Classification Through The Analytic Hierarchy Process Aiding

Cheol-Soo Park and Ingoo Han
Graduate School of Management
Korea Advanced Institute of Science & Technology

1. Introduction

Case-Based Reasoning(CBR) systems support ill structured decision-making. The measure of the success of a CBR system depends on its ability to retrieve the most relevant previous cases in support of the solution of a new case . One of the methodologies widely used in existing CBR systems to retrieve previous cases is that of the Nearest Neighbor(NN) matching function. The NN matching function is based on assumptions of the independence of attributes in previous case and the availability of rules and procedures for matching. The NN algorithm has been extensively analyzed in the literature on pattern recognition and machine learning, where it is viewed as an instance-based learning algorithm. The NN prediction function simply predicts that the given case's class is the same as that of its most similar case. Several studies on machine learning, CBR, statistics, pattern recognition, and other topics have used this algorithm in empirical comparison studies as a straw-man due to its simplicity and popularity. It is well known that while the NN algorithm is relatively robust classifier its primary drawbacks include an inability to tolerate irrelevant attributes, large storage requirements, and relatively high computational complexities for classifying new cases. This is because its similarity function, the Euclidean distance metric, assumes that all features are equally relevant. That is, each feature has equal impact on similarity computations. Several weight learning methods have been proposed, including algorithms based on incremental training, genetic algorithms, decision trees, information theory, ones for symbolic-valued attributes, and several others.

Most CBR systems make use of general domain knowledge in addition to knowledge represented by cases. Representation and use of that domain knowledge involves integration of the case-based method with other methods and representations of problem solving, for instance rule-based systems or deep models.

In this study, We proposed the AHP weighted *k*-NN classification. The AHP is a general theory of measurement in expert judgment methods. It is used to derive ratio scales from both discrete and continuous paired comparisons. These comparisons may be taken from actual measurements or from a fundamental scale which reflects the relative strength of preferences and feelings. The AHP has a special concern with departure from consistency, its measurement and on dependence within and between the groups of elements of its structure. Its has found its widest applications in multicriteria decision making, planning and resource allocation and in conflict resolution. In its general form the AHP is a nonlinear framework for carrying out both deductive and inductive thinking without use of the syllogism by taking several factors into consideration simultaneously and allowing for dependence and for feedback, and making numerical tradeoffs to arrive at a synthesis or conclusion.

Consequently, this paper proposes an extension to the *k*-NN algorithm. The *k*-NN algorithm has been shown to achieve comparable accuracy with the *k*-NN algorithm. However, *k*-NN algorithm has a very low time complexity compared to *k*-NN. The extension to *k*-NN algorithm introduced here assign weight to features using AHP, therefore it is called AHP W*k*-NN, for AHP Weighted *k* Nearest Neighbor. The paper also introduces a weight algorithm, called AHP, for Analytic Hierarchy Process. It is based on the concept of 'trade-off' and enables the decision-maker to develop the trade-off implicitly in the course of structuring and analyzing a series of reciprocal pairwise comparison matrices. An empirical evaluation of the AHP method on credit evaluation shows that it achieves an important improvement in the classification accuracy of the AHP W*k*-NN algorithm.

2. Overview of Case-based Reasoning

Case-based reasoning(CBR) makes data more accessible by organizing it as a set of examples from past experience that can be generalized and applied to current problems. It employs a combination of artificial intelligence technologies such as advanced knowledge representation and inductive machine learning to structure the information and discover useful knowledge. CBR also integrates traditional approaches to data analysis into its methodology. The many CBR applications illustrate an emerging pattern of using CBR to expand and apply corporate experience.

CBR provides a technique for capturing the experience in a large set of historical cases. It is a paradigm for knowledge-based systems development that solves new problems by retrieving and adapting old solutions. In addition to problem solving, its case library can be used for training and additional research into the trends and anomalies in the data. Case-based systems derive their power from their ability to retrieve relevant cases from a case library efficiently. However, for most applications, more sophisticated approaches are required. Three major types of indexing have been applied. They include:

1. Nearest-neighbor matching retrieves cases based on a weighted sum of features in the input case. The cases with the "closest" overall match according to some similarity metric are returned from the match process. This approach is best if the retrieval needs are not focused tightly on solving a specific problem.
2. Inductive retrieval methods are best when the retrieval goal is well-defined, such as the bond rating problem. Case is indexed based on the most important features affecting the outcome as induced from the data itself. The resulting decision tree provided for considerably faster

retrieval times than nearest-neighbor retrieval.

3. Knowledge-based retrival applies existing domain knowledge to locate relevant case. This approach is similar to rule-based expert systems, in which an expert determines the features used to classify cases. The knowledge need not be complete, and frequently systems combine a partial model of the domain with other indexing methods to retrieve accurate solutions.

2.1. The Nearest Neighbor(NN) Algorithms and The k-Nearest Neighbor(k-NN) Algorithms

The NN classification is based on the assumption that examples which are closer in the instance space are of the same class. Namely, unclassified ones should belong to the same class as their NN in the training dataset. After all the training set is stored in memory, a new example is classified with the class of the NN among all stored training instances. Although several distance metrics have been proposed for NN algorithms, the most common metric is the Euclidean distance metric. A primary weakness of the NN function is that it is sensitive to the presence of irrelevant features in the case representation. This is because its similarity function, the Euclidean distance metric, assumes that all features are equally relevant. That is, each feature has equal impact on similarity computations. Dynamic feature selection algorithms alleviate this problem. Most of them assign weights to each feature. The most relevant features are assigned the highest weights. For example, a typical weighed-Euclidean similarity function is

$$Similarity(x,y) = \sqrt{\sum_{iu=1}^{F} w_i \times (x_i - y_i)^2}$$

where w_i is the weight of feature i. Using this function, features with weights of zero are effectively ignored during similarity computations, whereas features whose weights are high have the most impact on determining similarity.

More generally, the k nearest neighbors are computed, and the new example is assigned the class that is most frequent among these k neighbors. Ties are broken arbitrarily in favor of the class with the smallest index among the ties. The optimal value of k can be estimated via leav-one-out cross-validation. Ties during cross-validation are broken in favor of smaller *k*s.

2.2 Estimating the Value of k

The previous study was concerned with whether any value of $k \neq 1$ would lead to a performance superior to that of the first nearest neighbor algorithm. In this section, we will investigate the issue of how the value of k that would lead to the best performance can be reliably estimated. One assumption that has been made throughout all research reported in the literature to this date is that a single value of k suffices to classify all queries.

3. Overview of the Analytic Hierarchy Process

The Analytic Hierarchy Process(AHP) has been applied to many varied and complex situations ranging from predicting oil price and planning for a national waterway, to a variety of financial decisions. Srinivasan and Kim illustrate the applicability of the AHP to many financial decision. For a comprehensive review of nonfinancial application of the AHP, see Zahedi. The methodology is based on the concept of 'trade-off' and enables the decision-maker to develop the trade-off implicitly in the course of structuring and analyzing a series of reciprocal pairwise comparison matrices.

The AHP was developed by Thomas L. Saaty[115] as a scaling procedure for priorities in hierarchical goal structure. The method's comparative advantage lies in areas too fuzzy, too unstructured, or too political for traditional techniques which require that measurement scales be explicit. The objective is to use the resulting priorities to allocate resources or important the most important project. The problem is to find the relative priorities of the projects with respect to each other, with respect to decision criteria, and to combine these priorities to a single overall ranking. While some analysis involve only the projects under consideration, more frequently analysis require the synthesis of the multiple decision makers assessments on multiple projects for multiple decision criteria.

<Table 1> AHP Four Step Implementation Approach.

Step 1 – Organize the analysis by breaking the problem into a hierarchy of interrelated decision elements.

Step 2 – Collect input data by pairwise comparison of decision elements.

Step 3 – Estimate the relative weights of the decision elements.

Step 4 – Consolidate the relative weights of decision elements to arrive at a set of ratings for the decision alternatives.

<Table 2> 9 Point Scale of Relative Importance used in the AHP

Intensity of Relative Importance	Definition	Explanation
1	Equal Importance	Two activities contribution equally to the objectives.
3	Moderate Importance of one over another	Experience and judgement slightly favor one activity over another.
5	Essential or Strong Importance	Experience and judgement strongly favor one activity over another.
7	Very strong importance	An activity is strongly favored and its dominance is demonstrated in practice.
8	Absolute importance	The evidence favoring one activity over another is of the highest possible order of affirmation.
2, 4, 6, 8	Intermediate values between two adjacent judgements	When compromise is needed.
Reciprocals of above Non-zero numbers	If activity i has one of the above non-zero numbers assigned to it when compared with activity j, the j has the reciprocal value when compared to i.	

<Table3> AHP Random Index

n	1	2	3	4	5	6	7	8	9	10
R.I.	0.00	0.00	0.058	0.90	1.12	1.24	1.32	1.41	1.45	1.49

If an estimate of $w_i/w_j = a_{ij}$ for each matrix entry

$$A = \begin{bmatrix} a_{11}\, a_{12} \cdots a_{1n} \\ a_{21}\, a_{22} \cdots a_{2n} \\ \vdots \quad \vdots \quad \vdots \\ a_{n1}\, a_{n2} \cdots a_{nn} \end{bmatrix}$$

and $a_{ij} = 1/a_{ij}$, the matrix A takes the form

This reciprocal ratio matrix can be translated to a largest eigenvalue problem since each A is a square matrix whose values are all real and positive. According to the Perron-Frobenius theorem[49,79], at least one real positive eigenvalue for any matrix with real positive entries exists and the eigenvalue is associated with a unique eigenvector of weighed. This eigenvector may be normalized such that the sum of its entries is one. If the pairwise comparison are perfectly consistent, that is if:

$$a_{ik}\, a_{kj} = a_{ij} \quad for\ all \quad i, j,\ k = 1,\ 2,\ \ldots,\ n$$

then any column may be normalized to yield the final matrix weight vector W. However, errors in judgement are typically made and the final result using column normalization approach would depend on which column was selected. Several authors discuss four competing methods for estimating W when errors in judgement exist. Those methods are the Least Square Methods(LSM), Geometric Means Methods(GMM), Logarithmic Least Squares(LLS), and the Eigenvector Method(EM). The LSM minimizes the objective:

Note, that, while there is no closed form solution to this problem, it is possible to solve it by Marquardt's iterative method. However, it is not a widely used approach. The second method GMM assumes the form:

$$w_i = \left[\prod_{j=1}^{n} (a_{ij})\right]^{1/n}$$

to generate the weight vector W where the vector W must be additionally normalized so the individual w_i sum to unity. It can be proved that the vector also provides a solution to the LLS, indeed, the GMM is sometimes used as an approximation method for the LLS approach. The third method, the LLS, estimates the weight vector by minimizing:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\ln a_{ij} - \ln w_i + w_j\right)^2$$

Finally, the eigenvector method computer W as the principal right eigenvector (or Perron right vector) of the matrix A by:

$$A_w = \lambda_{max} w$$

Where λ_{max} is the maximum eigenvalue of the matrix and decision alternative weights are computed by

$$w_i = \frac{\sum_{j=1}^{n} a_{ij} w_j}{\lambda_{max}} \quad for\ all\ i = 1, 2, \ldots, n.$$

Harker and Vargas and Saaty and Vargas note that the LLS and the EM have their advantage, however, Vargas assert the EM is a simple averaging process and thus is "a natural method for computing the weights."

Finally, The eigenvector method has the

developed capability to measure inconsistency in the pairwise comparison. As shown by Saaty, λ_{max} is always greater than or equal to n for positive, reciprocal matrices and is equal to n if and only if A is a consistent matrix. Following normalization for the size of the matrix, this value is called the consistency index(C.I.) where C.I. is computed by

Matrices of size n, for i = 1, ... 15, were populated with random values and their mean C.I. shown in Table 4 was called a random index(R.I.) by Saaty. Using the ratio of C.I. to R.I., Saaty defined the consistency ratio(C.R.) as a measure od how a give matrix compare to a purely random matrix. A value of C.R. ≤ 0.1 is typically considered acceptable.

Finally, Step four above consolidates the weight vectors from the various hierarchy levels in order to obtain a single vector of weights which serves as a ranking of the decision alternatives. Saaty computes this composite relative weight vector at the *k* th hierarchy level with respect to

$$C[1,k]\prod_{i=2}^{k} B_i$$

the first level by:

where C[1, k] = the vector of composite weights of elements hierarchy level k with respect to the element on hierarchy level 1

B_i = the n_{i-1} by n_i matrix rows of the estimated vectors of actual relative weights

n_i = the number of elements at level I

i = the hierarchy level

Following a basic review of the methodology, it seems appropriate to discuss some of the surrounding issues. Harker and Vargas assert there are four criticisms of the AHP theory.

Those are:

1) The axiomatic foundation of the AHP is not clearly defined
2) The ambiguity of the questions which the decision maker must answer
3) The use of a ratio scale to measure the decision maker's preference intensity
4) The validity of the Principle of Hierarchy Composition and resulting rank reversals in decision matrices

As observed by those following the literature, there is an on-going exchange of article regarding these four issues. The final resolution is yet to be determined.

A review of the recent articles expresses a polarity in basic philosophy. On one extreme are the traditionalist contending the AHP methodology fails to fully address the issues surrounding the decision making process. The focus of arguments appears to be on the need for and validity of new theory in light of the existing utility theory. On the opposite pole, the AHP developers and supporters contend the theory is new and not completely understood, but it fully addresses all the relevant issues. Saaty goes so far as to equate the refusal of utility theory supporters to accept the AHP to those of the pre-Einstein physicists and their refusal to accept relativity and the Einsteinian concepts of absolute and relative space-time physics.

Other than to examine the issues, acknowledge their existence, and insure the issues will not compromise the integrity of this research, the above issues will not bear further review. Whether the AHP paradigm is totally new or an extrapolation of existing theory is not a consideration of this research. It is presupposed that, while some methodological issues may remain, the basic paradigm is appropriate as a proven baseline.

4. AHP weighted k-Nearest Neighbor

Integration of domain knowledge into case indexing and retrieving process is important in building a useful case-based reasoning. In this study, we employ the weighted k-nearest neighbor approach. The weighted k-nn has difficulty in deciding a set of feature weights that could accurately retrieve cases in a given domain.. Since the feature weights for most problem domains are context dependent, each case should have its own set of feature weights for determining the relevance of that case to a new problem.

The importance associated with each field tells us how much attention to pay to the match. Although Kolodner(1993) suggests several ways of assigning the importance values such as knowledge of human experts and statistical evaluation, it is difficult to tell a priori which set of weigh would be the most effective to solve a specific problem. Considering that a function computing the degree of match can only be as good as the knowledge it has of the importance of dimensions, it is an importance of dimensions, it is an important task to find an optimal set of weights.

As Kolodner's suggests, one way to assign importance values is to have a human expert assign them as the case library is being guilt. The expert is expected to have the knowledge and experience required to decide which dimension make good predictors. So, Our central idea is hybrid of AHP and case-based reasoning that CBR transfer the burden of knowledge assignment of the indexing and retrieving process.

AHP model is effective method to obtain expert knowledge. Our study is an integrated approach using AHP and CBR to retrieve more relevant cases. This approach is aimed at unifying case specific and general domain knowledge within the system. We integrate the AHP and CBR. As the first step, we build a hierarchy structure for AHP weight. The second step is to apply the derived weight vector in step 1 to case indexing scheme for the case-based retrieval process and evaluated the resulting model with the additional validation cases for which the outcome is also known. A weight vector is used in the nearest neighbor matching function to rank and retrieve useful cases.

5. Application

Our study performs the assignment of AHP importance values to each dimension of case features in the problem of credit evaluation. We have seen that the AHP weighted k-nearest neighbor algorithm can outperform the first-nearest neighbor algorithm on specific data set. This study, therefore, devoted to studying methods and modifications of kNN that can improve kNN's performance. We will investigate the issues of how to assign the value of knowledge importance that will maximize the predictive accuracy of kNN on the test set. This is followed by a comparison of simple kNN versus kNN with AHP weight.

<Table 4> Classification Accuracy for Credit Evaluation

Methods	Performance
Pure CBR	69%
Weighted NN	71%
AHP Weighted k-NN	
1, 5-NN	75.5%
2, 10-NN	73%
3, 12-NN	72.8%
4. 15-NN	73.2%

6. Conclusions

In this study, we have proposed by hybrid methodology using AHP and CBR to the problem of credit evaluation. The AHP is used to assign relative importance of feature weights for case indexing and retrieving. We have shown that these hybrid approach support an effective retrieval of case and increases overall classification accuracy significantly.

References

[1] Panel of CBR Workshop, Case-based reasoning from DARPA machine learning program, in: *Proc. Of DARPA Workshop on Case-Based Reasoning* (Morgan Kaufman Publishers, San Mateo, CA, 1989) 1-14.

[2] J.L. Kolodner, *Case-Based Reasoning* (Morgan Kaufman Publishers, San Mateo, CA, 1993).

[3] B.V. Dasarathy, *Nearest Neighbor(NN) Norms: NN Pattern Classification Techniques* (IEEE Computer Society Press, Los Alamitos, CA, 1991).

[4] D.W Aha, D. Kibler and M.K Albert, Instance-Based Learning Algorithms, *Machine Learning* 6, 37-66.

[5] P.T. Hacker, and Vargas, L.G. (1987), "The theory of ratio scale estimation: Saaty's analytic hierarchy process," *Management Science*, 33(11), pp.1383-1403.

[6] T.L. Saaty, *The Analytic Hierarchy Process* (McGraw-Hill, New York, 1980)

[7] T.L. Saaty, The Analytic Hierarchy Process-What It is and How It is Used, *Mathematical Modelling* 9, 161-176.

[8] V. Srinivisan, and Ruparel, B. (1990), "CGX: An expert support system for credit granting," European Journal of Operational Research, 45, pp.293-308.

[9] F. Zahedi, (1986), " The analytic hierarchy process-A survey of the method and its applications," *Interface*, 16(4), pp.96-116.