CONSENSUS CLUSTERING AND FUZZY CLASSIFICATION FOR BREAST CANCER PROGNOSIS

Jonathan M. Garibaldi and Daniele Soria
School of Computer Science
University of Nottingham, Jubilee Campus
Wollaton Road, Nottingham, NG14 5AR, UK
Email: jmg@cs.nott.ac.uk

Khairul A. Rasmani
Faculty of Information Technology
Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia
Email: khairulanwar@tmsk.uitm.edu.my

KEYWORDS

Clustering, Validity Indices, Consensus Clustering, Fuzzy Classification, Breast Cancer, Prognosis

ABSTRACT

Extracting usable and useful knowledge from large and complex data sets is a difficult and challenging problem. In this paper, we show how two complementary techniques have been used to tackle this problem in the context of breast cancer. Diagnosis concerns the identification of cancer within a patient; in contrast, *prognosis* concerns the prediction of the ongoing course of the disease, including issues such as the choice of potential treatments such as chemotherapy or drug therapy, in combination with estimation of chances (or length) of survival. Reliable prognosis depends on many factors, including the identification of the *type* of this heterogeneous disease. We first use a consensus clustering methodology to identify core, well-characterised sub-groups (or *classes*) of the disease based on a large database of protein biomarkers from over a thousand patients. We then use fuzzy rule induction and simplification algorithms to generate a simple, comprehensible set of rules for use in future model-based classification. The methods are described and their use is illustrated on real-world data.

INTRODUCTION

Breast cancer, the most common cancer in women (Parkin et al., 2001; Kamangar et al., 2006), is a complex disease characterized by multiple molecular alterations. Current routine clinical management relies on availability of robust clinical and pathologic prognostic and predictive factors to support decision making. Recent advances in high-throughput molecular technologies have supported the evidence of a biologic heterogeneity of breast cancer. We and others have applied protein biomarker panels with known relevance to breast cancer, to large numbers of cases using tissue microarrays, exploring the existence and clinical significance of distinct breast cancer classes (Abd El-Rehim et al., 2005; Ambrogi et al., 2006; Callagy et al., 2003; Jacquemier et al., 2005; Diallo-Danebrock et al., 2007).

Clustering has become a widely used approach to extrapolate important information from data and to separate different groups that share similar characteristics within them. Cluster analysis may be thought of as the discovery of distinct and non-overlapping sub-partitions within a larger population (Monti et al., 2003). Many different clustering techniques are known today, but often only a few selected methods are used in any given domain. Choosing which method to use is not an easy task, as different clustering techniques return different groupings. Consequently, it has been demonstrated (Ambrogi et al., 2006; Soria et al., 2010) that the use of several methods is preferable in order to extract as much information as possible from the data.

When using more than one algorithm, it is then common to define a consensus across the results (Kellam et al., 2001) in order to integrate diverse sources of similarly clustered data (Filkov and Skiena, 2003) and to deal with the stability of the results obtained from different techniques. Several approaches have been proposed for this task. Kellam and colleagues (Kellam et al., 2001) identified robust clusters by the implementation of a new algorithm called 'Clusterfusion'. It takes the results of different clustering algorithms and generates a set of robust clusters based upon the consensus of the different results of each algorithm. In essence, a clustering technique is applied to the clustering results. Another approach, suggested by Monti and colleagues (Monti et al., 2003), deals with class discovery and clustering validation tailored to the task of analysing gene expression data. The new methodology, termed 'consensus clustering', provides a method, in conjunction with resampling techniques, to represent the consensus across multiple runs of a clustering algorithm and to assess the stability of the discovered clusters. Filkov and Skiena suggested to exploit the popularity of cluster analysis of biological data by integrating clusterings from existing data sets into a single representative clustering based on pairwise similarities of the clusterings. Their proposed representative clustering was the one that minimised the distance to all the other partitions (Filkov and Skiena, 2003). In another approach, Swift and colleagues used consensus clustering to improve confidence in gene-expression analysis, on the assumption that microarray analysis using clustering algorithms can suffer from lack of inter-method consistency in assigning related gene-expression profiles to clusters (Swift et al., 2004).

We adopted an alternative approach, based on calculating a number of external cluster validity indices across a range of cluster solutions produced by alternative clustering algorithms, and using consensus across the cluster validity indices and across methods to reach the overall 'best' number of clusters (Soria and Garibaldi, 2010). This methodology results in a number of well characterised (separate and distinct) groups of breast cancer cases, which may be interpreted as different classes (or types) of breast cancer, with corresponding alternative treatment regimes.

There are many non-fuzzy classification algorithms currently available, for example (Witten and Frank, 2000). However, many of these classification algorithms may be very good in generalisation ability and so be very useful for classifying new instances, but lack of comprehensibility of the generated models. In fact, most of the models generated by non-fuzzy classification algorithms contain numerical values and may not be linguistically interpretable. This makes it harder for the user to utilise the models for decision making purposes. Note that an automated-system, or decision support system, is normally considered as a tool to assist experts or non-experts in decision making. Hence, interpretability of such a system is normally regarded as highly important (Castellano et al., 2006). With interpretability in mind, we recently proposed a novel algorithm to induce a simplified set of linguistic rules (Rasmani et al., 2009) suitable for use in a quantifier-based fuzzy classification system (Rasmani and Shen, 2004). This methodology was applied to the breast cancer classes obtained by our consensus clustering in order to obtain a model-based fuzzy classification system suitable for new cases.

CONSENSUS CLUSTERING

The three-step methodology for elucidating core, stable classes (groups) of data from a complex, multidimensional dataset was as follows:

1. A variety of clustering algorithms were run.

2. The most appropriate number of clusters was investigated by means of cluster validity indices.

3. Concordance between clusters, assessed both visually and statistically, was used to guide the formation of stable 'core' classes of data.

The methodology was applied to a well-known set of data concerning breast cancer patients (Abd El-Rehim et al., 2005) in order to obtain core classes. Once these core classes were obtained, the clinical relevance of the corresponding patient groups were investigated by means of associations with related patient data. All statistical analysis was done using *R*, a free software environment for statistical computing and graphics (Maindonald and Braun, 2003).

Clustering Algorithms
Five different algorithms were used for cluster analysis:

1. Hierarchical (HCA)
2. K-means (KM)
3. Partitioning around medoids (PAM)
4. Adaptive resonance theory (ART)
5. Fuzzy c-means (FCM)

Hierarchical clustering: The hierarchical clustering algorithm (HCA) begins with all data considered to be in a separate cluster. It then finds the pair of data with the minimum value of some specified distance metric; this pair is then assigned to one cluster. The process continues iteratively until all data are in the same (one) cluster. A conventional hierarchical clustering algorithm (HCA) was utilised, utilising Euclidean distance on the raw (unnormalised) data with all attributes equally weighted.

K-means clustering: The K-means (KM) technique aims to partition the data into K clusters such that the sum of squares from points to the assigned cluster centres is minimised. The algorithm repeatedly moves all cluster centers to the mean of their Voronoi sets (the set of data points which are nearest to the cluster centre). The objective function minimised is:

$$J(V) = \sum_{j=1}^{k} \sum_{i=1}^{c_j} ||x_i - v_j||^2$$

where x_i is the i-th datum, v_j is the j-th cluster centre, k is the number of clusters, c_j is the number of data points in the cluster j and $||x_i - v_j||$ is the Euclidean distance between x_i and v_j.

The j-th centre v_j can be calculated as:

$$v_j = \frac{1}{c_j} \sum_{i=1}^{c_j} x_i, \quad j = 1, ..., k.$$

K-means clustering is dependent on the initial cluster centres setting (which, in turn, determines the initial cluster assignment). Various techniques have been proposed for the initialisation of clusters (Al-Daoud and Roberts, 1996), but for this study we used a fixed initialisation of the cluster centres obtained with hierarchical clustering. The number of clusters is an explicit input parameter to the K-means algorithm.

Partitioning around medoids: The partitioning around medoids (PAM) algorithm (also known as the k-medoids algorithm) is a technique which attempts to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the K-means algorithm, PAM chooses data points as centers (the so-called medoids) and then assigns each point to its nearest medoid. A medoid is defined as the

object within a cluster for which the average dissimilarity to all other objects in the cluster is minimal, i.e. it is the most centrally located datum in the given cluster. Dissimilarities are nonnegative numbers that are close to zero when two data points are 'near' to each other and large when the points are very different (Kaufman and Rousseeuw, 1990). Usually, a Euclidean metric is used for calculating dissimilarities between observations.

The algorithm consists of two phases: the *build* phase in which an initial set of k representative medoids is selected and the *swap* phase in which a search is carried out to improve the choice of medoids (and hence the cluster allocations). The *build* phase begins by identifying the first medoid, the point for which the sum of dissimilarities to all other points is as small as possible. Further medoids are selected iteratively through a process in which the remaining points are searched to find that which decreases the objective function as much as possible. Once k medoids have been selected, the *swap* phase commences in which the medoids are considered iteratively. Possible swaps between each medoid and other (non-medoid) points are considered one by one, searching for the largest possible improvement in the objective function. Thiscontinues until no further improvement in the objective function can be found. The algorithm is described in detail in (Kaufman and Rousseeuw, 1990), pp.102–104. The number of clusters is an explicit input parameter to the PAM algorithm.

Adaptive resonance theory: The adaptive resonance theory (ART) algorithm has three main steps (Carpenter and Grossberg, 1987). First, the data are normalised to a unit hypersphere, thus representing only the ratios between the various dimensions of the data. Second, data allocated to each cluster are required to be within a fixed maximum solid angle of the group mean, controlled by a so-called 'vigilance parameter' ρ, namely $X_k \cdot P^i \leq \rho$. However, even when the observation profile and a prototype are closer than the maximum aperture for the group, a further test is applied to ensure that the profile and prototype have the same dominant covariates. This is done in a third step by specifying the extent to which the nearest permissible prototype allocation for the given observation must be on the same side of the data space from the diagonal comprising a vector of ones, $\hat{1}$, using a preset parameter, λ:

$$X_k \cdot P^i \leq \lambda X_k \cdot \hat{1}.$$

The ART algorithm is initialised with no prototypes and creates them during each successive pass over the data set. It has some, limited, sensitivity to the order in which the data are presented and converges in a few iterations. In the ART algorithm the clusters are determined automatically: the number of clusters is not an explicit parameter, although there are parameters that can adjust the number obtained.

Fuzzy c-means: The fuzzy c-means (FCM) algorithm is a generalisation of the K-means algorithm which is based on the idea of permitting each object to be a member of *every* cluster to a certain degree, rather than an object having to belong to only one cluster at any one time. It aims to minimise the objective function:

$$J(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{i,j})^m \|x_i - v_j\|^2$$

where n is the number of data points, x_i and v_j are the data points and cluster centres and $\mu_{i,j}$ is the membership degree of data x_i to the cluster centre v_j ($\mu_{i,j} \in [0,1]$). m is called the 'fuzziness index' and the value of $m = 2.0$ is usually chosen. An exhaustive description of this method can be found in (Bezdek, 1974). As for K-means, the number of clusters is an explicit input parameter to FCM.

Cluster Validity

Clustering validity is a concept that is used to evaluate the quality of clustering results. If the number of clusters is not known prior to commencing an algorithm, a cluster validity index may be used to determine the best number of clusters for the given data set. Although there are many variations of validity indices, they are all either based on considering the data dispersion in a cluster and between clusters, or considering the scatter matrix of the data points and the one of the clusters centers. In this study, the following indices were applied to those algorithms for which the number of clusters is an explicit parameter, over a range of number of clusters:

1. Calinski and Harabasz (Maulik and Bandyopadhyay, 2002)
2. Hartigan (Hartigan, 1975)
3. Scott and Symons (Scott and Symons, 1971)
4. Marriot (Marriot, 1971)
5. Trace$W$ (Edwards and Cavalli-Sforza, 1965; Friedman and Rubin, 1967)
6. Trace$W^{-1}B$ (Friedman and Rubin, 1967)

For each index, the number of clusters to be considered was chosen according to the rule reported in Table 1 where i_n is the validity index value obtained for n clusters (Weingessel et al., 1999).

CLUSTERING RESULTS

Patients and Clinical Methods

A series of 1076 patients from the Nottingham Tenovus Primary Breast Carcinoma Series presenting with primary operable (stages I, II and III) invasive breast cancer between 1986-98 was used to evaluate the methodology. Immunohistochemical reactivity for twenty-five proteins, with known relevance in breast cancer including those

Table 1: Different validity indices and their associated decision rules

Index	Decision rule
Calinski and Harabasz	$\min_n((i_{n+1} - i_n) - (i_n - i_{n-1}))$
Hartigan	$\min_n((i_{n+1} - i_n) - (i_n - i_{n-1}))$
Scott and Symons	$\max_n(i_n - i_{n-1})$
Marriot	$\max_n((i_{n+1} - i_n) - (i_n - i_{n-1}))$
Trace$W$	$\max_n((i_{n+1} - i_n) - (i_n - i_{n-1}))$
Trace$W^{-1}B$	$\max_n(i_n - i_{n-1})$

used in routine clinical practice, were previously determined using standard immunocytochemical techniques on tumour samples prepared as tissue microarrays (Abd El-Rehim et al., 2005). Levels of immunohistochemical reactivity were determined by microscopical analysis using the modified H-score (values between 0-300), giving a semiquantitative assessment of both the intensity of staining and the percentage of positive cells.

HCA, K-means, PAM and ART Clustering

The HCA results from our previous study (Abd El-Rehim et al., 2005) were utilised, unaltered. Both the K-means and PAM algorithms were run with the number of clusters varying from two to twenty, as the number of clusters is an explicit input parameter of the algorithms. Given that both algorithms can be sensitive to cluster initialisation and in order to obtain reproducible results, both techniques were initialised with the cluster assignments obtained by hierarchical clustering. For the ART algorithm, the parameters were adjusted to obtain six clusters to match the number of clusters previously obtained by HCA. The best validity index obtained for repeated runs of the algorithm with 20 random initialisations was used to select the final clustering assignment.

Fuzzy C-means Clustering

The fuzzy c-means algorithm did not perform as hoped. When the number of clusters was set as two and three, it appeared that reasonable results were obtained. However, from examination of the membership function of each point assigned to these clusters, it could be seen that it was very close to either $\frac{1}{2}$ or $\frac{1}{3}$, respectively. In other words, every data point was assigned to all the clusters with the same membership. Moreover, when the number of clusters was above three, non-zero memberships were evident for only three clusters and these memberships were similar to the three cluster solution — i.e. for $n > 3$, the $n = 3$ cluster solution was obtained, but with $n - 3$ empty clusters.

The fuzziness index m was altered in an attempt to improve the results obtained, but it was found that little difference in the results was observed until m was close to one. Given that when $m = 1$ fuzzy c-means is equivalent to K-means, this result was not useful. As there are many applications for which the fuzzy c-means technique has been successful (see, for example, (Wang and Garibaldi, 2005)), these results are not easy to explain, but they may have been caused by the fact that our data contains a lot of values close to the extremes of each variable. Although the fuzzy c-means algorithm is widely used in literature, we decided to drop it from further analysis due to its poor performance on our data.

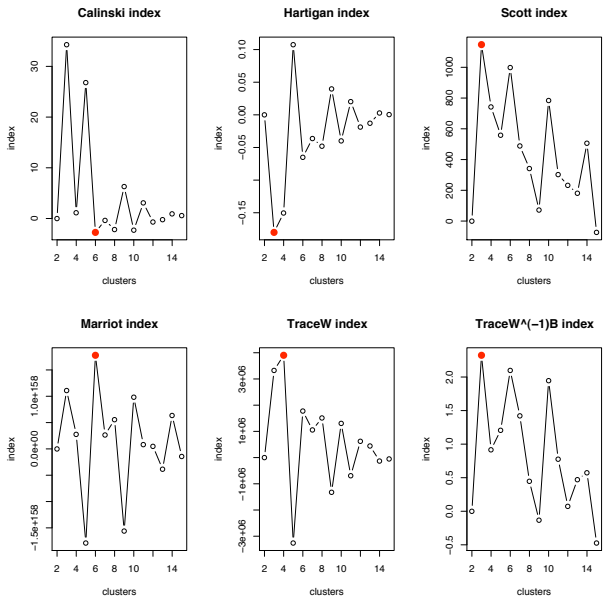

Figure 1: Cluster validity indices obtained for K-means for varying cluster numbers from 2 to 20.

Table 2: Optimum number of clusters estimated by each index for K-means and PAM methods

Index	K-means	PAM
Calinski and Harabasz	6	4
Hartigan	3	4
Scott and Symons	3	4
Marriot	6	4
Trace$W$	4	4
Trace$W^{-1}B$	3	4
Minimum sum of ranks	6	4

Cluster Validity

The values of the decision rule obtained for various values of the validity indices for K-means, for 2 to 20 clusters, are shown in Figure 1. The best number of clusters according to each validity index, for each clustering algorithm, is shown in Table 2, as indicated by the solid circle in Figure 1.

It can be seen that, while there was not absolute agreement among the indices as to which was the best number of clusters for the K-means method, there is good agreement that the best number of clusters for the PAM method is four. Although the best number of clusters varies according to validity index for K-means, on further inspection, it can be seen from Figure 1 that there is more agreement than might be immediately apparent. For example, the Scott and Symons index (which indi-

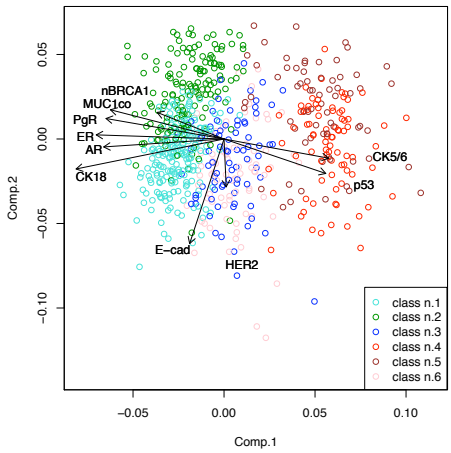

Figure 2: Biplot of classes projected on the first and second principal component axes

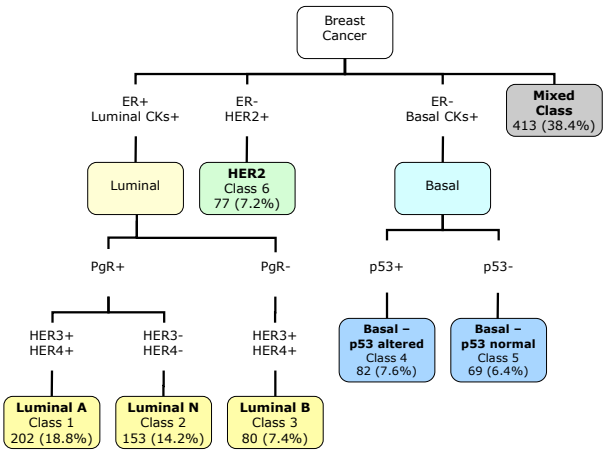

Figure 3: A summary of the classes of breast cancer obtained, with indicative class interpretations.

cated that the best number of clusters was three) indicated that the second best number of clusters was six. Consequently, the indices were used to rank order the number of clusters and the minimum sum of ranks was examined. It was found that the minimum sum of ranks (a form of consensus among the indices) indicated that the overall best number of clusters was six for K-means and four for PAM. However, it was subsequently found that the four cluster solution obtained by PAM was not as clinically interesting as the six cluster solution and it was dropped from further analysis.

Characterisation of Classes

Biplots of the six consensus classes were produced and are shown in Figure 2, in order to provide a visualisation of the separation of the classes. A proposed summary of the essential characterisations of the classes obtained is given in Figure 3, according to the available biopathological knowledge. It is worth noting that class 2, labelled as Luminal-N, and the split of the basal group into two different subgroups depending on p53 levels, appear to be novel findings not previously emphasised in literature.

FUZZY CLASSIFICATION

Fuzzy Subsethood Measures

A fuzzy subsethood measure was originally defined as the degree to which a fuzzy set is a subset of another. However, the definition of fuzzy subsethood value can be extended to calculate the degree of subsethood for linguistic terms in an attribute variable V to a decision class D (Yuan and Shaw, 1995). For linguistic terms $\{A_1, A_2, \ldots, A_n\} \in V$ and $(V, D) \subseteq U$:

$$S(D, A_i) = \frac{\sum_{x \in U} \nabla(\mu_D(x), \mu_{A_i}(x))}{\sum_{x \in U} \mu_D(x)} \qquad (1)$$

where ∇ can be any *t-norm* operator. It should be noted that, to be used for classification problems, both V and D must be defined under the same universe of discourse U (Yuan and Shaw, 1995). Although the decision class is represented by fuzzy sets, this definition allows the decision class with zero fuzziness where the membership value is either one or zero.

Rule Induction

FuzzyQSBA is a rule induction algorithm that was developed by extending the Weighted Subsethood-based Algorithm (WSBA) (Rasmani and Shen, 2006). WSBA has the significant advantage, as compared to previous subsethood-based methods, of not relying on the use of predefined threshold values in generating fuzzy rulesets. The development of WSBA was based on fuzzy subsethood values as defined in Equation (1). Given a training dataset, WSBA induces a fixed number of rules according to the number of possible classification outcomes. To avoid the use of any threshold values in the rule generation process, crisp weights generated using fuzzy subsethood values are created for each of the linguistic terms appearing in the resulting fuzzy rule antecedents. In FuzzyQSBA, fuzzy quantifiers are applied to replace the crisp weights within the rules learned by WSBA. As small changes in the training dataset might cause a change to the entire ruleset, developing a fuzzy model that employs continuous fuzzy quantifiers may be more appropriate compared to two-valued or multi-valued crisp quantifiers (Rasmani and Shen, 2004). (Vila et al., 1997) proposed a continuous fuzzy quantifier which applies linear interpolation between the two classical, extreme cases of the existential quantifier $\exists$ and the universal quantifier $\forall$. In particular, the quantifier was defined such that:

$$Q(A_{ij}, D_k) = (1 - \lambda_Q) T_{\forall, A/D} + \lambda_Q T_{\exists, A/D} \qquad (2)$$

where Q is the quantifier for fuzzy set A relative to fuzzy set D and λ_Q is the degree of neighbourhood of the two extreme quantifiers. The truth values of the existential quantifier $T_{\exists, A/D}$ and the universal quantifier $T_{\forall, A/D}$

were defined as:

$$T_{\exists,A/D} = \Delta_{k=1}^{N}\mu(a_k)\nabla\mu(d_k) \quad (3)$$

$$T_{\forall,A/D} = \nabla_{k=1}^{N}(1-\mu(d_k))\Delta\mu(a_k) \quad (4)$$

where a_k and d_k are the membership functions of fuzzy sets A and D respectively, ∇ represents a *t-norm* and Δ represents a corresponding *t-conorm*. By using fuzzy subsethood values as the *degree of neighbourhood* (λ_Q) of the quantifiers, any possible quantifiers that exist between the existential and universal quantifiers can be created in principle. Initially, all linguistic terms of each attribute are used to describe the antecedent of each rule. This may look tedious, but the reason for keeping this complete form is that every linguistic term may contain important information that should be taken into account. The continuous fuzzy quantifiers are created using information extracted from data and behave as a modifier for each of the fuzzy terms. The resulting FuzzyQSBA ruleset can be simply represented by:

$$R_k = \underset{i=1\ldots m}{\nabla}\big(\underset{j=1\ldots n}{\Delta}(Q(A_{ij},D_k)\nabla\mu_{A_{ij}}(x))\big), \quad k=1,2,\ldots,n \quad (5)$$

where $Q(A_{ij},D_k)$ are fuzzy quantifiers and $\mu_{A_{ij}}(x)$ are fuzzy linguistic terms. As both the quantifiers and the linguistic terms are fuzzy sets, choices of *t-norm* operators can be used to interpret $\nabla(Q(A_{ij},D_k),\mu_{A_{ij}}(x))$ whilst guaranteeing that the inference results are fuzzy sets. Based on the definitions of the fuzzy subsethood value, fuzzy existential quantifier and fuzzy universal quantifier (Equations (1,3,4)), it can be proved that if λ_Q is equal to zero then the truth-value of quantifier Q will also equal zero. Thus, during the rule generation process, the emerging ruleset is simplified as any linguistic terms whose quantifier has the truth-value of zero will be removed automatically from the fuzzy rule antecedents, reducing considerably the seeming complexity of the learned ruleset. As commonly used in rule-based systems for classification tasks, the concluding classification will be that of the rule whose overall weight is the highest amongst all.

Rule Extraction

Fuzzy quantifiers have been employed in FuzzyQSBA with the intention to increase the readability of the resulting fuzzy rules and to improve the transparency of the rule inference process. However, the structure of the rules is still very complex. Thus, although the use of quantifiers will make the rules more readable, it seems that it does not increase the comprehensibility of the fuzzy rules. As an alternative, a rule simplification process that is based on fuzzy quantifiers is proposed below. In (Bordogna and Pasi, 1997), fuzzy quantifiers are suggested to be used as a fuzzy threshold. The basic idea of a fuzzy threshold is extended here to conduct the rule simplification process for FuzzyQSBA. This is to offer flexibility in accepting or rejecting any particular linguistic term to represent a particular linguistic variable in a fuzzy rule. To employ the rule simplification, the following fuzzy quantifiers and fuzzy antonym quantifiers are proposed:

$$T_Q(\eta) = \begin{cases} 1 & \text{if } T_Q(\lambda) \geq \eta, \\ \frac{T_Q(\lambda)}{\eta} & \text{if } T_Q(\lambda) < \eta \end{cases} \quad (6)$$

$$T_{antQ}(\eta) = \begin{cases} 1 & \text{if } T_Q(\lambda) \leq 1-\eta, \\ \frac{1-T_Q(\lambda)}{\eta} & \text{if } T_Q(\lambda) > 1-\eta \end{cases} \quad (7)$$

where $T_Q(\lambda)$ is the truth value of quantifier (TVQ) associated with each linguistic term in Equation (5) and η is a threshold value that can be defined as:

$$\eta = p \times \omega \quad (8)$$

where p is a factor for the maximum TVQ, ω. In this technique, the decision to accept a particular linguistic term is made locally without affecting other variables. The aim of using a fuzzy threshold is to soften the decision boundary in the process of accepting or rejecting any terms to be promoted as antecedents of a fuzzy rule, whilst at the same time significantly reducing the number of terms in the induced fuzzy rules. The fuzzy quantifiers mentioned above can be interpreted as 'at least η' and its antonym 'at most $1-\eta$'.

Rule Simplification

The rule simplification algorithm is as follows:

1. For each variable, select the maximum TVQ and calculate $T_Q(\eta)$ and $T_{antQ}(\eta)$ for each linguistic term.
2. For $i = 1, 2, \ldots, l$ where l is the number of linguistic terms for a variable, and for $m \neq n$, calculate:

$$\delta(T_{Q_i}(\eta)) = |T_{Q_m}(\eta) - T_{Q_n}(\eta)|$$

$$\delta(T_{antQ_i}(\eta)) = |T_{antQ_m}(\eta) - T_{antQ_n}(\eta)|$$

3. Conduct the following test: if $\min_i\{\delta(T_{Q_i}(\eta))\} \geq \{\delta(T_{antQ_i}(\eta))\}$ then choose the negation of terms with the lowest TVQ to represent the conditional attribute; else choose the term with the highest TVQ.
4. Create a simplified rule using the accepted linguistic terms (or negation of the terms).

Note that when $\eta = 1$, the fuzzy quantifier and its antonym will become 'most' and 'least', and when $\eta = 0$ the quantifier and its antonym will become 'there exists at least one' and 'for all'. By using the technique proposed above, the primary terms with higher TVQs are accepted to represent the antecedents of the fuzzy rules. By lowering the value of η, the primary terms with a lower TVQ will gradually be accepted. The idea behind this technique is that only the dominant linguistic term (or its negation) will be chosen to represent a particular linguistic variable.

CLASSIFICATION RESULTS

The results of the automated rule induction and simplification obtained using the FuzzyQSBA algorithms described above is shown in Table 3. It can be seen that there is a very good correspondence between the automatically induced rules and the characterisation of the classes obtained from clinical experts shown in Figure 3. Note that the term 'luminal CKs' refers to CK5/6, CK14 and CK18, whereas 'basal CKs' refers to CK7/8 and others. However, in Table 3, the absence of luminal CKs defines membership of classes 4 and 5, as opposed to the presence of basal CKs as mentioned in Figure 3.

CONCLUSIONS

In this paper, we have illustrated the use of consensus clustering to elucidate six separate and distinct classes from the original data set. Further clinical investigations have confirm that these classes form well-chacterised sub-types of breast cancer with distinct clinical characteristics (Soria et al., 2010). We have then presented a rule simplification process (Rasmani et al., 2009) to accompany the FuzzyQSBA rule induction algorithms described previously (Rasmani and Shen, 2006) which results in a simple, comprehensible classification table for each of the six classes based on only ten biomarkers.

In future, we aim to implement the resultant fuzzy rule table in a model-based classification system that can be used to determine the type (class) of cancer in new patients presenting with breast cancer. We hope to thereby create a clinically useful decision support tool for assisting in the choice of treatment(s) for breast cancer, to improve patient survivability and quality of life (by ensuring appropriate treatments) and to reduce health service costs (by reducing unnecessary treatments).

REFERENCES

Abd El-Rehim, D., Ball, G., Pinder, S., Rakha, E., Paish, C., Robertson, J., Macmillan, D., Blamey, R., and Ellis, I. (2005). High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *Int. Journal of Cancer*, 116:340–350.

Al-Daoud, M. and Roberts, S. (1996). New methods for the initialisation of clusters. *Pattern Recognition Letters*, 17(5):451–455.

Ambrogi, F., Biganzoli, E., Querzoli, P., Ferretti, S., Boracchi, P., Alberti, S., Marubini, E., and Nenci, I. (2006). Molecular subtyping of breast cancer from traditional tumor marker profiles using parallel clustering methods. *Clinical Cancer Research*, 12(3):781–790.

Bezdek, J. (1974). Cluster validity with fuzzy sets. *Journal of Cybernetics*, 3(3):58–73.

Bordogna, G. and Pasi, G. (1997). Application of OWA operators to soften information retrieval systems. In *The Ordered Weighted Averaging Operators: Theory, Methodology and Applications*. LNCS.

Callagy, G., Cattaneo, E., Daigo, Y., Happerfield, L., Bobrow, L., Pharoah, P., and Caldas, C. (2003). Molecular classification of breast carcinomas using tissue microarrays. *Diagn Mol Pathol*, 12:27–34.

Carpenter, G. and Grossberg, S. (1987). ART2: Stable self-organization of pattern recognition codes for analogue input patterns. *Applied Optics*, 26:4919–4930.

Castellano, G., Fanelli, A., and Mencar, C. (2006). Classifying data with interpretable fuzzy granulation. In *Proceedings of the 3rd International Conference on Soft Computing and Intelligent Systems*.

Diallo-Danebrock, R., Ting, E., Gluz, O., Herr, A., Mohrmann, S., Geddert, H., Rody, A., Schaefer, K., Baldus, S., Hartmann, A., Wild, P., Burson, M., Gabbert, H., Nitz, U., and Poremba, C. (2007). Protein expression profiling in high-risk breast cancer patients treated with high-dose or conventional dose-dense chemotherapy. *Clin Cancer Res*, 13:488–497.

Edwards, A. and Cavalli-Sforza, L. (1965). A method for cluster analysis. *Biometrics*, 21(2):362–375.

Filkov, V. and Skiena, S. (2003). Integrating microarray data by consensus clustering. In *Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence*, pages 418– 426.

Friedman, H. and Rubin, J. (1967). On some invariant criteria for grouping data. *Journal of the American Statistical Association*, 62(320):1159–1178.

Hartigan, J. (1975). *Clustering Algorithms*. Wiley series in probability and mathematical statistics. Applied Probability and Statistics. New York: Wiley.

Jacquemier, J., Ginestier, C., Rougemont, J., Bardou, V.-J., Charafe-Jauffret, E., Geneix, J., Adélaïde, J., Koki, A., Houvenaeghel, G., Hassoun, J., Maraninchi, D., Viens, P., Birnbaum, D., and Bertucci, F. (2005). Protein expression profiling identifies subclasses of breast cancer and predicts prognosis. *Cancer Res*, 65:767–779.

Kamangar, F., Dores, G., and Anderson, W. (2006). Patterns of cancer incidence, mortality, and prevalence across five continents: Defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol*, 24:2137–2150.

Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: an Introduction to Cluster Analysis*. Wiley series in probability and mathematical statistics. Applied Probability and Statistics. New York: Wiley.

Kellam, P., Liu, X., Martin, N., Orengo, C., Swift, S., and Tucker, A. (2001). Comparing, contrasting and combining clusters in viral gene expression data. In *Proceedings of 6th Workshop on Intelligent Data Analysis in Medicine*.

Maindonald, J. and Braun, W. (2003). *Data Analysis and Graphics Using R — An Example-Based Approach*. Cambridge University Press.

Marriot, F. (1971). Practical problems in a method of cluster analysis. *Biometrics*, 27(3):501–514.

Maulik, U. and Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(12):1650–1654.

Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, 52:91–118.

Table 3: Simplified ruleset created using automated FuzzyQSBA pruning

Classification	Variables									
	ER	PgR	CK18	CK5/6	CK14	HER2	HER3	HER4	P53	MUC1
Class 1	HIGH	HIGH	HIGH	LOW	-	LOW	HIGH	HIGH	-	-
Class 2	HIGH	HIGH	HIGH	LOW	-	LOW	-	LOW	-	HIGH
Class 3	HIGH	LOW	HIGH	LOW	LOW	LOW	-	-	-	-
Class 4	LOW	LOW	-	-	-	LOW	HIGH	-	HIGH	-
Class 5	LOW	LOW	-	-	-	LOW	-	-	LOW	-
Class 6	LOW	-	HIGH	-	-	HIGH	-	HIGH	-	-

Parkin, D., Bray, F., Ferlay, J., and Pisani, P. (2001). Estimating the world cancer burden: Globocan 2000. *Int J Cancer*, 94:153–156.

Rasmani, K., Garibaldi, J., Shen, Q., and Ellis, I. (2009). Linguistic rulesets extracted from a quantifier-based fuzzy classification system. In *Proceedings of IEEE International Conference on Fuzzy Systems*.

Rasmani, K. and Shen, Q. (2004). Modifying fuzzy subsethood-based rule models with fuzzy quantifiers. In *Proceedings of the 13th IEEE International Conference on Fuzzy Systems*.

Rasmani, K. and Shen, Q. (2006). Data-driven fuzzy rule generation and its application for student performance evaluation. *Applied Intelligence*, 24:305–309.

Scott, A. and Symons, M. (1971). Clustering methods based on likelihood ratio criteria. *Biometrics*, 27(2):387–397.

Soria, D. and Garibaldi, J. (2010). A novel framework to elucidate core classes in a dataset. In *Proceedings of the World Congress on Computational Intelligence*.

Soria, D., Garibaldi, J. M., Ambrogi, F., Green, A. R., Powe, D., Rakha, E., Macmillan, R. D., Blamey, R. W., Ball, G., Lisboa, P. J., Etchells, T. A., Boracchi, P., Biganzoli, E., and Ellis, I. O. (2010). A methodology to identify consensus classes from clustering algorithms applied to immunohistochemical data from breast cancer patients. *Computers in Biology and Medicine*, 40:318–330.

Swift, S., Tucker, A., Vinciotti, V., Martin, N., Orengo, C., Liu, X., and Kellam, P. (2004). Consensus clustering and functional interpretation of gene-expression data. *Genome Biology*, 5:R94.

Vila, M., Cubero, J., Medina, J., and Pons, O. (1997). Using OWA operators in flexible query processing. In *The Ordered Weighted Averaging Operators: Theory, Methodology and Applications*. LNCS.

Wang, X.-Y. and Garibaldi, J. M. (2005). A comparison of fuzzy and non-fuzzy clustering techiques in cancer diagnosis. In *Proceedings of the 2nd International Conference in Computational Intelligence in Medicine and Healthcare*, pages 250–256, Lisbon, Portugal.

Weingessel, A., Dimitriadou, E., and Dolnicar, S. (1999). An examination of indexes for determining the number of clusters in binary data sets. Working Paper No.29.

Witten, I. and Frank, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco.

Yuan, Y. and Shaw, M. (1995). Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, 69(2):125–139.

AUTHOR BIOGRAPHIES

JONATHAN M. GARIBALDI received the BSc (Hons) degree in Physics from Bristol University, UK, and the MSc and PhD degree from the University of Plymouth, UK, in 1984, 1990, and 1997, respectively. He is currently an Associate Professor and Reader within the Intelligent Modelling and Analysis (IMA) Research Group in the School of Computer Science at the University of Nottingham, U.K. Dr Garibaldi has published over 80 papers on fuzzy expert systems and fuzzy modelling, including three book chapters, and has edited two books. His main research interests are modelling uncertainty in human reasoning and especially in modelling the variation in normal human decision making, particularly in medical domains. He has created and implemented fuzzy expert systems, and developed methods for fuzzy model optimisation. His email is `jmg@cs.nott.ac.uk` and his personal webpage is at `http://ima.ac.uk/garibaldi`.

DANIELE SORIA is currently a post-doctoral research associate within the Intelligent Modelling and Analysis Research Group, School of Computer Science, University of Nottingham. Daniele Soria received his BSc and MSc in Applied Mathematics from the University of Milan, Italy, in 2004 and his PhD, entitled 'Novel Methods to Elucidate Core Classes in Multi-Dimensional Biomedical Data', in Computer Science from the University of Nottingham, UK, in 2010. His research interests include data mining, bioinformatics and medical applications. His email is `dqs@cs.nott.ac.uk` and his personal webpage is at `http://ima.ac.uk/soria`.

KHAIRUL A. RASMANI is a lecturer at the Faculty of Information Technology and Quantitative Sciences, Universiti Teknologi MARA, Malaysia. He received his Masters Degree in Mathematical Education from University of Leeds, UK in 1997 and his Ph.D. degree from University of Wales, Aberystwyth, UK in December 2005. His research interests include fuzzy approximate reasoning, fuzzy rule-based systems and fuzzy classification systems. In 2009, he was a visiting researcher in the IMA Research Group, University of Nottingham, UK, during which time the majority of the work on fuzzy rule simplification was carried out.