Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference

Modeling Competence for Case Based Reasoning Systems Using Clustering

Abir Smiti and Zied Elouedi
LARODEC, Université de Tunis, Tunisia
Institut Supérieur de Gestion de Tunis
smiti.abir@gmail.com
zied.elouedi@gmx.fr

Abstract

The success of the Case Based Reasoning (CBR) system depends on the quality of the case data. This quality is dedicated to the study of the case base competence which is measured by the range of problems that can be satisfactorily solved. In fact, modeling case-base competence is a clamorous issue in the discipline of CBR. However, the existence of erroneous cases as noises and the non uniform problem distributions has not been considered in the proposed computing competence.
In this paper, we propose a novel case base competence model based on Mahalanobis distance and a clustering technique named DBSCAN-GM. The advantage of this newly proposed model is its high accuracy for predicting competence. In addition, it is not sensitive to noisy cases and it takes account the situation of the distributed case-base. Withal, we contest that this model has a conspicuous role to play in future CBR research in fields such as the development of new policies for maintaining the case base.

Introduction

One of the great aspirations of Artificial Intelligence (AI) is to create smart methods and systems able to understand and emulate human reasoning. Among the various intelligent system paradigms, Case Based Reasoning (CBR) (Kolodner 1992) (Aamodt and Plaza 1994) (Hahn and Chater 1998) is a relatively recent technique that is attracting increasing attention. It is a diversity of reasoning by analogy where it is a technique to model the human way in reasoning and thinking. CBR is able to find a solution to a problem by employing its luggage of knowledge or experiences which are presented in form of cases. To solve the problems, CBR system calls the past cases, it reminds to the similar situations already met. Then, it compares them with the current situation to build a new solution which, in turn, will be incorporated it into the existing case base (CB) (see Fig. 1).
Literally, CBR has been used to concoct multitudinous applications in a wide range of domains including financial analysis, risk assessment, manufacturing, medicine, law, technical maintenance, quality control, etc.

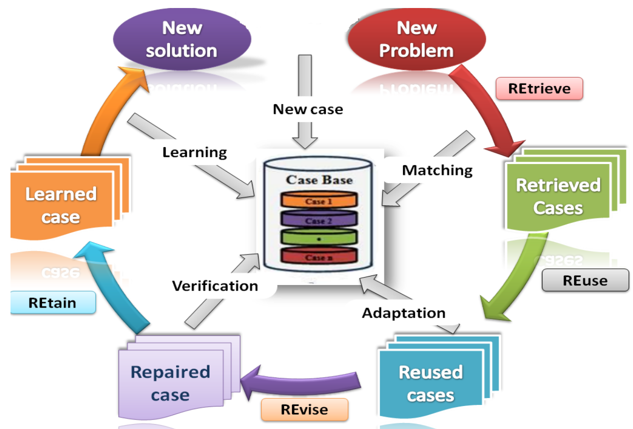

Figure 1: Case based reasoning cycle

Figuratively, CBR system is built to work for long periods of time, it is developed to deal with large amounts of information and cases, it adds cases to the case base through the retain process. As a result, the case base can grow very fast in the sense that it can affect negatively the CBR's quality results and can slow the speed of the query execution time concerning case-research phase. Resultantly, there has been a significant increase in the research area of Case Base Maintenance (CBM). Its objective is to guarantee a good operating in time of an information processing system and to facilitate future reasoning for a particular set of performance objectives (Leake and Wilson 2001). Recently, the case base maintenance issue has drawn more and more attention to a major gauge which is case base competence or case base coverage that supply to the evaluation of a case base. It is a decisive determinant contributing to the performance of a CBR system in the sense that a good quality case base must have high competence. It can be defined as the range of problems that can be satisfactorily solved (McKenna and Smyth 1998).
Several previous approaches for case base coverage can be sensitive to the presence of disagreeable cases such as noises which are those whose descriptions are academic in nature and if memorized in a case base, may cause the solutions to be spurious. In addition, many cases have approximately uniform coverage and others have very small coverage; thus,

it is difficult to distinguish between these cases types.
In this paper, we present a novel case base competence model based on Mahalanobis distance and a clustering technique named DBSCAN-GM. The advantage of this newly proposed model is its high accuracy for predicting competence. In addition, it is not delicate to noisy cases, as well as, it is virtuous for more general distributed case-bases.
The rest of the paper is organized as follows: In Section 2, some of strategies for modeling the competence of the case base will be approached. Section 3 describes in detail our new approach for coverage case base. Section 4 details and analyzes experimental results carried out on data sets from the U.C.I. repository (Asuncion and Newman 2007). Finally, Section 5 ends this work and presents future works.

Case-base Competence: Related work

The competence of case base which is the range of problems the CBR can solve, has been given much attention in the literature. This measure is an essential tool for use in all stages of system development. It particularly important during system maintenance, where knowledge is added, deleted and modified to effect system adaptation and improvement. However, it is difficult to measure the competence of the system, in addition the precise nature of the relationship between the case base and competence is complex and not well understood. So, we need a theoretical model that allows the competence of a case-base to be evaluated and predicted. Many different approaches to model the competence (Smyth and McKenna 1999b; 1999a; Reinartz, Iglezakis, and Roth-Berghofer 2000; Smiti and Elouedi 2011) have shown how different cases can make very different types of competence contribution:

(Smyth and Keane 1995) and (Smyth and McKenna 1999a) defined two key fundamental concepts which are coverage and reachability.

- Coverage is an important competence property. Coverage of a case is the set of target problems that it can be used to solve. The overall coverage of a case base in relation to a set of queries is the total number of covered queries divided by the total number of queries in the query set.
- Reachability is an important competence property. Reachability of a target problem is the set of cases that can be used to provide a solution for the target.

In order to have a CB with good competence, its coverage ratio must be high.
(Grachten, García, and Arcos 2005) consider a case is significant in the CB if it covers many similar cases: its similarity value (sim) should be greater than a verge Ψ. Hence, the concept of coverage can be extended in the following way:

$$Cov(c_i \in CB) = \{c_j \in CB | sim(c_i, c_j) > \Psi\} \quad (1)$$

Where Ψ is the quality criterion, it can take values in the interval $[0, 1]$. Based on many tests, the verge Ψ can be defined using an hierarchical competence model.

The competence model proposed by (M&S) (Smyth and McKenna 2001) is a copacetic contribution of the analysis of case base structure by assessing the local competence contributions of cases and their interactions. It is assuming that the competence is based on a number of factors including the size and the density of cases. The number and density of cases can be readily measured. In fact, The individual competence contribution of a single case within a dense collection will be lower than the contribution of the same case within a sparse group; dense groups contain greater redundancy than sparse groups (Smyth and McKenna 1999a). The density of an individual case, that we named $Dens$, can be defined as the average similarity (Sim) between this case and other clusters of cases called competence groups (Equation 2). Hence, the density of a cluster of cases is measured as a whole as the average local density over all cases in the group (Equation 3).

The coverage of each competence group is an estimate of the problem space area that the group covers. As indicated above group coverage must be directly proportional to the size of the group but inversely proportional to its density (Smyth and McKenna 1999a). This leads to the definition of group coverage shown in Equation 4.

$$Dens(c, G) = \frac{\sum_{c' \epsilon G-c} Sim(c, c')}{|G-1|} \quad (2)$$

$$Dens(G) = \frac{\sum_{c \epsilon G} Dens(c, G)}{|G|} \quad (3)$$

$$Cov(G) = 1 + ||G| \times (1 - Dens(G))| \quad (4)$$

Where $|G|$ is the number of cases in the group G.
In the final step, the overall competence of a case-base can be defined as the sum of the coverage of all competence groups. As a result, for a given case-base, with competence groups $G = G_1, ..., G_n$, the total coverage is defined as following:

$$TotalCoverage(G) = \sum_{G_i \in G} Cov(G_i) \quad (5)$$

(Yang and Zhu 2001; Salamó and Golobardes 2003) described case coverage based on a rather rough concept of case neighborhood.

These works have highlighted the importance of modeling CBR competence. However they suffer from some shortcomings such as they are not always meticulous, especially in the situation of non-uniform distributed case-bases, as shown in (Shiu, Yan, and Wang 2001; Massie, Craw, and Wiratunga 2007). Besides, they scan the entire case base for the categorization which is not evident and they are hypersensitive to erroneous cases as noises (Pan, Yang, and Pan 2005).

To alleviate this potential problem, we propose, in this paper, a novel approach for computing case base coverage, named CMDC- Coverage model based on Mahalanobis distance and Clustering. We hold a different point of view for modelling the competence. The innovation of our work consists of proposing efficient techniques of machine learning to distinguish the important cases, which invoke the quality of the system, whether noisy cases or isolated cases or similar cases.

CMDC- Coverage model based on Mahalanobis distance and Clustering

Coverage of a case is the set of target problems that it can be used to solve. Computing this set for every case and target problem is not a feasible option. Sharply, the best way is to find some approximations to this set.
Hypothetically, we consider that the case base is a representative sample of the problem space. Under this circumstance and in order to facilitate the competence computing, a given case base can be decomposed into groups of closely related cases. The competence of the case base as a whole is computed as the sum of these group coverage values. This is valid because we suppose, that each group of cases is considered as an autonomous set makes an independent coverage contribution.
As was seen in previous efforts, the competence of the case base is proportional to the individual coverage contribution of a single case within a determined groups distribution. which is related to the size of the case base.

$$Comp\%(CB) = |1 - \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} Cov(x_{ij})}{SizeCB}| \quad (6)$$

where k is the number of groups and Cov is the coverage contribution of each case in one cluster j with given distribution. This value depends on the type of the case and its role in the CB. Authentically, to obtain a good approximation of the coverage computing, we have to consider these notes where we define three types of cases:

- CN_i: Noisy cases are a distortion of a value or the addition of the spurious object. They are disagreeable cases, they can dramatically slow the classification accuracy. As a result, the CBR's quality will be reduced. They mislead the computation of the CB's coverage because there is no other case that can solve them and they can not cover other cases. In analytical tasks, CN_i are cases that do not belong to any set of similar cases. The best choice in this situation, is to detect cases expected to be noisy and affect them an empty set as a coverage value.

$$Cov(CN_i) = \emptyset \quad (7)$$

- CS_i: Each case from a group of similar cases and which is near to the group centroid, provides similar coverage values, because they are close to each other, they cover the same set of cases. Hence, the coverage value of each case equals to the number of cases in this group (n).

$$Cov(CS_i) = n \quad (8)$$

- CI_i: In a set of similar cases, there are cases that are much distant to the other members in this group. We can consider them as isolated cases. They belong to one set of similar cases not like those of type CN_i but they are farther to the set's centroid than the cases of type CS. They cover only themselves. As a result, the coverage of each case of this type equals to one.

$$Cov(CI_i) = 1 \quad (9)$$

Based on these definitions, we create a new coverage model named CMDC- Coverage model based on Mahalanobis distance and Clustering, our model distinguishes these three types of cases and affects the appropriate coverage value to each type.

To apply this idea, we need first to create multiple, small groups from the case base that are located on different sites. Each small group contains cases that are closely related to each other. This can be done only by a clustering technique because it ensures that each group is small, independent and contain similar cases, so it is easier to detect the different types of cases.
After that, for each small cluster: the cases, which are near to the cluster's center and close to each other, are considered as cases of the type of CS_i. The cases, which are far away from the center, are considered as cases of the type of CI_i. Finally, the cases, which are outside the clusters and have not affected to a determined cluster, are considered as cases of the type of CN_i (See Fig. 2).

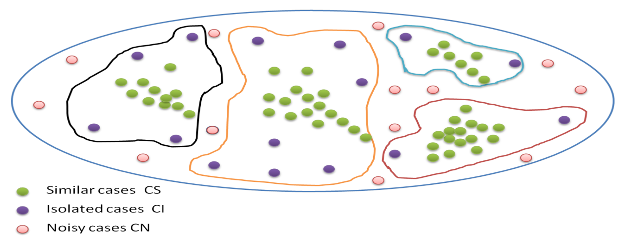

Figure 2: Different types of cases for CMDC model

Among the proposed clustering approaches, we should, ideally, use a method that while clustering and creating groups of similar cases, can smooth the discover of the different types of cases in such data sets. So, it should have the maximum of the following main properties:

- It is totally automatic; in particular, a principled and intuitive problem formulation, such that the user does not need to set parameters, especially the number of clusters K. In fact, if the user is not a domain expert, it is difficult to choose the best value for K and in our method of coverage, we need to have the best number of clusters to determinate the groups of similar cases.
- It has the capability of handling noisy cases like that we detect the cases of type CN_i
- It supports arbitrary shape of clusters and it can handle non-uniform distributed case-bases.
- It scales up for large case base.

To overcome all these conditions, we use a new clustering method called "DBSCAN-GM" (Smiti and Elouedi

2012). It combines Gaussian Means (Hamerly and Elkan 2003) and density-based clustering method (DBSCAN) (Ester et al. 1996) methods. DBSCAN-GM clustering method benefits from the advantages of both algorithms to cover the conditions cited above: The first stage, DBSCAN-GM runs Gaussian-Means to generate automatically a set of clusters with their centers, in purpose to estimate the parameters of DBSCAN. In this manner, the parameter identification problem of DBSCAN is solved. The second stage, it runs DBSCAN with their determined parameters to handle noises and discover arbitrary size and shaped clusters. In this fashion, the noisy data shortcoming of Gaussian-Means is unraveled. In briefly, the steps of the DBSCAN-GM is described as follows:

Algorithm 1 Basic DBSCAN-GM Algorithm

1: Begin
2: Run GMeans and find the center of all sets M_j
3: /* Estimate the parameters EPS and Minpts of DBSCAN-GM:*/
4: For each cluster j with center M_j do:
  /* The radius r_j of the cluster j*/
  Calculate $r_j = \sqrt{\frac{\sum_{i=1}^{|G|} distance^2(M_j, x_{ij})}{N_j}}$
  /* The $Minpts_j$ of the cluster j*/
  Calculate $Minpts_j = \frac{\Pi \times r_j^2}{Total_V olume_j} \times N_j$
  Where $TotalVolume_j = \frac{4}{3} \times \Pi \times r_j^3$
5: The EPS value is: $Eps = Min(r_j)$
6: The Minpts value is: $Minpts = Min(Minpts_j)$
7: Run DBSCAN(Eps, Minpts)
8: End

The DBSCAN-GM is an appropriate clustering method for our CMDC coverage technique. This occurs because this clustering method shows good performance on large databases, it relieves the problem of the global parameters; it automatically generates a set of clusters. Besides, it has the possibility of detecting noisy instances which we will affect them an empty set as coverage value. In addition, it generates non-uniform distribution and different shapes for the clusters, comparing to other clustering algorithms.

Once we have partitioned the original case memory by DBSCAN-GM and we have detected cases expected to be noisy (CN_i) and accorded them an empty set as a coverage value, CMDC directs attention to finding the other types:
Based an intrinsic assumption, the CS_i are in the center of the cluster space and follow a normal distribution and occur in a high probability region of this cluster. However, the isolated cases CI_i are located at the border of the cluster space and deviate strongly from the cluster distribution. They have a low probability to be generated by the overall distribution and they deviate more than the standard deviation from the mean.

For each cluster, the cases which are distant from the clusters center are considered as cases of type CI. The best technique to detect them, is Mahalanobis distance (Filzmoser, Garrett, and Reimann 2005) because it takes into account the covariance among the variables in calculating distances. With this measure, the problems of scale and correlation inherent in the other distance such as Euclidean one are no longer an issue. In addition, Mahalanobis distance is an efficient for the non uniform distribution and arbitrarily shaped clusters because it deals with clusters of different densities and shapes.
Given p-dimensional multivariate sample (cases) x_i $(i = 1; 2...; n)$, the Mahalanobis distance is defined as:

$$MD_i = ((x_i - t)^T C_n^{-1}(x_i - t)^{1/2} \tag{10}$$

Where t is the estimated multivariate location and C_n the estimated covariance matrix:

$$C_n = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X_n})(x_i - \overline{X_n})^T \tag{11}$$

Where $\overline{X_n}$ is the mean of the cluster and n is the number of cases in this cluster.
Accordingly, those observations with a large Mahalanobis distance in a cluster are selected as CI type. The threshold of large distance depends on when the similarity between cases and the center starts raising.
For that, we need to measure how closely the cases cluster around the mean and how are spread out in a distribution of the cluster. Hence, the case whose Mahalanobis distance is superior to this threshold, will be consider as CI case, else it will be CS type.
This measure is substantiated by the standard deviation that is able to know how tightly cases are clustered around the center. It indicates how much, on average, each of the cases in the distribution deviates from the center of the distribution because it depends on the average of the squared deviations from the mean of the cluster. The standard deviation is therefore a good measure of the categorization of CI and CS cases. The formula is as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X_n})^2}{n}} \tag{12}$$

Analogously, we can define the similar cases as a cases with a tiny Mahalanobis distance and close to the center.
$\{x_i, z_j\}\ \epsilon\ Cluster_j$ with $center(z_j)$:

$$Cov(x_i) = \begin{cases} 1 & \text{if } MD(x_i, z_j) > \sigma \\ |CS| & \text{otherwise} \end{cases}$$

As a result, we have affected for each case the appropriate coverage value depending on its type.

Experimental Analysis

In previous sections, a new model using clustering and Mahalanobis distance for modeling the coverage of case bases has been presented. In this section, empirical evidence is needed to support this model. In short, we mention that the propose model in this paper carefully matches the actual competence and we think that our CMDC benefits from superior effectiveness and performance in terms of competence. We test our algorithm on a number of case bases. Our experiments are performed on several publicly available datasets. Actually, in this paper, we use public datasets obtained from the U.C.I. repository of Machine Learning databases (Asuncion and Newman 2007). In fact, our competence model was implemented on three classification datasets: Iris with size of 150 cases, Ionosphere with 351 cases and Mammographic with the number of 961 cases.

Experiments have been carried out using it applied to classification issues. In this case, we use the nearest neighbors (K-NN) for the classification. In general, the retrieval technique, which is used by the major CBR applications, is (K-NN) algorithm. It is a simple approach that computes the similarity between stored cases and new input case. Hence, we choose to select the (1-NN) to compute the percentage of correct classification. We apply the 1-NN algorithm to the same datasets and the same task to obtain the average accuracy rate.

In the first part of our experimentation, our competence model was applied to each case-base and its predicted competence compared to the test set accuracy, as following:
This experiment uses different size of cases bases. Each of these datasets was split into training set and test set approximately. The training set contains 80% of case base and the test set contains 20% of cases. Initially, the training set was partitioned into five independent sets. Each case is chosen randomly such that the case base satisfies non-uniform distribution. Every case is a two-dimension vector. The smallest case-base was created using one of these sets, and a growing case-base was created by successively adding one of these sets. For each of the five case-bases two measurements are taken: the case-base accuracy yielding a real competence value for each case-base. Second, our competence model is built for each case-base and case-base coverage is measured, yielding a predicted competence value for each case-base.
We use, in this situation, the correlation coefficient, to measure the relationship between the CB accuracy and the predicted competence model. Actually, this coefficient is a number between 0 and 1. If there is no relationship between the predicted values (our CMDC competence) and the actual values (PCC), the correlation coefficient is 0 or very low. As the strength of the relationship between the predicted values and actual values increases so does the correlation coefficient. Thus the higher the correlation coefficient the better.
Figs. 3 shows the experimental corollaries of three different datasets which have been presented in different five case bases. They present the accuracy and case base competence plotted against case base size:

The results afford meritorious support in favour of our competence model. There appears to be a very abutting connection between the two curves and hence a strong correlation between predicted competence provided by our CMDC model and the test set accuracy. It can be seen from these graphs that there is a strong correlation between the model's predictions and test set accuracy. For instance, the correlation coefficient between the two curves, in Iris example, is 0.94, which is statistically powerful. Moreover, for Mammograph dataset, the correlation is greater than 0.92, and for Ionosphere, its is equal to 0.88.

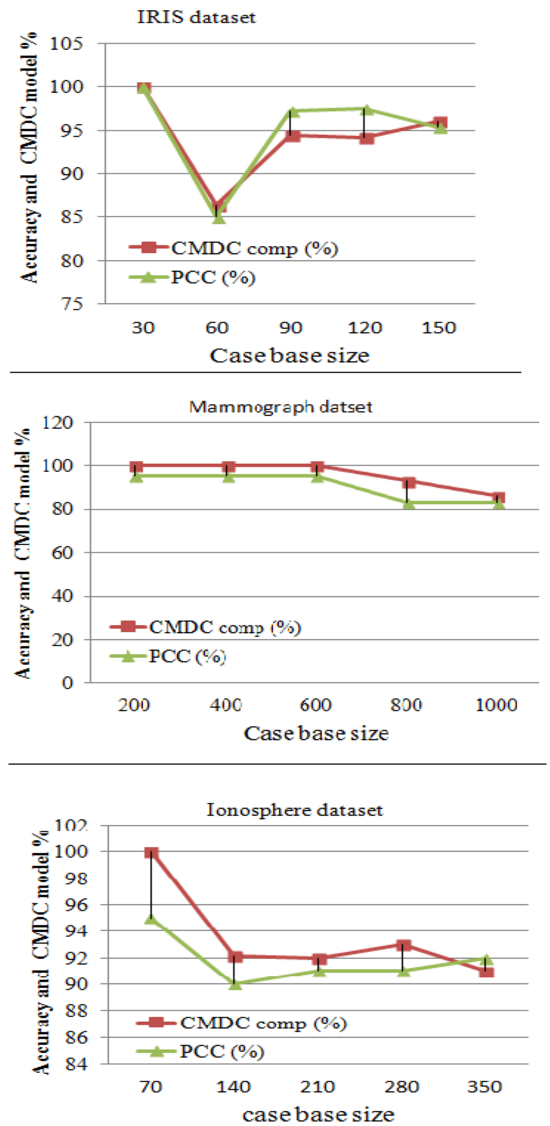

Figure 3: Comparing predicted competence (CMDC comp) to the case-base accuracy (PCC) for different Case-base Sizes of the Iris, Ionosphere and Mamograph datasets

Looking at the plots in more detail, we find that the plot of the model's prediction and the one of the accuracy express a direct relationship, if one increases the other increases, one decreases the other decreases. Likewise, We can observe that the plot of the true competence (PCC) is matched by model's prediction in some sizes. That's prove that the evaluation results exhibit a quintessential correlation between the predictions of the model and true competence for a spacious case-base sizes.

In the second part of our experimentation, we realized that our CMDC model can be appreciated when compared with the well-known competence model (M&S) (Smyth and

McKenna 2001), using the same benchmark data sets as described above. For this comparison, we use Percentage error as evaluated index (Equation 13), which represents the relative error of coverage computed by using the (M&S) model and our new CDCM model respectively.

$$Error(\%) = \frac{|EstimateComp - PCC|}{PCC} * 100 \qquad (13)$$

The experiment results are computed and are shown in Table 1. The results positively support our model.

Table 1: Comparing CMDC model to (M&S) technique

Dataset	(M&S)	CMDC
IRIS	4.010	0.927
Ionosphere	3.544	0.287
Mammograph	21.10	13.820

Conspicuously, the results are very encouraging. The Percentage error of our CMDC model is rather lower than using the (M&S) model. This is due to the fact that the (M&S) model suffers from two lacks: The group case density does not give a good measure of competence and problem complexity is not adequately reflected by the model. However, our model alleviates these problems. Its point of strength is that it takes into account the type of cases and is not sensitive to noise or the nature of the group's distribution.

Conclusions

The research in case base reasoning has spotlighted the momentous of modeling CBR competence in order to reduce the compulsion of case base evaluation experiments and to pamper case base maintenance research.

In this paper, a novel competence model of case-bases is proposed. It is based on Mahalanobis distance and a clustering technique named DBSCAN-GM. The advantage of this newly proposed model is its high accuracy for predicting competence. In addition, it takes account the nature of the distribution and the different types of cases such as the noises, isolated cases and similar cases.

The results of experiments escorted are very forceful and positive for our model. It shows that the new model proposed in this paper has extended the scope of modeling case-base competence. Future tasks include applying the model in other CBR domains. In particular we will begin to focus on maintaining the CBR systems that will apply this model.

References

Aamodt, A., and Plaza, E. 1994. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications* 7(1):39–52.

Asuncion, A., and Newman, D. 2007. UCI machine learning repository. http://www.ics.uci.edu/mlearn.

Ester, M.; Kriegel, H.-P.; Sander, J.; and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceeding of 2nd International Conference on Knowledge Discovery and*, 226–231.

Filzmoser, P.; Garrett, R. G.; and Reimann, C. 2005. Multivariate outlier detection in exploration geochemistry. *Computers and Geosciences* 31:579–587.

Grachten, M.; García, F. A.; and Arcos, J. L. 2005. Navigating through case base competence. *Proceedings of the 6th international conference on Case-Based Reasoning Research and Development* 282–295.

Hahn, U., and Chater, N. 1998. Understanding similarity: A joint project for psychology, case-based reasoning and law. *Artificial Intelligence Review* 393–427.

Hamerly, G., and Elkan, C. 2003. Learning the k in k-means. volume 17. MIT Press.

Kolodner, J. L. 1992. An introduction to case-based reasoning. *Artif. Intell. Rev.* 6(1):3–34.

Leake, D. B., and Wilson, D. C. 2001. Maintaining case-based reasoners: Dimensions and directions. *Computational Intelligence* 17:196–213.

Massie, S.; Craw, S.; and Wiratunga, N. 2007. When similar problems don't have similar solutions. In *ICCBR*, 92–106.

McKenna, E., and Smyth, B. 1998. A competence model for case-based reasoning. *9th Irish Conference on Artificial Intelligence and Cognitive Science* 208–220.

Pan, R.; Yang, Q.; and Pan, S. J. 2005. Competence driven case-base mining. In *Proceedings of the Twentieth National Conference on Artificial Intelligence, AAAI*, 228–233.

Reinartz, T.; Iglezakis, I.; and Roth-Berghofer, T. 2000. On quality measures for case base maintenance. In *Proceedings of the 5th European Workshop on CaseBased Reasoning*, 247–259. SpringerVerlag.

Salamó, M., and Golobardes, E. 2003. Hybrid deletion policies for case base maintenance. In *The Florida Artificial Intelligence Research Society FLAIRS*, 150–154.

Shiu, C.; Yan, L.; and Wang, X. Z. 2001. Using fuzzy integral to model case-base competence. In *Workshop on Soft Computing in Case Based Reasoning*, 301–310.

Smiti, A., and Elouedi, Z. 2011. Article: Overview of maintenance for case based reasoning systems. *International Journal of Computer Applications* 32(2):49–56. Published by Foundation of Computer Science, New York, USA.

Smiti, A., and Elouedi, Z. 2012. Dbscan-gm: An improved clustering method based on gaussian means and dbscan techniques. In *International Conference on Intelligent Engineering Systems (INES)*, 573–578. IEEE Computer Society.

Smyth, B., and Keane, M. T. 1995. Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems. In *the 14th International Joint Conference on Artificial Intelligent*, 377–382.

Smyth, B., and McKenna, E. 1999a. Building compact competent case-bases. In *Proceedings of the Third International Conference on Case-Based Reasoning*, 329–342. Springer.

Smyth, B., and McKenna, E. 1999b. Footprint-based retrieval. In *ICCBR*, volume 1650 of *Lecture Notes in Computer Science*, 343–357. Springer.

Smyth, B., and McKenna, E. 2001. Competence models and the maintenance problem. *Computational Intelligence* 17(2):235–249.

Yang, Q., and Zhu, J. 2001. A case addition policy for case-base maintenance. *Computational Intelligence Journal, A Special Issue on Case-Base Maintenance* 17(2):250–262.