Assessing Confidence in Cased Based Reuse Step

F. Alejandro García [a,1], Javier Orozco [b] Jordi Gonzàlez [c] and J. Lluís Arcos [a]

[a] *Artificial Intelligence Research Institute IIIA-CSIC,*
Campus UAB, 08193 Bellaterra, Spain
[b] *Computer Vision Center CVC & Dept. de Ciències de la Computació,*
Edifici O, Campus UAB, 08193 Bellaterra, Spain
[c] *Institut de Robòtica i Informàtica Industrial UPC-CSIC,*
C. Llorens i Artigas 4-6, 08028, Barcelona, Spain

Abstract. Case-Based Reasoning (CBR) is a learning approach that solves current situations by reusing previous solutions that are stored in a case base. In the CBR cycle the *reuse* step plays an important role into the problem solving process, since the solution for a new problem is based in the available solutions of the retrieved cases. In classification tasks a trivial reuse method is commonly used, which takes into account the most frequently solution proposed by the set of retrieved cases. We propose an alternative reuse process; we call confidence-reuse method, which make a qualitative assessment of the information retrieved. This approach is focused on measuring the solution accuracy, applying some confidence predictors based in a *k*-NN classifier with the aim of analyzing and evaluating the information offered by the retrieved cases.

Keywords. Case-Based Reasoning, Reuse, Confidence reuse, *k*-NN, Classification

1. Introduction

Case-Based Reasoning (CBR) is a learning approach that imitates human problem solving behavior by solving current situations while reusing previous solution knowledge [1]. The development of CBR systems has increased the necessity of supporting the analysis of the Case-Base (CB) structure by providing solutions with an estimated confidence. Cheetam and Price emphasized the importance for the CBR systems to be capable of attaching a solution confidence [3,2], which means that the system will produce both the solution for the target problem and a value estimating the confidence the system has with the solution proposed .

CBR systems use different classifiers within the problem solving process, where most of them produce numeric scores based on similarity measures. These systems should provide a solution beyond the quantitative value produced by a classifier, this must be a considerably and significantly quality rate to take into account in the decision

---

process. When the numeric scores become to quality scores, it could help to attach a confidence criterion solution proposed. Some approaches in CBR have been focused on keeping the stability and accuracy of traditional artificial intelligence problem-solving systems[10], and have been suggested to deal with confidence performance[7,4,9]. We highlight those works aimed to evaluate confidence in classification tasks as for example, a Spam filter case-based reasoning[5] system employed to identify Spam mail.

In this paper we propose a confidence assessment approach applied to expressive face recognition in a facial expression recognition domain. This domain presents higher complexity for the difficulty of clearly assess the different human face expressions. We have developed a CBR confidence classification system in a multi-class problem while analysing the confidence performance into each class by using confidence predictors based on k-NN classifiers in order to assess the probability of a given target problem to belong to a corresponding class.

This paper is organized as follows: in section 2 the case-based reasoning representation and the classification task into the CBR cycle. Section 3 presents some experimental results and discussions. We conclude in the section 4 summarising the main contributions and introducing some perspectives to extend the proposed approach.

2. CBR Representation

Case-Based Reasoning (CBR) is a learning approach that imitates human problem solving behavior by means of reasoning about current situations and reusing past situations. CBR solves a new problem by retrieving a previous similar situation and by reusing information and knowledge of that situation [1]. The CBR paradigm uses previous solved situations called cases, which are stored in a case-base (CB).

In CBR systems the most commonly used classification technique is the k-NN classifie, which decides the belonging class by assessing its k Nearest Neighbours.

The main tasks that describes a system CBR process is called CBR cycle[1], which contains four steps. Bellow, the two first are described according to facial expression analysis by assessing confidence. The last two steps of the CBR cycle are out of the scope of this paper.

1. **Case Retrieve:** The goal of this step is to retrieve the cases close to the target problem. The k-NN classifier is used in order to retrieve from the CB the k most similar cases to the target problem. This process is based on a similarity measure. We apply an Euclidean distance to compare the $problem\text{-}description$ attributes of the cases.
2. **Case Reuse:** in this step, the goal to achieve is to determine a possible solution for the target problem, by using the previously collected information from the retrieved cases. We could apply a reuse solution method by applying the majority solution from the retrieved cases. This majority-reuse has some drawbacks: the first one is that the k-NN needs to employ an additional strategy for choosing the solution in case of ties (e.g. having a list of retrieve cases as A, A, B, B, C is not easy to solve). The second one is that the classification efficiency is decreasing when noise into the case-base increases. Because of this we use an alternative decision method that we call Confidence-Reuse, which is based on previous confidence assessment to provide a current solution for the target problem. In other words, an evaluation of the confidence of the proposed solution.

2.1. Confidence Assessment

The confidence evaluation is based on four additional similarity measures that we call predictors. They are applied with respect to the k-NN classifier while providing additional scores. Although the predictors could be computed assuming that the target problem belongs to each one of the classes into the CB, we analyze only probable solutions acquired from the retrieved neighbours. Hence, if the retrieved solutions are A, A, B, B, C, C, the target problem is analyzed only with respect classes A, B and C.

We assess the confidence of the solution making some measures based on the k-NN classification. The measures used are similar to others existing in previous work [5], with some differences, our problem has to reason about more than the two classes used in [5] and we applied only four measures in order to keep the system *eficciency*. We assume that both the amount of neighbours and the similarity relationship between cases play an important role in the assessment of the predictors. The four confidence predictors try the sample data in a neighbourhood around to the target problem according to an additional set of similarity measures based on k-NN, Euclidean distance and solution indexing. These predictors follow the next notation:

- For a given target problem t, the neighbours are ordered by similarity distance S, which is the inverse of the Euclidean distance between two cases. The closest neighbour is the most similar.
- The set of retrieved cases is composed by relevant-neighbours (RN) (cases belonging to the same class that the target problem is *assuming*) or irrelevant-neighbours (IN) (cases belonging to different class that the target problem is *assuming*).

The confidence measures are defined as follows:

(a) Average-index, S_1, captures how close are the first n-IN neighbours to the target problem t. This predictor gets the average of the position or index I' for the first n-IN.

$$S_1(t, n) = \frac{1}{n} \sum_{i=1}^{n} I'_i(t) \tag{1}$$

(b) Average RN-similarity, S_2, measures the average of the similarity for the first n-RN neighbours to the target problem t.

$$S_2(t, n) = \frac{\sum_{i=1}^{n} S[t, RN_i(t)]}{|RN|} \tag{2}$$

(c) Similarity ratio, S_3, calculates the ratio of the similarity between the target problem t and its n-RN to the similarity between the target problem t and its n-IN.

$$S_3(t, n) = \frac{\sum_{i=1}^{n} S[t, RN_i(t)]}{\sum_{i=1}^{n} S[t, IN_i(t)]} \tag{3}$$

(d) Similarity Ratio within k, S_4, which is similar to the above measure but only considering the first k-RN and the first k-IN from the k-NN set.

$$S_4(t,k) = \frac{\sum_{i=1}^{k} S[t, RN_i(t)]}{\sum_{i=1}^{k} S[t, IN_i(t)]} \tag{4}$$

The effectiveness of each measure depends on the proportion of cases correctly predicted and the high confidence depends on the predictoros agreement. On the other hand, each predictor produces a positive score, in this way, we identify the measure that extracts the majority of accuracy by using weights. Each one of the predictor's measurements are standardised into the range [0,1] according to the theoretical maximum value. Posterior confidence value λ is computed according to a average function as follows:

$$\lambda = \frac{1}{4}\sum_{i=1}^{4} S_i \tag{5}$$

The $confidence\text{-}reuse$ method, is used to compute the confidence value λ, (Eq. 5). Every λ value offered by (Eq. 5) is between the range [0,1], thus we calculated the percentage of confidence values with ($\lambda * 100$). In this manner we set a confidence-threshold in 80% in order to filter cases, and regarded those cases that achieved the threshold into the CB. This filter was carried out by performing a leave-one-out process[1]. Subsequently, we assume that all λ values follow a Gaussian distribution for each class, $N(\mu_c, \sigma_c)$ for $c = 1, ..., 7$ according to the seven classes.

Once the λ is calculated in each class, we learned the mean μ_c and the standard deviation σ_c of λ values for each class c.

In order to solve a new target problem t, we firstly run a k-NN process to get a set of retrieved solutions. Then we compute the maximum probability of t to belong to each retrieved solutions as follow:

$$P(L_t = c) = P(\lambda_t | \gamma_t) = \prod_{i=1}^{m} N(\lambda_i; \mu_c, \sigma_c) \tag{6}$$

where L_t is the corresponding solution for the target problem, λ_t the confidence value, c is every class in the retrieved cases, and m is the length of the class c. Therefore the classification decision is made according to the minimum distance of each λ_t to the class paramters:

$$Min \left| \frac{\lambda_t(c) - \mu_c}{\sigma_c} \right| \tag{7}$$

3. Experimental Results

The experiments were aimed to analyze the behaviour of classification tasks in a multi-class case-base system. The shown results in this section were collected from the multi-class facial expression data base the FGnet[2], which is an image database that contains face image sequences, showing a number of subjects performing the six different basic emotions and the neutral emotion defined by Ekman and Friesen [6].

---

[1]Leave-one-out process, leaves every case out of the case base in order to solve it as a new problem.
[2]http://www-prima.inrialpes.fr/FGnet/
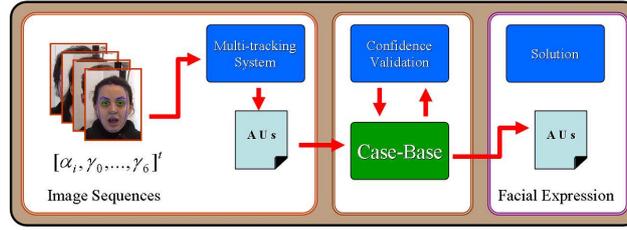
Figure 1. Problem solving system structure

Figure 2. Image sequence for happiness expression.

3.1. Case base building

By using a Multi-tracking system [8], we made use of a Facial Action Coding System (FACS)[3] to describe facial movements and categorise different facial expressions, see Fig.1.

We constructed the case base with a case-structure, which employs the FACS as attributes and adopts the FGnet facial expression label as the solution or class for each element, see Fig.2. Consequently, the case-base is a categorised database which assume $\Theta = [\gamma, \mathbf{L}]$, as case structure, where γ is a seven dimensional continues valued vector for FACS, $\gamma = [$*Upper Lip, Lower Lip, Lip Stretcher, Lip Corners, Inner Brow, Outer Brow, Eyelid*$]$ and $\mathbf{L}$ is the class label (Solution). All the cases are distributed in seven classes.

3.2. System-Training

The first training step is to build up a case-base CB with the same amount of elements per class. The information is coming from image sequences developed by different actors in each class in order to avoid redundant information. In deterministic problems the amount of elements into the case base play an important role. Even thought we demonstrate that working with a small case base with learning approach it is possible to achieve high proportion of example correctly classified, see Fig.3. The case base was cautiously constructed taken into account the solving credibility associated to each element.

3.3. System-Testing

A discrete comparison towards the classification tasks was analyzed by applying confidence-reuse method explained in section 2 and comparing classification results against majority-reuse results. The tests were designed in two ways; the first one was

---

[3]Facial Action Coding System (FACS) is the most widely used and versatile method for measuring and describing facial behaviours.
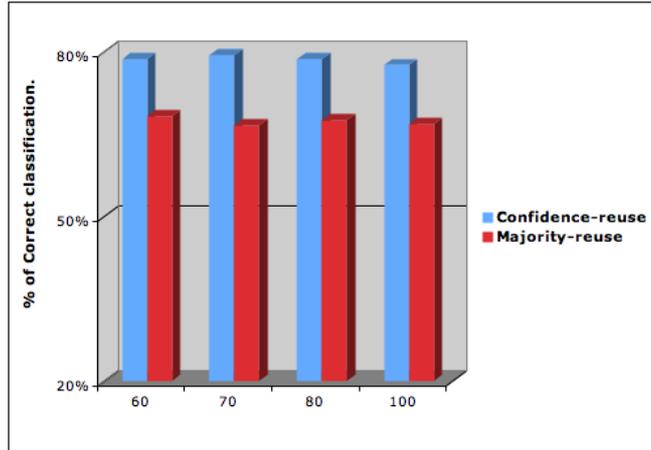
Figure 3. Average Classification using different case base sizes.

aimed to analyze robustness of the method, for that reason we compared results of k-NN classifier, with confidence evaluation and without applying it. We observed that even using different case base sizes, the confidence-reuse method did not suggest worse results than the results proposed by the majority-reuse method. Can be observed in Fig.3 in average 78% of classification correctness achieved by confidence-reuse using all classes of the case base, whilst the majority-reuse achieved in average only 67% using the same cases. It is worth to mention that each image sequence offered by FGnet-database is developed from low expressiveness until achieve the best expressiveness. Some information in the beginning and even in the middle of a sequence could appear as noise in the CB, as we can see in Fig.2, the two first images have the label *happy* but its attributes assume information similar to other classes and this information is easy to misclassify. Due to noise, enhancing the length of elements in each class does not always allow better classification results.

The rest of the tests were focused on analyzing unseen data. We consider as unseen data the image sequences of new actors in each class. We classified at least three sequences by actor, completing a total of 574 new cases. The classification per class is detailed in (Fig.4), where we can see that two important classes *sadness* and *fear* reach better performance using accuracy evaluation. In addition is possible to see the percentage of cases in those classes considered with confidence is low because of the agreement of the predictors is as low. The percentage of confidence is determined by the agreement of the solution proposed by predictors. The aim of each predictor is to evaluate the solutions (without duplicates) nearest to the target problem and re-ranking them according to its own evaluation. Finally the decision making is carried out based in all predictor's agreement.

To evaluate the confidence have an advantage to highlight, it lets know if the solution is more than good or bad, for instance we found with the results of a confusion matrix that too much cases of the class *sadness* are misclassified and confused with the class *neutral*. In this situation, even if the classification improved using confidence-reuse, the confidence evaluation shows that only 7% of cases reach high confidence, this information could be used by system manager in the final decision making, see Fig. 4.
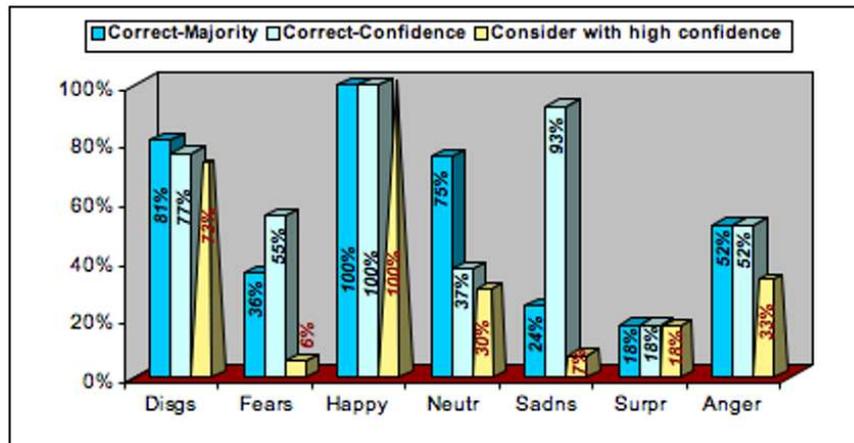
Figure 4. Classification performance by using the Confidence Reuse method.

4. Conclusions

We highlight two advantages of applying confidence evaluation in reuse step in CBR systems, the first one is that to solve some problems it is needed to know more than the solution offered by the nearest neighbours, an additionally, assessment of the information coming from the nearest neighbours lets to reach better classification result as we have demonstrated.

We have shown an efficient sampling process to compare data while enhancing the k-NN classification process. By using the confidence assessment process and the posterior comparison among classes, we could get a higher classification rate, which provide two classification descriptors; the solution and the corresponding confidence. As well, we can compare experimental results for different k-values and case-bases through the classification error and the classification confidence.

The Gaussian assumption for the confidence value in each class have allowed us to set a decision process based on previous confidence solutions and the confidence for the proposed solution, which means an alternative decision according to the quality of the solution. This parameter is worth for maintenance process, since it reveals the quality of the database for each class.

As future work, we attempt to provide an optimal k-value according to a confidence threshold in order to minimize the computational complexity related with the solving process against to the whole database. Additionally, we want to provide a strong aggregation method based on confidence assessment, which will enhance the dynamical learning capabilities of this CBR system.

References

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

[2] W. Cheetam and J. Price. Measure of a solution accuracy in case-based reasoning systems. *In Funk, P.,González-Calero, P.,eds.: 7th European Conference on Case-Based Reasoning (ECCBR 2004)*, 3155:106–118, 2004.

[3] W. Cheetham. *Case-Based Reasoning with Confidence*. Book Advances in Case-Based Reasoning, 2000.

[4] J. Chua and P. Tischer. Determining the trustworthiness of a case-based reasoning solution. *In International Conference on Computational Intelligence for Modeling, Control and Automation:7th European Conference on Case-Based Reasoning (ECCBR 2004)*, 3155(1740881885):12–14, 2004.

[5] S. Delany, P. Cunningham, and D. Doyle. Generating of classification confidence for a case-based spam filter. In *In 6th International Conference on Case-Based Reasoning (ICCBR 2005).Springer*, volume 3620, pages 177–190, 2005.

[6] P. Ekman and W. V. Friesen. Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press, Palo Alto*, 1978.

[7] M. Grachten, F.A. Garcia, and J.L. Arcos. Navigating through case base competence. In *In eds.:7th European Conference on Case-Based Reasoning (ECCBR 2004), Springer*, volume 3155, pages 282–295, 2004.

[8] J. Orozco and F.A. García and J.L. Arcos and J. Gonzàlez. patio-Temporal Reasoning for Reliable Facial Expression Interpretation. In *Proceedings of the 5th International Conference on Computer Vision Systems (ICVS'2007)*, Bielefeld, Germany, March 2007.

[9] T. Reinartz, I. Iglezakis, and T. Roth-Berghofer. Computational intelligence. *Artificial Intelligence Communications*, 2001(17):214–234, 2001.

[10] B. Smyth and E. McKenna. Competence models and the maintenance problem. *Computational Intelligence*, 17(2), 2001.