Genetic Algorithm and VR for Assessing the Level of Expertise of Maintenance Operator

Axel Foltyn[1(✉)], Christophe Guillet[2], Florence Danglade[1], and Frédéric Merienne[1]

[1] Arts et Métiers Institute of Technology, LISPEN, 71100 Chalon-Sur-Saône, France
axel.foltyn@ensam.eu

[2] Université de Bourgogne, LISPEN, 71100 Chalon-Sur-Saône, France

Abstract. The study aims to find the features for assessing the level of a maintenance operator. A genetic algorithm is used to identify the most relevant features and reduce their size. Based on 30 different features entered, we demonstrate that only three operator-level evaluation features provide a good classification. Virtual reality was used to simulate maintenance operations, collect data, and validate our method for identifying the most relevant features.

Keywords: Feature selection · Skill Assessment · VR

1 Introduction

In some industries, know-how is passed on between an expert and a novice operator [2], particularly for complex manual operations such as maintenance. However, this approach faces challenges such as training duration and experts' availability. To overcome these difficulties, one solution could be a digital assistant tailored to operators' level of expertise. This personalized assistance would offer comprehensive support to beginners and clear, succinct information to experts.

A maintenance operation can be decomposed into a sequence of movements. Each movement can be characterized by a set of features.

Although some studies have identified features for assessing operator expertise in other contexts [5,7,9,10,12], the large number of features poses a considerable challenge. This challenge often translates into less accurate classification due to the phenomenon of the curse of dimensionality [6] or slower evaluation due to the calculation of irrelevant features. With this in mind, our aim is to develop an expertise-level evaluation system using cluster algorithms [3].

Features extraction algorithms such as Principal Component Analysis (PCA) are common methods. However, they are only projections. Since we also want to reduce computation time, we plan to use a combination of motion evaluation features proposed in the literature and feature selection techniques. Some feature selection techniques are fast but may lead to a local but not global optimal solution, while others provide optimal solutions but with longer times [11].

L. T. De Paolis et al. (Eds.): XR Salento 2024, LNCS 15027, pp. 367–375, 2024.
https://doi.org/10.1007/978-3-031-71707-9_31

The fact that a large number of features can be selected to evaluate motion raises the following research question: What are the most relevant features for evaluating the behavior of a maintenance operator by using VR ?

2 Method

The proposed method (Fig. 1) aims to find the relevant features for assessing a maintenance operator's level of expertise.

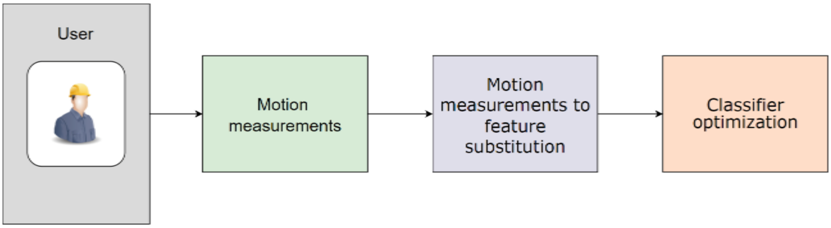

Fig. 1. Proposed pipeline for extracting relevant features.

To achieve this goal, several steps are proposed. First, the motion measurements of each user are collected. From these motion measurements, features are defined to evaluate the operator's level of expertise. Then, we substitute the motion measurements in values taken by these features.

The final step consists of building a classification model from a set of optimized features. This optimization is implemented via a dimension reduction performed by our feature selection algorithm. Motion measurements use tracking technology to capture head and hands positions and rotations at a fixed frequency. Positions are expressed in cartesian coordinates, while rotations are expressed in quaternion coordinates [4]. The substitution step makes it possible to obtain features qualifying the operator's level of expertise from the motion measurements. Five features (Table 1) have been chosen from the state of the art [5,7,9,10,12]. To avoid outliers and simplify calculations for the classifier, each feature is averaged. This average is calculated for each feature over the duration of the movement. Table 1 shows the chosen features. Classification into different expertise-level groups will be achieved by separation using K-Means and SVM algorithms. Given the risk of encountering a non-linearly separable problem, placing the data associated with each subject in a feature space of different dimension is necessary. To achieve this, we apply a kernel trick [1] of the form $f(x_1, x_2, ..., x_N) \Rightarrow (x_1, x_1^2, ...x_N, x_N^2)$ with N=5 (the number of features). This kernel trick doubles the number of features. In the end, with three sensors (the two hands and the head), the total number of features equals 30. The dimension reduction method is based on a main loop similar to that of the SBS (sequential backward selection) feature selection method [8] (Fig. 2), which tests each combination obtained by deleting a feature from the set of features and retains the one that achieves the best result. This loop is repeated until the desired dimension is

Table 1. Table containing formulas for each feature used.

feature name	Formula and Related term
Control (C)	$C = \frac{1}{N}\sum_{i=1}^{N}\frac{v_{max,i}}{v_{min,i}}$
Stability (S)	$S = \frac{1}{N}\sum_{i=1}^{N}\sigma_i$
Motion frequency (F)	Number of movements per minute on average
Velocity (v)	$v = \frac{1}{N*M}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{dp_i}{dt}(t_{i,j})$
Acceleration (a)	$a = \frac{1}{N*M}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{d^2p_i}{dt^2}(t_{i,j})$ with $\{t_{i,j}\}_{1\leq i\leq N,1\leq j\leq M}$: the set of recording instants of the N*M positions over the duration $\{W_i=[0;t_{i,M}]\}_{1<i<N}$: The set of time windows over which we calculate the mean $\overline{\left(\frac{d^4p_i}{dt^4}\right)^{-1}}$ σ_i : The standard deviation of measurements $\left\{\left(\frac{d^4p_i}{dt^4}(t_{i,j})\right)^{-1}\right\}_{1\leq j\leq M}$ on window $W_i=[0;t_{i,M}]$ $\sigma_i = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(\left(\frac{d^4p_i}{dt^4}(t_{i,j})\right)^{-1}-\overline{\left(\frac{d^4p_i}{dt^4}\right)^{-1}}\right)^2}$ v_{max} : The maximum velocity value on window $W_i=[0;t_{i,M}]$ v_{min} : The minimum velocity value on window $W_i=[0;t_{i,M}]$ p : Cartesian coordinate position t : Time in seconds

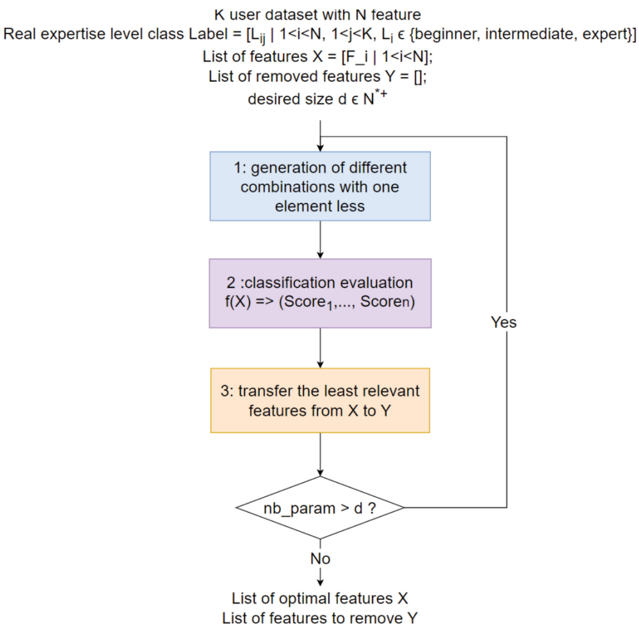

Fig. 2. SBS steps.

reached. However, in our approach, the step 1 uses of a genetic algorithm that is prematurely stopped (Fig. 3). This makes it possible to observe the directions in feature space towards which solutions are heading and to find the least relevant feature. Genetic algorithms are programming tools that simulate the process of natural selection. Their aim is to find an optimal solution to a complex problem. This search for a good solution is associated with a fitness function measuring the quality of a solution expressed as a number or vector.

The steps in this algorithm are illustrated in Fig. 3. We initialize our lists of weighting weights randomly (1.1). This step is commonly referred to as genesis. Then, our data are weighted by N weights $(w_{1,i}, w_{2,i}..., w_{N,i})$ (1.2). Our model ranks each data set weighted by our weights (1.3). An evaluation is performed to find the weight lists with the best scores (1.4). These lists are then modified (1.5). The algorithm is rerun from step 1.2 until the set number of epochs is reached. For the classification model (1.3), we decided to use K-means. This choice stems from the fact that the algorithm is popular for its simplicity and is used in other feature selection studies. This will enable us to make a future comparison with these algorithms. K-Means is an unsupervised clustering algorithm that divides a data set into K clusters, with K defined in advance, minimizing intra-cluster variance. For evaluation (1.4), we use a fitness function with three criteria (misclassification, number of features, and novelty). The first two criteria should be minimized. For misclassification evaluation (see Fig. 5), the process begins with the creation of k groups (for example k=3 in Fig. 4). In each group, minority labels are identified and counted as errors. For example in the Fig. 4, three misclassifications are identified, which are the circled dots. We impose a penalty on the number of features, which is a value used to induce a reduction in the number of features. To do this, we calculate the sum of the weights associated with each feature and increment each weight by one when it exceeds a minimum threshold (set at 0 in our case). This penalizes solutions that use more features. Novelty is calculated with a maximized score called novelty search. Novelty search is a computing method that promotes the discovery of new solutions by encouraging diversity rather than focusing on a specific goal.

3 Experimentation

The experiment consisted in assessing the operator's level for an activity involving the assembly (Fig. 3) of part of a landing gear. To test our model, we set up an experiment using virtual reality (Fig. 7). This allows us to put subjects in an ecological situation at a lower cost and with precise motion measurements. The assembly activity was performed in virtual reality (VR) using the HTC Vive Pro headset to recover sub-millimeter-precision hands positions and rotations using the joysticks, and head positions and rotations using the headset. Hands and head positions and rotations were recorded at a frequency of 60 Hz. Assembly was simulated 83 times at three different skill levels (beginner, intermediate and expert). When the subject completes the task, the system saves all recorded observations as a CSV file. These data can be used as input for the feature selection algorithm presented in the method section.

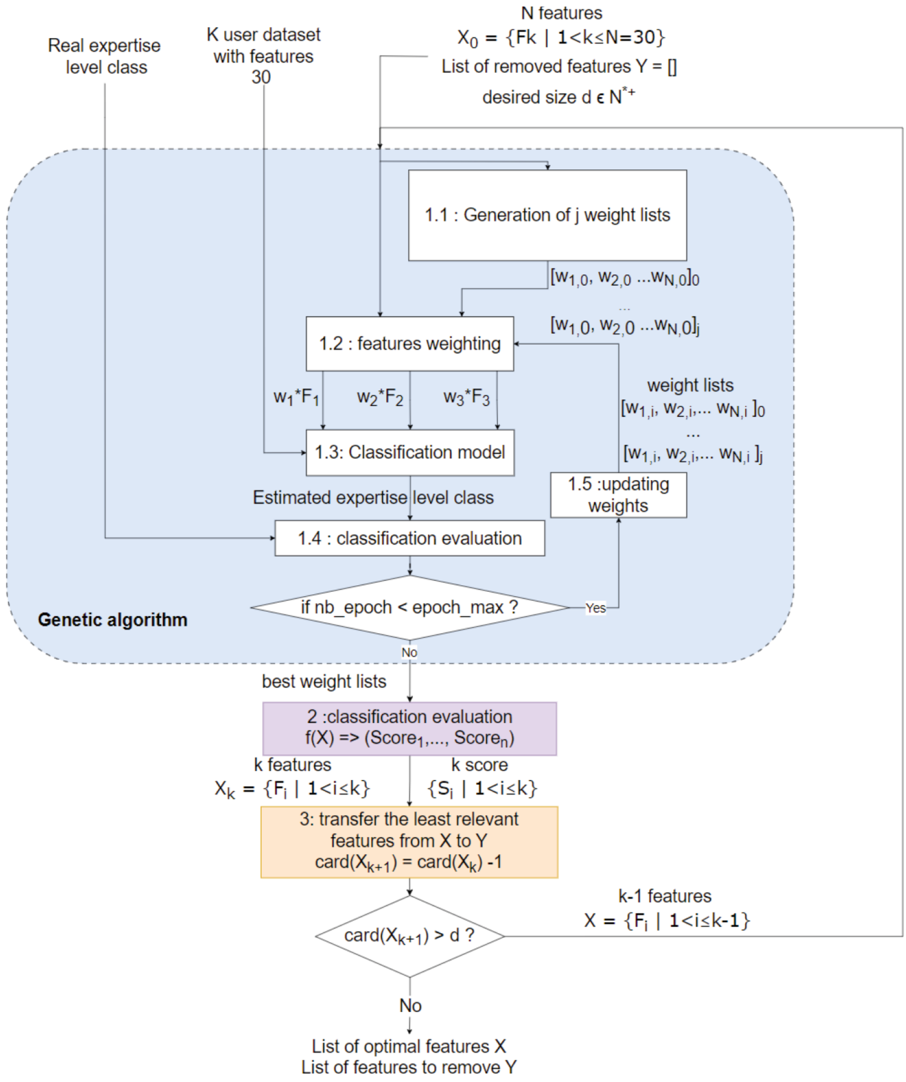

Fig. 3. Diagram of our prosed method for feature selection.

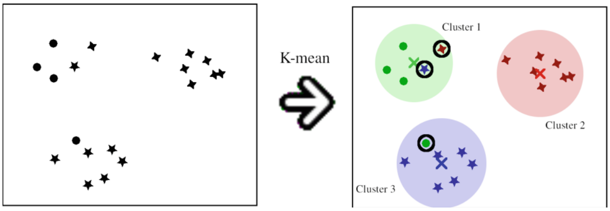

Fig. 4. Visual example of K-Means grouping, with crosses representing the centers of groups named centroids.

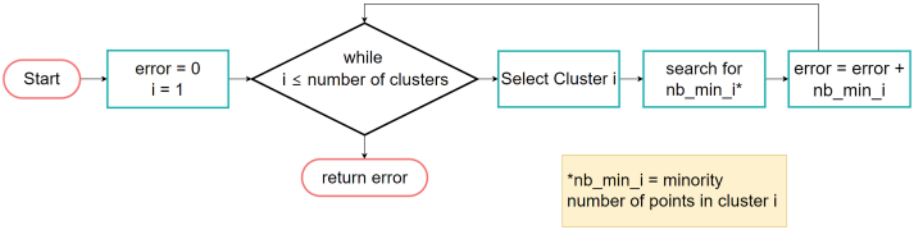

Fig. 5. Schematic diagram of classification misclassification evaluation steps.

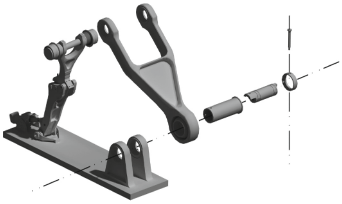

Fig. 6. Exploded view of assembly.

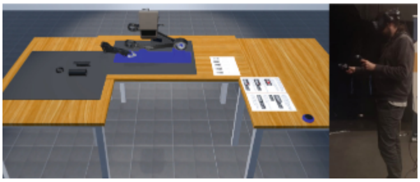

Fig. 7. Image of a virtual environment with the subject doing maintenance operation.

4 Results and Discussion

On the 83 simulation subjects, split 80% for the learning phase and 20% for the test phase, all with 5-fold cross-validation increasing the robustness of our algorithm reduced the number of features by one at each epoch. Figure 8 shows the average errors for each epoch.

It can be seen that the average error varies little when the number of features is greater than ten and decreases after that. In fact, if the number of features is too large, classification is more complicated. There are two possible reasons for this: 1) the ratio between the number of samples and the number of features is too low, 2) the dimension of the feature space is too high (curse of dimensions [6]).

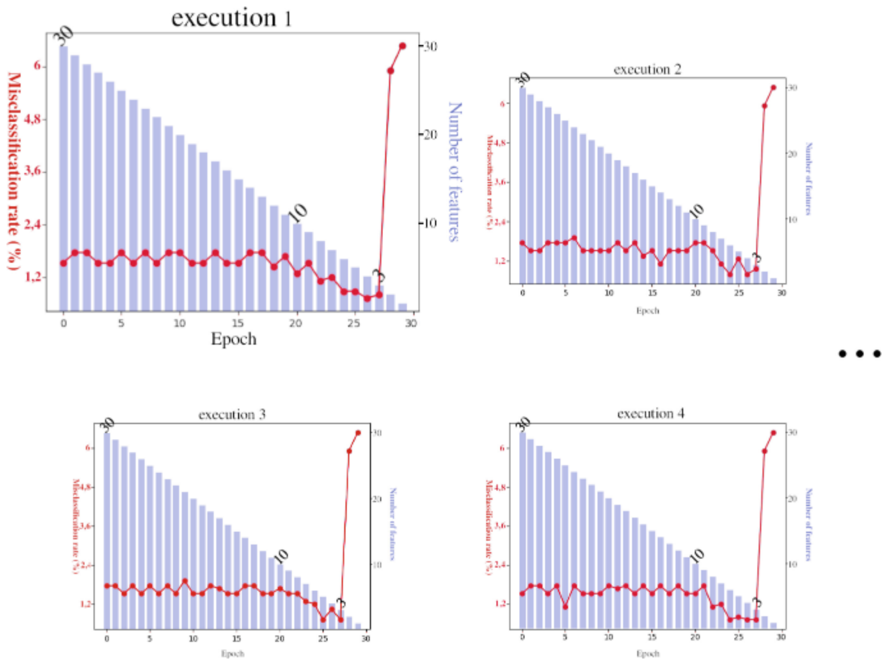

Fig. 8. Evolution of the misclassification rate for each epoch.

The relevant feature selection criteria are, on the one hand, the choice of a threshold above which any value of the mean classification error is rejected and, on the other hand, the repeatability rate of the result over ten runs of our algorithm. In our case, the chosen threshold is 2.4, and the repeatability rate is 100%. The epoch at which these criteria are met is epoch 27. This epoch has retained three features (Fig. 8) which are the squares of the head and hands motion frequencies.

Table 2. Classification error rates.

	case with all features	case with relevant features
Number of features	30	3
Misclassification rate	0,92%	0,72%

Table 2 shows classification error rates of 0.92% with all features and 0.72% with the three relevant features. These rates are almost identical whether we take the set of features or the three optimal evaluation features. However, in the latter case, the number of features to be calculated has been reduced by a factor of 10. This proves that our feature selection method is robust and is linked to the possibility of returning to the same score, regardless of the number of trials.

Moreover, with the simplified study framework we have chosen, we have shown that the most relevant features are the square of the head and hands motion frequencies.

5 Conclusion

For the study, motion measurements were collected for each subject in a virtual reality application simulating an industrial maintenance task. Based on these motion measurements, 30 features were defined to assess the operator's level of expertise. Then, the motion measurements are substituted by the values of these evaluation features in our study. Based on these evaluation features, we proposed a new feature selection algorithm, a hybrid between the SBS and genetic algorithms. We could show that three features were sufficient to assess a maintenance operator's expertise level. However, the simplified framework of our study, due to the choice of averaging all the features, the absence of any study of the relevance of the size of the time window chosen, and the fact that the data collected comes from simulation may induce various biases. Regardless of the bias induced by our study framework, we have shown that our algorithm is robust. Having verified the robustness of our method in a simplified case, it would be interesting to test it in a real case. Moreover, the choice of motion measurements was limited by our XR system. In the future, it would be interesting to measure new movements such as gaze, fingers and back using tools such as varjo and leap motion. In conclusion, the results obtained validate the algorithm based on two of the three existing families of feature selection algorithms: filter and wrapper selection. It would therefore be interesting to test embedded selection, which is the last of these families.

Acknowledgements. This work was supported by French government funding managed by the National Research Agency in the framework of the project VIMACO - ANR-21-CE10-0009. This work was supported by French government funding managed by the National Research Agency under the Investments for the Future program (PIA) grant ANR-21- ESRE-0030 (CONTINUUM).

References

1. Kernels and Feature maps: Theory and intuition—Data Blog. https://xavierbourretsicotte.github.io/Kernel_feature_map.html
2. Bellu, S.L., Lahlou, S., Nosulenko, V., Samoylenko, E.: Studying activity in manual work: a framework for analysis and training. Le travail humain **79**(1), 7–30 (2016)
3. Cahapin, E.L., Malabag, B.A., Jr, C.S.S., Reyes, J.L., Legaspi, G.S., Adrales, K.L.: Clustering of students admission data using k-means, hierarchical, and DBSCAN algorithms. Bull. Electr. Eng. Inf. **12**(6), 3647–3656 (2023). https://doi.org/10.11591/eei.v12i6.4849
4. Dam, E.B., Koch, M., Lillholm, M.: Quaternions, Interpolation and Animation (1998)

5. Hammond, T., et al.: It's not just about accuracy: metrics that matter when modeling expert sketching ability. ACM Trans. Interact. Intell. Syst. **8**(3), 19:1–19:47 (2018). https://doi.org/10.1145/3181673
6. Hastie, T., Friedman, J., Tibshirani, R.: The Elements of Statistical Learning. Springer Series in Statistics, Springer, New York (2001). https://doi.org/10.1007/978-0-387-21606-5
7. Kiang, C.T., Yoong, C.K., Spowage, A.C.: Local sensor system for badminton smash analysis. In: 2009 IEEE Instrumentation and Measurement Technology Conference, pp. 883–888 (2009). https://doi.org/10.1109/IMTC.2009.5168575
8. Kumar, V.: Feature selection: a literature review. Smart Comput. Rev. **4**(3) (2014). https://doi.org/10.6029/smartcr.2014.03.007
9. Ladha, C., Hammerla, N., Olivier, P., Ploetz, T.: ClimbAX: skill assessment for climbing enthusiasts (2013). https://doi.org/10.1145/2493432.2493492
10. Nordin, N., Xie, S.Q., Wünsche, B.: Assessment of movement quality in robot-assisted upper limb rehabilitation after stroke: a review. J. Neuroeng. Rehabil. **11**(1), 137 (2014). https://doi.org/10.1186/1743-0003-11-137
11. Venkatesh, B., Anuradha, J.: A review of feature selection and its methods. Cybern. Inf. Technol. **19**(1), 3–26 (2019). https://doi.org/10.2478/cait-2019-0001
12. Zollo, L., Rossini, L., Bravi, M., Magrone, G., Sterzi, S., Guglielmelli, E.: Quantitative evaluation of upper-limb motor control in robot-aided rehabilitation. Med. Biol. Eng. Comput. **49**(10), 1131–1144 (2011). https://doi.org/10.1007/s11517-011-0808-1