Feature Selection for Complex Systems Monitoring: an Application using Data Fusion

Gwénolé Le Moal, George Moraru, Philippe Véron
Arts et Métiers ParisTech, CNRS, Laboratoire LSIS
Aix-en-Provence, France
gwenole.lemoal@ensam.fr
george.moraru@ensam.fr
philippe.veron@ensam.Fr

Patrice Rabaté, Marc Douilly
EADS Innovation Works
Nantes/Méaulte, France
patrice.rabate@eads.net
marc.douilly@eads.net

Abstract – Emergence of automated and flexible production means leads to the need of robust monitoring systems. Such systems are aimed at the estimation of the production process state by deriving it as a function of critical variables, called features, that characterize the process condition. The problem of feature selection, which consists, given an original set of features, in finding a subset such the estimation accuracy of the monitoring system is the highest possible, is therefore of major importance for sensor-based monitoring applications. Considering real-world applications, feature selection can be tricky due to imperfection on available data collections: depending on the data acquisition conditions and the monitored process operating conditions, they can be heterogeneous, incomplete, imprecise, contradictory, or erroneous. Classical feature selection techniques lack of solutions to deal with uncertain data coming from different collections. Data fusion provides solutions to process these data collections altogether in order to achieve coherent feature selection, even in difficult cases involving imperfect data. In this work, condition monitoring of the tool in industrial drilling systems will serve as a basis to demonstrate how data fusion techniques can be used to perform feature selection in such difficult cases.

Keywords: *Sensor-based monitoring; feature selection; data fusion; uncertainty representation; evidence theory; drill condition monitoring*

I. Introduction

A. *Feature selection for complex systems monitoring*

Manufacturing industry increasing needs of quality, productivity and flexibility coupled with costs reduction objectives made monitoring of complex production processes a subject of major importance. Monitoring systems have to perform considering variable operating conditions, harsh environments and complex decision making situations. These constraints make the design of robust monitoring systems a challenging task.

An important step of the design of a monitoring system is feature selection. As estimations of the process state are based upon some features of interest, a good selection of these features is essential. A good feature set will improve the prediction performance, provide faster and more cost-effective estimations and a better understanding of the underlying process [1], whereas the use of irrelevant features will lead to downgraded estimation performances and increase computation time. A reduced feature set is often wished as it implies the use of fewer sensors and because it reduces risks of estimation performance degradation due to the curse of dimensionality.

When no clues are available about the relevance of features at the designing step of a monitoring system, feature selection is usually performed following a basic feature selection process (BFSP), as depicted in Fig. 1. It consists in designing experiments dedicated to emphasize the phenomena to be monitored, and to collect raw data from sensors. Sensors are the same that will be used by the monitoring system. Large scale feature extraction procedures are then performed on the acquired data, being as exhaustive as possible in order not to miss useful features. Finally, a feature relevance characterization procedure is applied in order to obtain a relevant feature set regarding the phenomena of interest. This BFSP procedure has been applied in [2] and [4] for instance.

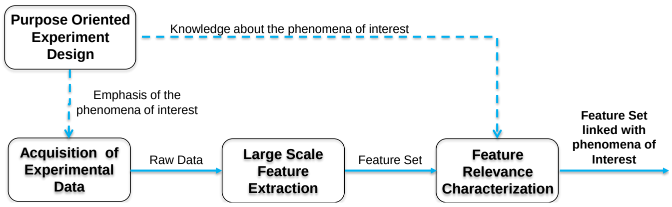

Figure 1. Basic feature selection process (BFSP)

When monitoring complex processes, this BFSP presents some limitations because of its sensibility to experimental conditions like process parameters or influence quantities that affect sensor measurements. Moreover, it is not suitable for dispersive systems showing significantly different behaviors when only slight changes, sometimes barely detectable, occurred in the operating conditions. In [2] for example, only one operating condition of the machining system has been investigated so findings about selected features should not be generalized, and in [8] testing the monitoring system in variable operating conditions in order to assess the proposed feature selection methodology sensibility and effectiveness will be part of future work. This lack of robustness of the method when quasi-exhaustive data about the monitored system are not available, which is likely to occur when studying flexible processes working in harsh environments, is an important issue.

Tracks to overcome these drawbacks and select features that allow improving the robustness and the flexibility of monitoring systems exist. The first one consists in considering that available data are not perfect but tainted with uncertainty coming from the monitored process dispersive behavior, and/or acquisition problems like noise, sensors failure, sensor mounting issues, harmful influence quantities… Keeping in mind that collected data may not exactly represent the process behavior should prevent misleading interpretations. Moreover, the presence of multiple data collections concerning the process, even if heterogeneous, incomplete or contradictory, should be considered as a chance for robustness improvement, and not as a drawback in order to obtain "easy to interpret" results. The processing of these different data collections in order to obtain a single feature set is then an important issue, and data fusion becomes necessary to incorporate all the data into the analysis [3].

B. Description of the drill condition monitoring problem

Precision drilling is considered a complex process and its monitoring is a challenging task [11]. Monitoring the drill bit condition is of major importance in many fields of manufacturing industry where high quality and high productivity are required, like in aeronautical industry for instance [12]. Numerous studies have been carried where tool condition monitoring systems have been designed and implemented successfully. However, precision manufacturing industry suffers a lack of flexible and widely deployed drill condition monitoring systems. This is due to lack of robustness of systems once they are implemented in the shop floor, because they have been developed in friendly lab environments and under steady operating conditions [11].

In this work, monitoring of drill bit cutting edge micro-chipping in industrial environment has been investigated. 2 drill bit cutting edge states have been considered: chipped or good shape. Drill bit cutting edge chippings are shown in Fig. 2. A clustering algorithm has been used as an estimator: it was aimed at finding two clusters within a dedicated feature space. If two clusters could be clearly identified, presence of the 2 drill bit possible states was stated. As drills were in good shape state at the beginning of their use, it implied that a chipping occurred. The presence of 2 clusters has been assessed using classical clusters separation measures on the results provided by the clustering algorithm on the available data. A clustering method has been preferred to classification approaches because it allows having flexibility regarding operating condition as no supervised learning is performed a priori.

Performance of chipping identification has been assessed for different feature spaces issued from different feature selection methods in order to compare them. Feature relevance characterization approaches have already been applied to tool condition monitoring (TCM) in milling and turning [2,9], showing promising results for tool wear monitoring. In this work, the complete feature selection process, from the data acquisition step to the obtaining of a dedicated feature set, has been investigated.

Test campaigns have been conducted in industrial conditions on 3 different drilling machines, a robot and 2 machine tools. Therefore, data collections were of limited size, not exactly the same sensors were used during each test campaign, and operating conditions were not always identical. Due to the harsh acquisition conditions, some measurements were affected by high noise levels and sometimes sensors failures occurred. Approximately 10 sensors of 5 different types have been used for each campaign, and between 350 and 500 features have been extracted from the raw data, depending on the number of mounted sensors during the test campaign.

Although the available data collections can be considered difficult to deal with due to their heterogeneity, they are representative of industrial conditions where a tool condition monitoring system is needed, and also justify the incorporation of imperfection in information modeling.

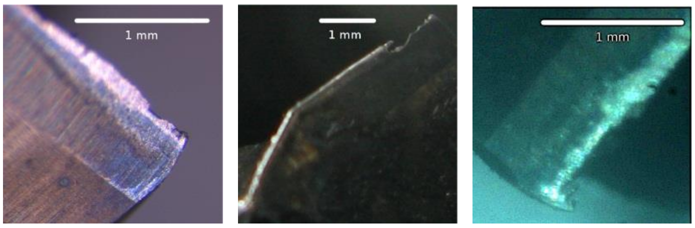

Figure 2. Examples of drill cutting edge micro-chippings

C. Paper organization

The feature relevance characterization step of the BFSP will first be discussed, leading to the choice of an established feature weighting algorithm. In a second time, data fusion based approaches aimed at analyzing the contents of every data collection altogether will be presented, and the different solutions it offers in terms of data imperfection representation will be detailed. Then, some results will be discussed, emphasizing the influence of data fusion and data imperfection representation on the monitoring performance.

II. FEATURE SELECTION

A. A brief introduction to feature relevance characterization

The feature relevance characterization step of the BFSP has been widely studied, and a good introduction and useful references can be found in [1,5,6,7]. The feature relevance characterization procedure can take 3 main forms: filters, wrappers and embedded [1,7]. Within the first one, features are selected or weighted (selection/rejection can be considered as special cases of weighting using binary weight values [7]) according to their relevance regarding the phenomena of interest. The weighting is performed independently of the estimator that will be used to perform monitoring. Feature weights are assigned following some relevance criterion, usually correlation with the phenomena of interest. Filters are often used because of their simplicity, scalability and good empirical success [1], but they present limitations: one common criticism of feature weighting is that it leads to the selection of a redundant subset of features, and the same or better estimation performance could be achieved with a smaller subset of complementary variables. Moreover, dependencies between features are not taken into account, although variables that are useless by themselves can sometimes be useful together [1]. Filtering can also be performed before applying any other feature relevance characterization procedure dedicated to reduce the dimensionality of the original feature space.

Wrapper approaches [5] use the estimator as a mean to evaluate relevance of features during the feature relevance characterization process. This form of feature relevance characterization guarantees coherence between the selected features and the estimator used by the monitoring system. It also allows taking into account the influence of the use of several features simultaneously. However, wrappers can be time consuming due to the multiple evaluations of the estimation algorithm, and when facing a large number of features, all feature combinations cannot be evaluated, requiring the use of heuristic search methods within the original feature power set.

Lastly, embedded methods integrate the feature relevance characterization process as a part of the estimator induction step. This form of feature relevance characterization implies the use of particular estimators and is not suitable when a clustering algorithm is used, as it is the case in this study.

Hybrid methods also exist that usually use filtering procedures as a preprocessing step in order to reduce the feature space dimensionality before using a wrapper or embedded approach. Examples of hybrid strategies implemented to face high number of features can be found in [3] and [10].

A filter method will be used to perform feature weighting in this work because it can be applied before using any estimation algorithm, or as a pre-processing step before implementing a wrapper or embedded feature relevance characterization method. Moreover, sensor-based monitoring of industrial production means often requires the use of a reduced feature set. The choice of a feature weighting algorithm is discussed in the next section.

B. Choice of a feature weighting algorithm

Several variations have been developed following the simplest filtering scheme. The FOCUS algorithm [13] involves a greater degree of search among the feature space as it begins by looking at each feature in isolation, then pairs, triplets, and so on, halting only when good enough performance is achieved. As it addresses the problem of feature interaction and feature redundancy, it has shown good robustness facing the presence of irrelevant features, but its search procedure is likely to become intractable as a function of the number of features in the original feature set to be analyzed.

The RELIEF algorithm [14] uses a statistical based feature evaluation function: it collects all statistically relevant features by assigning, one instance at a time, high weights to features showing strong separation power between closest instances of different classes and keeping closest instances belonging to the same class close in the feature space. The final ranking is obtained by averaging those weights assessed for a statistically relevant number of instances. This algorithm combines several advantages compared to other feature weighting techniques: first, it handles the problem of features interaction by working within the whole feature space, so no time consuming exploration of the feature space is needed. Then, it allows obtaining good results even when working with noisy data and/or in feature space containing a lot of irrelevant features [7,14], which is particularly interesting in our application case. However, RELIEF does not help with redundant features.

Due to its simplicity and empirical success, the RELIEF algorithm has been widely used and extended to feature weighting for multiclass classification/clustering and regression applications [15,16]. An interesting analysis was performed in [17] that allowed identifying two weaknesses of the algorithm: relevance evaluation is performed in the original feature space, but can be significantly different in the resulting weighted feature space. Then, as features weights are averages of their separation power over instances classes membership, the presence of outliers in the data set can lead to severe misleads. In our application case where imperfect data exist, these drawbacks can significantly affect the feature relevance characterization results. Solutions have been proposed to address them: first, instead of using only the closest instances to assess the separation power of a feature, influences of several neighbors is taken into account via the use of a kernel function, which allows reducing harmful influence of outliers. Then, the last weighted feature space that has been calculated is used to evaluate features separation power at each iteration, leading, under easy-to-achieve conditions, to the convergence of the algorithm to an optimal weighted feature space. This last property gave its name to the new algorithm: IRELIEF for Iterative RELIEF. It has shown superior performance than the RELIEF algorithm in most cases [17].

Because of its clear theoretical foundations, good empirical success and robustness facing some data imperfections, the IRELIEF algorithm has been chosen to perform the feature relevance characterization step in this work.

III. DATA FUSION FOR FEATURE SELECTION

A. Basic fusion approach and issues

Considering the existence of several data collections, data fusion is needed to perform feature selection taking all available information into account. The basic idea is to perform the BFSP using IRELIEF on every available data collection, and then merge the obtained weighted feature sets to obtain a generic weighted feature set, as depicted in Fig. 3.

As the feature relevance characterization step is identical for every available data collection, weighted features sets are given in the same formalism to the fusion step. This allows avoiding a tricky task in data fusion which consists in a conversion of the data into a common coordinate frame before being merged. This task is called data alignment and is considered a difficult problem for which a general theory does not exist [18]. For instance, feature weights issued of each BFSP can be averaged in order to obtain the generic weighted feature set.

However, data collections do not always suit to such direct fusion. First, they may not allow extracting exactly the same features. In that case, several strategies can be implemented depending on the fusion algorithm used, following the principle that a feature which is not present in every data collection should not be disadvantaged. Using feature weights for fusion without any preprocessing can also lead to erroneous conclusions because of the different scaling given by the feature relevance characterization algorithm as a function of the data collection specificities. Finally, all data collections may not be considered equally informative: for instance, if a large test campaign has been conducted in good conditions, using high quality sensors with high sampling rates and no problem occurred, it should be considered more informative than another one that has given little amount of noisy data, where acquisition problems occurred, and conducted in conditions that are different of the ones that are likely to be encountered by the monitoring system.

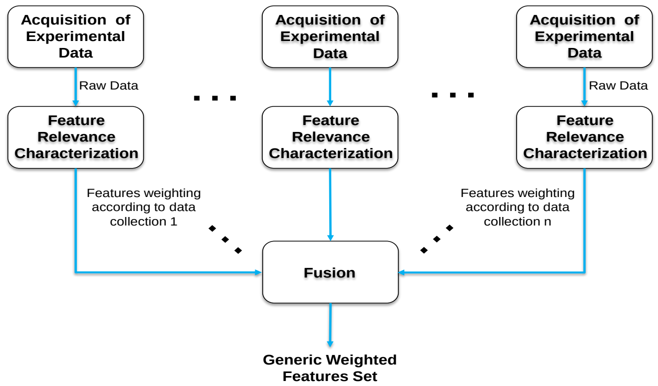

Figure 3. Global feature selection process using data fusion

In order to address these issues, it is possible to incorporate a data alignment step to the BFSP before the fusion process occurs, as depicted in figure 4. The data alignment step has several objectives:

- to give weights to features that are independent of the scaling resulting from the feature relevance characterization step
- to perform a coherent feature weighing regarding the fusion method used at the next step
- to let the *monitoring system designer* incorporate meta-knowledge concerning the quality of information provided by a data collection, if any

If the first objective is straightforward, the following ones are to be considered with particular attention. The last one raises the question of the representation of data imperfection because of its close link with the quality of information provided by data collections.

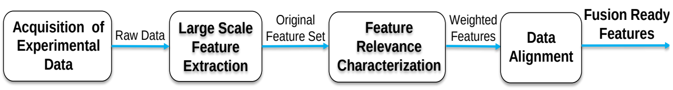

Figure 4. Basic feature selection process incorporating a data alignment step

B. *Uncertainty sources and representations within the global feature selection process*

The IRELIEF algorithm allows handling low level stochastic uncertainty on features values and attenuating consequences of the presence of outliers in the data. Still, epistemic uncertainty due to lack of knowledge, which comes from the inability for a BFSP to draw conclusions about some features relevance, cannot be modeled at this sub-level of the global feature selection process.

Origins of epistemic uncertainty in sensor-based monitoring are various. First of all, sensor failures lead to lack of knowledge, as no information is available to draw conclusions. Then, everything that affects sensors detection ability, like mounting issues or too low sample rates increases the level of epistemic uncertainty. Moreover, uncertainty does not always fall precisely into either stochastic or epistemic uncertainty [19]: when so much stochastic perturbations, or noise, affect measurements making statements based upon acquired data difficult, a lack of detection ability is considered, so stochastic perturbations induce epistemic uncertainty. Lack of information of a BFSP should be compensated at the fusion step by more accurate statements issued from the processing of other data collections, taking advantage of the data collections redundancy and complementarity. Therefore, the choice of a good information modeling scheme allowing a precise representation of data imperfection is critical to achieve good fusion performance.

Evidence theory has been shown to be a practical framework to handle both stochastic and epistemic uncertainty and suits information fusion contexts. It has been used in this work at both the data alignment and fusion steps, following previous works of the authors aimed at singularity detection in difficult contexts using multiple information sources [20,21]. In this work, the focused singularity is feature relevance and the information sources are the data collections. This approach proposes a data alignment procedure which has been designed to favor most informative information sources statements at the fusion step. The modeling of information coming from each BFSP at the data alignment step allows representing epistemic uncertainty explicitly by the use of belief functions. Information modeling and fusion have been done within the Transferable Belief Model [22], that followed original works of Dempster [23] and Shafer [24] on evidence theory. It has been shown to be superior to classical probabilistic approach in multi-source difficult contexts when following informative sources was important [20,21]. The quality of information issued from a BFSP is a parameter of the data alignment step that influences transition from certain statements to uncertain ones, then transferring more influence to other sources at the fusion step. In this study, this parameter has been set-up manually as a function of the quality of the test campaign that generated each data collection in terms of operating conditions, acquisition parameters, and the number of available instances for each class.

C. *Information fusion algortihms*

Issues evocated in the previous section emphasized the importance of the fusion algorithm. 4 different *{information modeling, fusion algorithm}* couples have been implemented. For the 3 first ones, very simple data alignment strategies and fusion algorithm were used. The features weights issued from the BFSPs have been respectively: conserved, squared, and cubed; and the weights issued from the so obtained fusion ready weighted feature sets (FRWFSs) have then been averaged at the fusion step. Power elevation of the feature weights is aimed at emphasizing the difference between highly relevant features and others, and doing so advantaging the most informative features as a function of the power elevation. No epistemic uncertainty has been modeled explicitly.

As explained in the previous section, the last couple has been designed in the evidential framework. The fusion has been done according to the classical conjunctive combination rule [24], giving a list of feature ranked according to their relevance. The evidential method presents a severe drawback that is encountered in applications of evidence theory: the complexity increases exponentially as a function of the number of features because calculations are done for every element of the features power set. Thus, in order to reduce computation time, a criterion has been set-up after observation of the IRELIEF results to eliminate irrelevant features, and a redundancy filter has also been set-up as redundant features decreases prediction performance.

IV. RESULTS AND DISCUSSION

A. Experiments design

For each of the 3 test campaigns that have been used, the BFSP has been done using the IRELIEF algorithm. Then, only features presenting an interest and that were not redundant have been selected. It represented a maximum of 12 features over the 500+ that have been extracted. Then, the 4 data alignment and fusion algorithms couples have been applied, each providing a ranked list of features. As the objective of the monitoring system is to perform the best clustering performance using the smallest feature set, clustering has been performed using an increasing number of feature according to their ranking. The ability of each method to identify good features has been assessed using the clustering error rate computed over results of a Monte-Carlo simulation, in order to overcome bias involved by the random clusters initialization. Respective influences of the {information modeling, fusion algorithm} couples and of the data collection number has been assessed with 3 data collections: 2 are issued from drilling test campaigns realized on CNC machines, with quite similar operating conditions but different acquisition set-up, and the 3rd one has been conducted on a robot, with very different operating conditions, quite different sensors and containing a lower amount data, so it has been considered less informative.

B. Some results and discussion

The global feature selection process, as depicted in Fig. 3, has first been applied on the two similar data collections, and the clustering performance achieved with the so obtained generic weighted feature set on each data collection has been computed. Results obtained for the 1st and 2nd data collections are visible in Fig. 5 and 6 respectively. Concerning the 1st test campaign, all approaches gave comparable results except the cubic one. One can notice that clustering performance decreased as the number of feature increased. This is due to the fact that new features did not provide useful information while increasing the feature space dimensionality and noise. Better results were obtained with the 2nd data collection: lower error rates have been achieved, even with low numbers of features. This is probably due to the apparition of a more impacting tool cutting edge chipping. Particularly, the proposed data alignment and fusion approach outperformed the 3 classical ones when using low numbers of features, showing a good ability for efficient feature ranking.

In a second time, the 3rd data collection, which differs significantly from the two others with respect to many points, has been used with the 2 other ones in the feature selection process. Results are depicted in Figs. 7, 8 and 9. If results were similar for the 2nd data collection, they were better for the 1st one: the BFSP on the 3rd data collection has selected features that are useful but that were not selected when using only 2 data collections. Results were comparable using either the proposed approach or classical ones as nearly all features are involved to achieve them. One can remark that the proposed approach performance increases since 8 features are involved instead of 9 when using classical approaches. Considering the 3rd data collection, good cutting edge chipping identification has been achieved, and advantage was also given to the proposed approach confirming performance improvements can be achieved by taking data imperfection into account in the feature selection process.

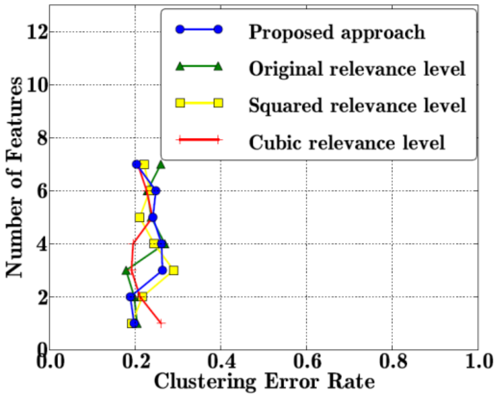

Figure 5. Clustering results obtained for the first data collection when merging 2 data collections to perform feature selection

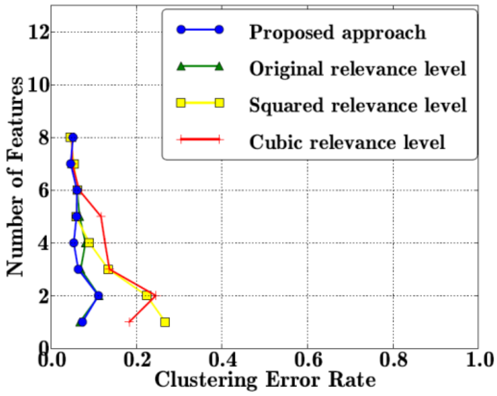

Figure 6. Clustering results obtained for the second data collection collection when merging 2 data collections to perform feature selection

V. CONCLUSION AND FURTHER WORKS

In this work, an attempt was made to tackle the problem of feature selection for complex systems monitoring using data fusion. Cases involving difficult contexts and data tainted with imperfections, that are classical in industrial contexts, have been emphasized, and solutions using data fusion have been presented and evaluated. An application allowed showing potential benefits of data fusion on estimation performance, even when using heterogeneous data, and also demonstrated that a feature selection method taking into account data imperfection explicitly could perform better than classical approaches.

Fusion of more data collections is planned and should allow drawing more general conclusions. In addition, an algorithm will be implemented to assess the quality level of information contained in a data collection in a deterministic way, as it is a critical parameter of the proposed data alignment approach. A fuzzy system is envisaged as it will allow easy translation from the monitoring system designer semantic thoughts about the quality of a data collection to a numerical parameter.

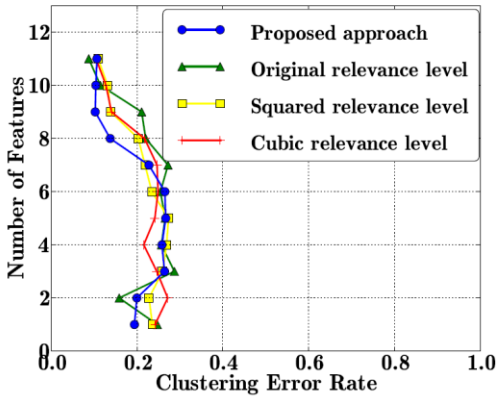

Figure 7. Clustering results obtained for the first data collection when merging 3 data collections to perform feature selection

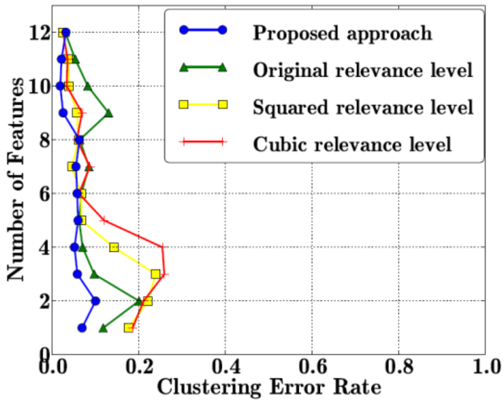

Figure 8. Clustering results obtained for the second data collection when merging 3 data collections to perform feature selection

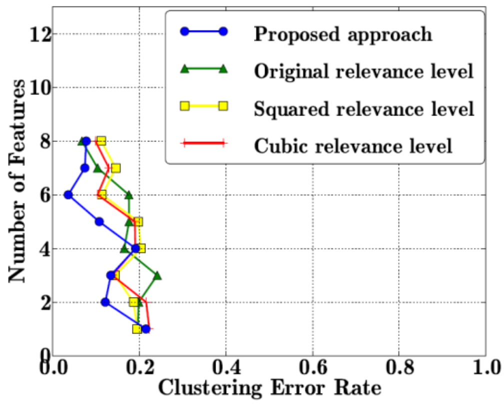

Figure 9. Clustering results obtained for the third data collection when merging 3 data collections to perform feature selection

REFERENCES

[1] I. Guyon and A. Elisseeff, "An Introduction to variable and feature selection," J. of machine learning research, vol. 3, 2003, pp. 1157-1182.

[2] K. Goebel and W. Yan, "Feature selection for tool wear diagnosis using soft computing techniques," Proc. of the ASME, Manufacturing in engineering division, vol. 11, 2000, pp.157-163.

[3] E. Cantù-Paz, S. Newsam, and C. Kamath, "Feature selection for scientific applications," Proc. of the 10th int. conf. on knowledge discovery and data mining, 2004, pp 788-793.

[4] L. Jack and A. Nandi, "Feature selection for ANNs using genetic algorithms in condition monitoring," Proc. of the 7th European symp. on artificial neural networks, 1999, pp. 313-318.

[5] R. Kohavi and G. John, "Wrappers for feature selection," Artificial intelligence, vol. 97, 1997, pp. 273-324.

[6] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," Artificial intelligence, vol. 97, 1997, p. 245-271.

[7] L. C. Molina, L. Belanche and A. Nebot, "Feature selection algorithms: a survey and experimental evaluation," Proc. of the int. conf. on data mining, 2002, pp 306-313.

[8] T.H Loutas, D. Roulias, E. Pauly and V. Kostopoulos, "The combined use of vibration, acoustic emission and oil debris on-line monitoring toward a more effective condition monitoring of rotative machinery," Mechanical systems and signal processing, vol. 25, 2011, pp. 1339-1352.

[9] R. G. Silva and W. J Wilcox, "Sensor based condition monitoring feature selection using a self-organizing map," Proc. of the world congress on engineering, vol. 2, 2008.

[10] J. Bins and B. Draper, "Feature selection from huge feature sets," Int. conf. on computer vision, vol. 2, 2001, pp 159-165.

[11] S. Y. Liang, S. L. Hecker and R. G. Landers, "Machining process monitoring and control: the state of the art," J. of manufacturing science and engeneering, vol. 126, 2004, pp. 397-408.

[12] C.E Everson and S.H Cheragi, "The application of acoustic emission for precision drilling process monitoring," Int. j. of machine tools and manufacture, vol. 39, 1999, pp. 371-387.

[13] H. Almuallim and T. G. Dietterich, "Learning with many irrelevant features," Proc. of the 9th int. conf. on artificial intelligence, 1991, pp. 547-552.

[14] K. Kira and L. Rendell, "A practical approach to feature selection," Proc. of the 2nd int. conf. on machine learning, 1992, pp. 249-256.

[15] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," Proc. of Euro. conf. on machine learning, 1994, pp. 171-182.

[16] M. Robnik and R. Kononenko, "An adaptation of RELIEF for attribute estimation in regression," Proc. of the 14th int. conf. on machine learning, 1997.

[17] Y. Sun, "Iterative RELIEF for feature weighting: algorithms, theories and applications," IEEE Trans. on pattern analysis and machine intelligence, vol. 29, 2007.

[18] D. Hall and S.A.H McMullen, "Mathematical techniques in multisensor data fusion," Artech house, Boston, 2004.

[19] W.L. Oberkampf, S.M. Deland, B.M. Rutherford, K.V. Diegert and K.F Alvin, "Estimation of total uncertainty in modeling and simulation," Sandia report SAND2000-0824, Albuquerque, 2000.

[20] G. Le Moal, G. Moraru, P. Veron, M. Douilly and P. Rabaté "Multisensor data fusion and belief functions for robust singularity detection in signals," Proc. of the 14th int. conf. on information fusion, 2011, pp. 961-968.

[21] G. Le Moal, G. Moraru, P. Veron, M. Douilly and P. Rabaté "Information fusion for singularity detection in difficult cases: a comparison of probabilistic and evidential approaches," Submitted to Information Fusion, 2012.

[22] P. Smets and R. Kennes, "The transferable belief model", Artificial intelligence, vol. 66, 1994, pp. 191-234.

[23] A.P. Dempster,"Upper and lower probabilities induced by multivalued mapping," Annals of mathematical statistics, vol. 38, 1967, pp. 325-339.

[24] G. Shafer, "A mathematical theory of evidence," Princeton university press, Princeton, 1976.