

# Ensemble creation and reconfiguration for activity recognition: An information theoretic approach

Ricardo Chavarriaga, Hesam Sagha and José del R. Millán

Chair in Non-Invasive Brain-Machine Interface (CNBI), Center for Neuroprosthetics

Ecole Polytechnique Fédérale de Lausanne EPFL, Lausanne, Switzerland

{ricardo.chavarriaga, hesam.sagha,jose.millan}@epfl.ch

**Abstract**—Advances in sensing, portable computing devices, and wireless communication has lead to an increase in the number and variety of sensing enabled devices (e.g. smartphones or sensing garments). Pervasive computing and activity recognition systems should be able to take advantage of these sensors, even if they are not always available or appear in runtime. These sensors can be integrated into an ensemble that fuse their information to obtain the final decision. There is therefore a need for mechanisms to select which sensors should compose the ensemble, as well as techniques for dynamically reconfigure the ensemble so as to integrate new sensors.

Sensors can be integrated into an ensemble where information from each of them is fused to obtain the final decisions. From the machine learning point of view, this corresponds to the combination of classifiers where measures of the accuracy and diversity of the ensemble are used to select the elements that may lead to the highest performance. Recent works propose measures of accuracy and diversity based on an information theoretical approach. In this paper we study the use of these measures for selecting ensembles in activity recognition based on body sensor networks. Besides a comparison with traditional diversity measures (e.g.,  $Q$ -,  $\kappa$ -statistics), we also present mechanisms to exploit these measures for the dynamic reconfiguration of the ensemble and detection of changes in the network.

**Index Terms**—Classifier combination, Mutual information, Diversity, Dynamic reconfiguration, Body sensor networks

## I. INTRODUCTION

Pervasive computing and activity recognition systems can provide proactive support by exploiting the knowledge of the user’s context, determined from sensors located on-body, on objects, or in the environment. Applications in health care assistance, manufacturing processes, or entertainment can benefit from such systems, enabled by factors such as technological advances in sensing and portable computing devices and wireless communication. Indeed, as sensing-enabled devices become more available the burden of the problem has partially moved from how to acquire the data to how to deal with it; an issue that is not only restricted to activity recognition [1].

In real life applications, sensors may appear or disappear, change location or be upgraded, thus requiring for dynamic network reconfiguration. In consequence, the traditional approach that conceives systems having a static setup decided at design time is no longer valid. This dynamic reconfiguration should be able to identify the available sensors and decide whether or not it will be useful to integrate them into the network.

In this work, we consider the case where each sensor or

node of the network has an associated classifier and decisions of the individual nodes are fused to compute the overall system output (e.g. classifier fusion). From a machine learning perspective, this corresponds to the case where a large set of classifiers are produced and fused to improve the recognition accuracy [2], [3]. The key problem lies on selecting a set of classifiers (i.e. an ensemble) among the available ones that yields the best performance. This problem is often formulated in terms of optimizing the accuracy and diversity of the classifier ensemble. However, there is still no formal method for performing this selection despite several efforts to characterize these properties [3].

Recently, information theoretic (IT) measures for accuracy and diversity have been proposed as a criteria to build classifier ensembles [4], [5]. In this paper, we test these approaches on activity recognition based on body-worn sensors. We further propose their use for online dynamic reconfiguration of the sensor network and detection of anomalous sensor behavior.

This paper is structured as follows, Section II presents the proposed methods. First, we formulate the classification problem from an information theoretical perspective. Then, we present the IT-based measures of accuracy and diversity proposed by Meynet and Thiran [4] as well as some approaches to exploit these measures for ensemble formation and dynamic reconfiguration. Experimental results using two datasets of activity recognition scenarios are detailed in Section III, followed by the conclusions and future work (Section IV).

## II. METHODS

### A. Information theoretic framework of classification

Previous works have modeled the classification problem following an information theoretic framework. Given a set of examples, we want to design a classifier that provides class labels  $\hat{C}$  that are as close as possible to the true labels  $C$ . Formally, a bound on the error probability  $P_e$ , depending on the mutual information can be obtained based on the Fano’s inequality [6]:

$$\frac{H(C) - I_\alpha(C; \hat{C}) - h(P_e)}{\log N_c - 1} \leq P_e \quad (1)$$

$$P_e \leq \frac{H(C) - I_\beta(C; \hat{C}) - h(P_e)}{\min_k H(C|e, c_k)} \quad (2)$$

where  $H(\cdot)$  is the Shannon entropy,  $h(\cdot)$  is the binary Shannon's entropy,  $N_c$  is the number of classes, and  $I_{\alpha,\beta}(Y;X)$  is the Renyi's mutual information ( $\alpha, \beta \in \mathbb{R}^+ \setminus \{1\}$ ). Consequently, maximizing the mutual information between the estimated and the true labels  $I(C; \hat{C})$  will decrease the probability of errors [4].

### B. Classifier combination

As it is the case for single classifiers, given an ensemble of  $K$  classifiers, the maximal accuracy of the fused output  $\hat{C}$  will be obtained by maximizing the mutual information  $I(C; \hat{C})$ . Obviously, this can be achieved by increasing the accuracy of the individual classifiers in the ensemble. In the same line of thought, maximizing the individual accuracies (therefore increasing  $I(C; C_i)$ ) will increase the mutual information between classifiers ( $I(C_i; C_j)$ ), as they will tend to provide the same (correct) class labels. Consequently, the diversity of the ensemble will decrease as well, highlighting the fact that the trade-off between individual classifier accuracies and diversity should be taken into account in the selection of the classifier ensemble.

In a general sense, the accuracy and diversity of the ensemble are related to the mutual information with respect to the ground truth labels, as well as between its members. Given an ensemble  $S = \{C_1, \dots, C_K\}$ , the mutual information between the true class labels and the output of the classifiers in the ensemble can be expressed as [5],

$$I(C; C_{1:K}) = \sum_{i=1}^K I(C; C_i) - \sum_{\substack{\mathbf{X} \subseteq S \\ |\mathbf{X}|=2..K}} I(\{\mathbf{X}\}) + \sum_{\substack{\mathbf{X} \subseteq S \\ |\mathbf{X}|=2..K}} I(\{\mathbf{X}\}|C) \quad (3)$$

where  $I(\{\mathbf{X}\}|C)$  is the conditional mutual information, i.e. the mutual information between the elements of  $\mathbf{X}$ , knowing the value of  $C$ . The first term is related to the accuracy of single classifiers while the remaining ones are related to the diversity within the ensemble. The second term reflects the interaction information among all possible subsets of the ensemble, while the third term measures the class-conditional correlations between subsets of the ensemble. Equation 3 can be rewritten as

$$I(C; C_{1:K}) = \sum_{i=1}^K I(C; C_i) - \sum_{|\mathbf{X}|=2} I(\{\mathbf{X}\}) + \sum_{|\mathbf{X}|=2} I(\{\mathbf{X}\}|C) - \sum_{|\mathbf{X}|=3} I(\{\mathbf{X}\}) + \sum_{|\mathbf{X}|=3} I(\{\mathbf{X}\}|C) - \dots - \sum_{|\mathbf{X}|=K} I(\{\mathbf{X}\}) + \sum_{|\mathbf{X}|=K} I(\{\mathbf{X}\}|C) \quad (4)$$

where the expression in the second line corresponds to a pair-wise measure of diversity, the third line to a 3-way diversity and so on [5].

---

### Algorithm 1 IT-based creation of classifier ensemble

---

First select the best individual classifier:

$$C_1^* = \arg \max_{C_i} I(C; C_i) \quad (8)$$

Then, select two more classifiers so as to maximize ITS:

$$(C_2^*, C_3^*) = \arg \max_{\substack{(C_i, C_j) \\ i, j=1, \dots, M \setminus \{C_1^*\}}} ITS(C_1^*, C_i, C_j) \quad (9)$$

Recursively add classifiers that maximize ITS until reaching the intended ensemble size.

---

Along the same line, Meynet and Thiran proposed a heuristic to estimate accuracy and diversity in the classifier ensemble by taking into account individual accuracies and pairwise mutual information among classifiers (i.e. the first two terms in equation 4) [4]. Considering the case where the ensemble decision is made by majority voting they define a measure of the information theoretic accuracy of the ensemble (ITA) as the average accuracy of individual classifiers,

$$ITA = \frac{1}{K} \sum_{i=1}^K I(C; C_i) \quad (5)$$

and an information theoretic diversity measure (ITD) inversely proportional to the pair-wise mutual information between individual classifiers,

$$ITD = \frac{\binom{K}{2}}{\sum_{i=1}^{K-1} \sum_{j=i+1}^K I(C_i; C_j)} \quad (6)$$

Using an empirical evaluation of the relation among these factors they conclude that  $ITA^2 \propto 1/ITD$ , and define an information theoretic score defined as,

$$ITS = (1 + ITA)(1 + ITA)^2(1 + ITD) \quad (7)$$

where the first factor aims at maximizing the average classifier accuracy while the remaining ones are intended to promote diversity given similar ITA values.

### C. Ensemble formation

The IT-based measures introduced in Section II-B reflect the accuracy and diversity of a given ensemble. Then, given a pool of  $M$  available classifiers, a classifier ensemble can be created so as to maximize the ITS (i.e. aiming at obtaining the highest accuracy and diversity). An iterative method was proposed for the creation of an ensemble  $S^* = \{C_1^*, C_2^*, \dots, C_K^*\}$ , where nodes are added incrementally as shown in Algorithm 1 [4]. This approach allows for the creation of the classifier ensembles without requiring an exhaustive search among all possible classifier combinations.

TABLE I  
PROBABILITIES OF COINCIDENT ERRORS FOR TWO CLASSIFIERS  $f_i, f_j$

	$f_j$ correct	$f_j$ wrong
$f_i$ correct	a	b
$f_i$ wrong	c	d

#### D. Other diversity measures

In the past, several measures have been empirically proposed to measure the diversity of a classifier ensemble [3]. Among them, we can cite the  $Q$ -statistic that is based on the probability of coincident errors for two classifiers (c.f. Table I). Given these probabilities it is defined as [7],

$$Q_{ij} = \frac{ad - bc}{ad + bc} \quad (10)$$

and the overall  $Q$ -statistic of an ensemble is computed as the average  $Q$ -value for all pairwise combination of classifiers. Similarly, another measure, termed the  $\kappa$ -statistic, has been defined as,

$$\kappa = \frac{2(ac - bd)}{(a + b)(c + d) + (a + c)(b + d)} \quad (11)$$

Other measures to assess or optimize diversity have been proposed using machine learning approaches [8], [9]. They have empirically shown to reflect the diversity of the ensemble, and somehow relate to the ensemble accuracy. However, a formal link between these diversity measures and the actual accuracy is yet to be formalized [10].

#### E. Ensemble reconfiguration

As mentioned before, real-life activity recognition systems are prone to changes due to several factors (e.g. sensor failure, loss of power, addition of new sensors, etc.). This raises the need for dynamic mechanisms that allow for the exploitation of new resources that were not originally considered at design time. These characteristics require the system to be able to reconfigure the classifier ensemble on runtime without the need for retraining the whole system [11]. A recent approach has been proposed to support this aspect by relying on nodes that are able to self-describe themselves and advertise their capabilities (e.g. activities they are able to recognize) [12]. Accordingly, upon the appearance of a new node, its self-description information can be used to include it into the pool of available classifiers and decide whether it should be part of the ensemble (i.e. by estimating the accuracy and diversity of ensembles that include that particular node).

Measures of ensemble accuracy and diversity are computed using training data as they require the availability of the true class labels. Moreover, they normally require the use of the same training data for all classifiers, as they are based on the quantification of coincident errors—as is the case for the  $Q$ - and  $\kappa$ -statistics (Eq. 10, 11, respectively)—or consider the class-conditional correlation among classifiers (Eq. 3). Regarding the ITS, only the accuracy term (ITA) requires the ground truth labels to be computed. In turn, the diversity measure

(ITD) only requires to know the output of the classifiers given the same input. In consequence, the two measures can be computed on different sets of data samples provided that they come from the same distributions.

According to this, when designing a system all available classifiers can be trained using the same dataset and can have an estimation of their ITA and ITD over the same samples. If a new node, trained over the same set of classes using a (possibly) different training set, is added in the future, its mutual information with the already available nodes,  $I(C_i; C_j)$ , can be computed using samples gathered on runtime, even if ground truth labels are not available. Therefore, the initial self-description is only required to have an estimation of the node's accuracy,  $I(C; C_i)$ . Furthermore, after proper estimation, the pair-wise mutual information can be added to the node's self-descriptor to be used in the future.

We test this approach for exploiting new resources based on the online estimation of node's mutual information. As mentioned before, this approach relies on the assumption that the training samples and those gathered on runtime come from the same feature distribution. Furthermore, it is clear that the proper estimation of the mutual information across classifiers will depend on the sampling done during the system operation (i.e. in terms of number of samples per class).

#### F. Anomaly detection

The previous sections show how the information theoretic measures can be used to create a classifier ensemble. These measures use the pair-wise mutual information between the classifiers to characterize their diversity. We will explore whether this measure  $I(C_i; C_j)$  can further be used to detect changes in the system. In general, these changes (e.g. due to sensor failures, displacements, etc.) are reflected in the feature distribution and consequently in the classifiers outputs. Previous approaches use training data to characterize the normal system behavior. Then deviations from these measures are used to detect *anomalous* behavior. This approach has been applied by characterizing the distributions at the feature level [13] as well as at the level of the classification outputs in classifier ensembles [14].

Similarly, the mutual information between classifiers will be affected by those changes. In consequence, we assessed how this mutual information  $I(C_i; C_j)$  changes when noise is added to the channels in the system. Using the values estimated during training as a reference ( $I_{i,j}^{tr} = I(C_i; C_j)$ ) deviations from these values would correspond to anomalous sensor behavior.

### III. EXPERIMENTAL RESULTS

We evaluated the proposed approach on two datasets corresponding to activity recognition scenarios using on-body sensors. In both cases we use a motion jacket equipped with nodes containing accelerometer, rate gyro and magnetic sensors mounted on different points on the arms and upper torso of the user. In all the experiments, data was segmented into meaningful actions and was classified using a Linear

Discriminant Analysis (LDA) classifier with the mean and variance of the signals as features. Then majority voting is used to compute the final decision of the ensemble.

The first dataset corresponds to a car manufacturing scenario including 20 activity classes, e.g. open/close hood, open/close door, etc [15]. It contains data from seven subjects performing 10 recording sessions each. For each subject, seven on-body locations were recorded resulting in a total of 21 sensors. Performance was evaluated using leave-one-subject-out cross validation. For each subject there were about 300 samples.

The second dataset included daily home activities in a breakfast scenario (also referred to as *Opportunity* dataset) [16], [17]. We performed classification of 24 classes corresponding to low-level actions (e.g. open /close Fridge, open/close door, reach cup) using sensors at five locations (i.e. 15 sensors). We report results for one subject using 3-fold cross validation. For each fold there were about 700 samples for training and 350 in the testing set.

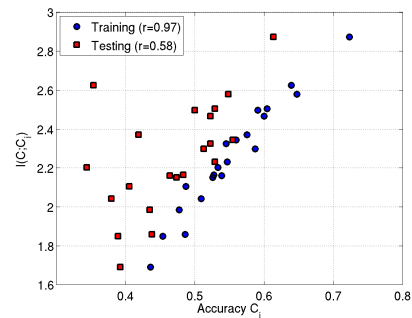
Figure 1 illustrates the relation between the individual classification accuracy and the mutual information between the classifier output labels and the true class labels  $I(C_i, C)$ , for one fold of each dataset. As expected, a clear correlation between these variables was observed. The average correlation in the car scenario was  $0.969 \pm 0.007$  and  $0.492 \pm 0.101$  for the training and testing datasets, respectively. In the case of the daily living scenario, these correlations were  $0.983 \pm 0.003$  and  $0.930 \pm 0.001$ . Unsurprisingly, correlation decreases in the testing sets, presumably due to non-stationarities on the used signals; moreover, this decrease is larger in the car manufacturing scenario in which the training and testing datasets were collected from different users.

#### A. Ensemble formation

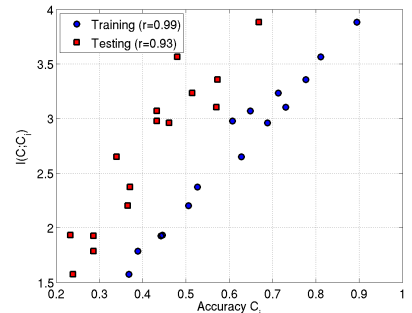
We compared the performance of different methods to create the classifier ensembles. As described in Section II-C given an existing ensemble, two new members are selected from the pool so as to maximize a pair-wise diversity measure. We compared 3 diversity measures corresponding to the Information theoretic score (*ITS*, Equation 7), *Q*-statistic (*QS*, Equation 10) and  $\kappa$ -statistic (*KS*, Equation 11).

The average testing performance of ensembles obtained using each method is shown in Table II. As a baseline, for each ensemble size we randomly generated up to 100 different ensembles, and report their mean and maximal accuracy (*mean*, and *max*, respectively). Unsurprisingly, classification accuracy increases as more classifiers are included in the ensemble. The average classification when all nodes are used is  $74.87 \pm 26.79$  and  $78.72 \pm 2.74$  for the car manufacturing and daily life scenario, respectively. Maximization of ITS consistently outperforms other ensemble creation methods. However, it should be noticed that maximizing ITS yields one of the best possible ensembles, but does not necessarily yield the best accuracy that can be achieved.

These results suggest that the ITS –computed using pair-wise mutual information across sensors– can be used reli-



(a) Car manufacturing scenario



(b) Daily living scenario

Fig. 1. Relation between the mutual information  $I(C_i, C)$  computed on the training set and the accuracy of individual classifiers for both training (blue dots) and testing set (red squares).

ably to build ensembles of a larger size. This is particularly interesting for opportunistic systems, as the self-description information of each node contains only pair-wise information with respect to other sensors in the network rather than a fully characterization of the possible ensembles that can be created.

#### B. Ensemble reconfiguration

As described in section II-E, the IT-based measures can be used to integrate new nodes into the ensemble, even if they have not been trained on the same dataset as the current nodes in the system. We tested this approach using the car manufacturing scenario assuming two sets of sensors trained on separate sets of samples. As before, we divided the dataset into 7 folds. In each case a separate subject was left for testing. Then, for each fold we created an initial ensemble  $S^1$  composed of 5 nodes trained on three subjects, while the set of remaining nodes  $S^2$  was trained on the other three subjects.

We simulated a scenario in which the system, already operating with the ensemble  $S^1$ , discovers the new set nodes which describe themselves with their estimated accuracy  $I(C; C_i)$ . Then, the ITD is estimated using half of the testing samples of the new subject (about 150 samples). This value, combined with the ITA is used to compute the ITS and to select ensembles of larger size as in the previous section.

Figure 2 shows the average accuracy of different ensemble sizes in the remaining samples in the testing set. The dotted blue line shows the mean performance of randomly generated ensembles, and the blue box-plots show the distribution of the

TABLE II  
AVERAGE ACCURACY ( $\pm$  STD) OF CLASSIFIER ENSEMBLES USING DIFFERENT STRATEGIES

Car manufacturing scenario							
K	3	7	9	11	15	19	
mean	38.75 $\pm$ 5.24	60.52 $\pm$ 12.55	64.35 $\pm$ 15.11	67.78 $\pm$ 17.99	73.33 $\pm$ 18.22	74.32 $\pm$ 25.19	
max	61.87 $\pm$ 4.52	77.67 $\pm$ 6.41	80.14 $\pm$ 6.07	81.67 $\pm$ 6.54	83.95 $\pm$ 6.28	84.92 $\pm$ 6.47	
ITS	40.68 $\pm$ 18.28	70.92 $\pm$ 8.89	73.62 $\pm$ 7.86	75.81 $\pm$ 6.96	78.97 $\pm$ 6.67	74.17 $\pm$ 27.10	
QS	33.22 $\pm$ 17.56	47.81 $\pm$ 18.90	42.64 $\pm$ 28.00	46.21 $\pm$ 30.80	66.98 $\pm$ 23.17	73.24 $\pm$ 26.18	
KS	37.88 $\pm$ 21.86	62.46 $\pm$ 22.12	55.19 $\pm$ 34.63	68.55 $\pm$ 24.45	80.64 $\pm$ 7.77	74.40 $\pm$ 26.92	

Daily life scenario							
K	3	5	7	9	11	13	
mean	46.14 $\pm$ 5.18	60.01 $\pm$ 8.85	65.97 $\pm$ 7.38	68.61 $\pm$ 8.98	74.05 $\pm$ 5.84	76.78 $\pm$ 3.92	
max	76.14 $\pm$ 4.17	80.12 $\pm$ 4.09	82.53 $\pm$ 2.32	81.67 $\pm$ 2.63	82.87 $\pm$ 2.45	81.12 $\pm$ 2.41	
ITS	77.37 $\pm$ 7.67	80.67 $\pm$ 4.37	78.34 $\pm$ 5.14	79.18 $\pm$ 4.65	81.30 $\pm$ 2.10	79.93 $\pm$ 2.44	
QS	75.69 $\pm$ 8.59	70.42 $\pm$ 8.82	72.09 $\pm$ 3.81	74.00 $\pm$ 2.78	76.94 $\pm$ 4.10	76.14 $\pm$ 1.78	
KS	74.06 $\pm$ 5.58	49.89 $\pm$ 28.89	50.83 $\pm$ 29.98	66.28 $\pm$ 6.11	70.33 $\pm$ 4.23	75.22 $\pm$ 3.12	

best ensembles of these sets. Red box-plots show the performance of the ensembles created by maximizing the IT score. It shows that despite using information about the accuracy and diversity obtained from different datasets, these ensembles generally perform better than average. Moreover, it should be taken into account that the estimation of the ensemble diversity is not very accurate given the small set of samples used for this purpose. Overall, this suggest the feasibility of using this approach for the dynamic reconfiguration of classifier ensembles. However, further studies are required to define criteria to decide when the estimation of the ensemble diversity is reliable enough to consider the new nodes as eligible to be included in the ensemble.

### C. Anomaly detection

We then assessed whether the mutual information can be used to detect changes in the network. In order to do this, we artificially added noise to the testing sets of the two scenarios and evaluate how the pair-wise classifier mutual information estimated on the noisy testing set  $I_{i,j} = I(C_i; C_j)$  changes with respect to the estimation done in the training data  $I_{i,j}^{tr} = I(C_i; C_j)$ .

For a given sensor  $i$  we define two vectors

$\mathbf{I}_i = \{I_{i,1}, I_{i,2}, \dots, I_{i,K}\}$  and  $\mathbf{I}_i^{tr} = \{I_{i,1}^{tr}, I_{i,2}^{tr}, \dots, I_{i,K}^{tr}\}$  containing the pair-wise mutual information between the sensor  $i$  and other sensors in the ensemble computed in the testing and training sets respectively. The Euclidean distance between these vectors was used to quantify these changes for both the original and noisy testing sets. We considered five levels of additive Gaussian noise (SNR of 100, 20, 10, 5, and 1dB).

As previously mentioned, the estimation of the mutual information between classifiers may depend on the used samples. Therefore any observed difference between  $\mathbf{I}_i$  and  $\mathbf{I}_i^{tr}$  may be due to a sampling problem instead of anomalous sensor behavior. To counter that, we assessed the variability of this estimation by computing it on different partitions of the training set. To this end, we split the training set in two parts A and B (1/5 and 4/5 of the data respectively) and compute the distance between the MI computed in each part  $\Delta \mathbf{I}_i = |\mathbf{I}_i^A - \mathbf{I}_i^B|$ . The process was repeated 100 times using random permutations of the data and the maximal values were used as a baseline to assess how much the estimated MI can change using data from the *same* training feature distribution.

Figure 3 shows the Euclidean distance between the MI vectors obtained in the training set and different testing conditions. Data was normalized for each sensor dividing by the maximum value of  $\Delta \mathbf{I}_i$  obtained in the training set. On each plot, the leftmost bar corresponds to the original testing set with no noise added. In both scenarios, the computed distance increases consistently as additive noise is included in the signals. Suggesting that it can be used to detect these types of anomalies in the network (e.g. instead of the Mahalanobis distance used in [14]).

## IV. CONCLUSION

In this work we assess the performance of information theoretic measures of accuracy and diversity for the creation of classifiers ensembles for activity recognition. Results on a daily living and a car manufacturing scenarios confirm the suitability of these measures, as it was previously shown in other pattern recognition problems [4]. Moreover, the presented mechanism outperforms other traditional measures of diversity. The use of an incremental algorithm that maximizes the IT-score leads to

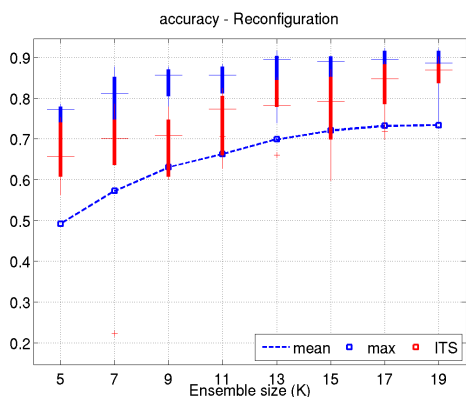
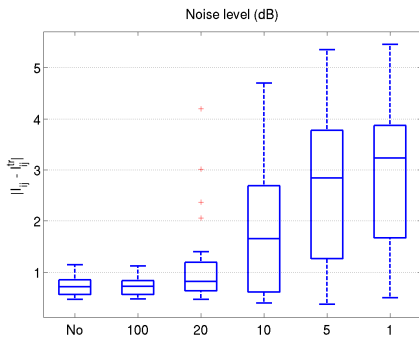
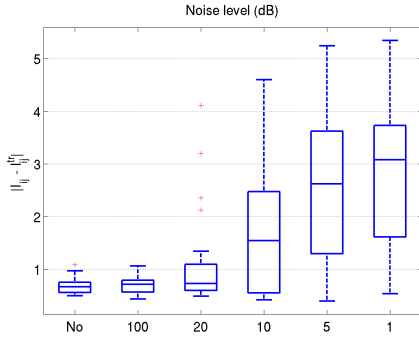


Fig. 2. Accuracy of the ensembles during dynamic reconfiguration. Car manufacturing scenario



(a) Car manufacturing scenario



(b) Daily living scenario

Fig. 3. Changes in the pair-wise classifier mutual information (with respect to the training data) when noise is added to one sensor in the network. Values are normalized for each node (See text for details).

ensembles with high accuracy without requiring an exhaustive search. Nevertheless, it should be noticed that the method may not lead to the highest possible performance.

The ITS is based on the pair-wise mutual information across classifiers in the ensemble, disregarding higher order relationships (e.g. 3-way diversity and higher). Despite this assumption, it seems to provide a reliable way to estimate the expected performance of the different ensembles. This is particularly important as it simplifies the computation of the score; moreover, it reduces the amount of information that has to be stored to characterize potential ensembles. In the case of an architecture relying on nodes self-descriptions, as proposed by Kurz et al. [12], each node has to store only its mutual information with respect to other sensors instead of all the possible node combination.

The use of a diversity measure that does not depend on coincident errors but only on the mutual information across classifiers make it possible to use it for dynamic reconfiguration of the ensemble. In this case, even if new sensors have been trained on a different dataset than the current ensemble, the diversity measure can be estimated using samples acquired on runtime. This allows for an estimation of the ITS of ensembles that incorporates new sensors as shown in section III-B. Nevertheless, the accuracy of such estimation will strongly depend on the number of samples used for such purpose.

Further work is therefore needed to define proper criteria for the online assessment of the diversity measure. Moreover, the proposed approach assumes that fusion is performed using majority voting. Current research is being undertaken to extend it to other fusion mechanisms as well as to make it more suitable for problems with a large number of classes.

## ACKNOWLEDGMENTS

This work was supported by the EU-FET project ICT-225938 *Opportunity: Activity and Context Recognition with Opportunistic Sensor Configuration*.

## REFERENCES

- [1] R. G. Baraniuk, "More is less: signal processing and the data deluge." *Science*, vol. 331, pp. 717–719, 2011.
- [2] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 226–239, 1998.
- [3] G. Brown, J. Wyatt, R. Harris, and X. Yao, "Diversity creation methods: A survey and categorisation," *Journal of Information Fusion*, vol. 6, pp. 5–20, 2005.
- [4] J. Meynet and J.-P. Thiran, "Information theoretic combination of pattern classifiers," *Pattern Recognition*, vol. 43, pp. 3412–3421, 2010.
- [5] G. Brown, "An information theoretic perspective on multiple classifier systems," in *Intl Workshop on Multiple Classifier Systems*, 2009.
- [6] D. Erdogmus and J. C. Principe, "Lower and upper bounds for misclassification probability based on Renyi's information," *The Journal of VLSI Signal Processing*, vol. 37, pp. 305–317, 2004.
- [7] G. Yule, "On the association of attributes in statistics," *Biometrika*, vol. 2, pp. 121–134, 1903.
- [8] V. Khare and X. Yao, "Artificial speciation of neural network ensembles," in *2002 UK Workshop on Computational Intelligence (UKCI'02)*, J. A. Bullinaria, Ed., 2002, pp. 96–103.
- [9] C. Villalonga, D. Roggen, and G. Tröster, "Shaping sensor node ensembles according to their recognition performance within a planning-based context framework," in *Eighth International Conference on Networked Sensing Systems*, 2011.
- [10] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine Learning*, vol. 51, pp. 181–207, 2003.
- [11] D. Roggen, K. Förster, A. Calatroni, T. Holleczeck, Y. Fang, G. Tröster, P. Lukowicz, G. Pirkel, D. Bannach, K. Kunze, A. Ferscha, C. Holzmänn, A. Riener, R. Chavarriaga, and J. Millán, "OPPORTUNITY: Towards opportunistic activity and context recognition systems," in *IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications*, 2009.
- [12] M. Kurz, G. Hözl, A. Ferscha, H. Sagha, J. d. R. Millán, and R. Chavarriaga, "Dynamic Quantification of Activity Recognition Capabilities in Opportunistic Systems," in *4th Conference on Context Awareness for Proactive Systems: CAPS2011*, 2011.
- [13] C. Alippi and M. Roveri, "An adaptive cusum-based test for signal change detection," in *IEEE Int Symposium Circuits and Systems (ISCAS)*. IEEE, 2006.
- [14] H. Sagha, J. d. R. Millán, and R. Chavarriaga, "Detecting anomalies to improve classification performance in an opportunistic sensor network," in *7th IEEE International Workshop on Sensor Networks and Systems for Pervasive Computing, PerSens 2011*, 2011.
- [15] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster, "Wearable activity tracking in car manufacturing," *IEEE Pervasive Computing Magazine*, vol. 7, pp. 42–50, 2008.
- [16] D. Roggen, A. Calatroni, M. Rossi, T. Holleczeck, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkel, A. Ferscha, J. Doppler, C. Holzmänn, M. Kurz, G. Holl, R. Chavarriaga, H. Sagha, H. Bayati, M. Creatura, and J. D. R. Millán, "Collecting complex activity data sets in highly rich networked sensor environments," in *Seventh International Conference on Networked Sensing Systems*, 2010.
- [17] P. Lukowicz, G. Pirkel, D. Bannach, F. Wagner, A. Calatroni, K. Förster, T. Holleczeck, M. Rossi, D. Roggen, G. Tröster, J. Doppler, C. Holzmänn, A. Riener, A. Ferscha, and R. Chavarriaga, "Recording a complex, multi modal activity data set for context recognition," in *Workshop on Context-Systems Design, Evaluation and Optimisation at ARCS, 2010*, 2010.