XVIII IMEKO WORLD CONGRESS
Metrology for a Sustainable Development
September, 17 – 22, 2006, Rio de Janeiro, Brazil

EVOLUTIONARY PATTERN RECOGNITION FOR MEASUREMENT OF VEHICLE EMISSION FACTORS IN CRITICAL DRIVING CONDITIONS

Pasquale Arpaia [1], *Fabrizio Clemente* [2], *Mario Rapone* [3], *Carmine Romanucci* [4]

[1] Dipartimento di Ingegneria, Università del Sannio, Corso Garibaldi 107, 82100 Benevento, ITALY. Ph : +39 0824 305804-17, Fax: +39 0824 305840, E-mail: arpaia@unisannio.it
[2] Consiglio Nazionale delle Ricerche, Istituto di Ingegneria Biomedica - Sez. di Roma, Area di Ricerca Roma 1, Montelibretti, via Salaria km 29,300 - 00016 Monterotondo S. (RM), Ph: +39 (0)690 67 26 92, Fax: +39 (0) 6 90 67 28 29, E-mail: clemente@isib.cnr.it;
[3] Consiglio Nazionale delle Ricerche, Istituto Motori, Viale Marconi 8, 80125 Napoli, ITALY;
[4] Dipartimento di Ingegneria dei Materiali e della Produzione, Università di Napoli Federico II, Piazzale Tecchio 80, 80125 Napoli, ITALY. Ph :+39 081 7177232, E-mail: romanuccicar@libero.it.

Abstract: A research project on the measurement of emission factors in critical driving conditions arising from a cooperation among University of Sannio, Istituto Motori of CNR, and some companies of Italian Regione Campania is started. In particular, in this paper, a procedure of automatic feature extraction and data classification of vehicle driving sequences by means of Wavelet analysis and Cultural Algorithms is proposed. The procedure improves the lack of sensitivity of the state-of-the-art statistical methods to variations of vehicles instantaneous kinematic parameters (e.g. speed, acceleration) through a simple heuristic link between vehicle speed and road traffic data.

Keywords: Cultural Algorithm, Feature Extraction, Vehicle Emissions.

1. INTRODUCTION

The evaluation of vehicle emission factors is necessary for quantifying the impact of traffic flows on air quality [1]-[2]. Thus, in last years, a scientific debate arises on the choice of reference *driving-cycles* (DCY) (Fig. 1) used to perform emission measurements in the laboratory or to associate on-road emission measurement to vehicle driving behavior. Many research associations proposed own definitions [3]. Several studies are based upon the collection and statistical analysis of data from large numbers of vehicles/trips in particular regions [4].

Nevertheless, in this approach, the definition of the so called *micro-trips* or *sequences* in a driving-cycle is of basic relevance [1]-[5]. Micro-trips are subsets of DCYs whose features contribute, through a clustering procedure, to associate the current traffic condition to reference DCY defined by different associations (EPA, ARTEMIS) [1]-[5], [7]. Thus a definition of such features is critical. State of the art exploits a kinematic approach computing for each sequence features such as: speed, acceleration, and some aggregated variables (e.g. time spent in acceleration, time spent in cruise mode and so on) [3]-[6].

In this paper, arising from the scientific cooperation among University of Sannio, Istituto Motori of CNR, and some companies of Italian Regione Campania, the problem is faced by a heuristic classification procedure. In particular, the procedure was based on a (i) *Feature Extraction* (FE) algorithm, exploiting *wavelet transforms*, and a (ii) *Cultural Algorithm* (CA).

The use of an advanced FE technique, such as wavelet transform [9]-[13], is necessary in order to improve the sensitivity of classical statistical methods. Wavelets are well known functions whose energy is concentrated both in time and frequency [13]; they are successfully applied to non-stationary signals and images [9]-[11], noise filtering, and feature extraction of automotive signals [9]-[13]. In our idea the CA has to produce, using an heuristic approach, an optimal evaluation criteria to classify the extracted features.

CAs represent the state of the art of *Evolutive Algorithms* (EA); culture is there exploited to accelerate evolutive process (*cultures evolves faster than coltures* [14]-[15]). Several papers have shown that CAs are more effective and accurate than classical EAs [12]-[17]; they are mainly composed by a *population space* and a *belief space* communicating through a communication protocol [15]. The former space contains a set of possible solutions of the problem [14]-[19], and it is usually constituted by a Genetic

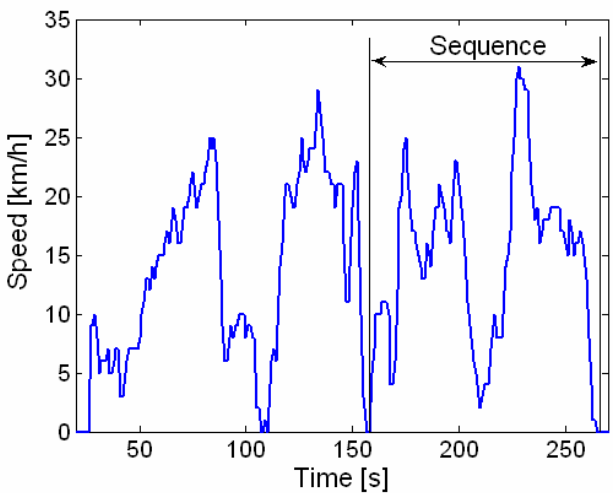

Fig. 1 – An example of driving cycle and sequence to be identified.

Algorithm (GA) [14]-[15], [18]; the latter is a specific mechanism of evolution pressure; the characteristics of best individuals updates the knowledge in the belief space by evolving the information for a faster and better solution search. In this way, the intrinsic resource waste of genetic evolution is avoided by driving evolution suitably through such a cultural mechanism [14]-[16].

2. THE PROPOSED PROCEDURE

The proposed procedure classifies the different sequences, obtained through a *signal segmentation* algorithm [9]-[10], by a given signal, in a prefixed number of clusters. The procedure is subdivided in two main parts: (i) the wavelet-based *Feature Extraction section*, and (ii) *the CA section*.

2.1. Feature Extraction Section

The main task of this section is to apply the wavelet transforms to each of the extracted sequence in order to obtain the features that will be subsequently processed by the proposed CA. The following notation is introduced [10]:

- D_l is the l-th detail level of wavelet coefficients;
- d_{li} is the i-th coefficient in the l-th level;
- $[a, b]$ represents a segment with boundaries at samples a and b whose midpoint is named m;
- t_i is the time of the sample normalized from 0 to $(b - a)$.

The subsequent features were extracted [10]:

1) *Wavelet Coefficient Average Level*:

$$WA_l = \frac{\sum_{i=a}^{b} d_{li}}{b-a} \tag{1}$$

representing the mean of detail coefficients for l-th level; this feature contains information about the signal slope.

2) *Wavelet Coefficient Energy*:

$$WE_l = \frac{\sum_{i=a}^{b} d_{li}^{\,2}}{b-a} \tag{2}$$

providing information on noise level and frequency over the segment.

3) *Wavelet Coefficient X-Centroid*:

$$XC_l = \frac{\left(\sum_{i=a}^{b} t_i \cdot d_{li}\right) - \min(D_l) \cdot \sum_{i=a}^{b} t_i}{\left\{\left(\sum_{i=a}^{b} d_{li}\right) - \min(D_l) \cdot (b-a)\right\} \cdot (b-a)} \tag{3}$$

giving an indication of the location in time at which an event occurs.

4) *Wavelet Quarters*:

$$WQ_l^1 = \frac{\sum_{i=a}^{m} d_{li}}{m-a} \qquad \text{if } d_{li} > 0 \tag{4}$$

$$WQ_l^2 = \frac{\sum_{i=a}^{m} d_{li}}{m-a} \qquad \text{if } d_{li} < 0 \tag{5}$$

$$WQ_l^3 = \frac{\sum_{i=m}^{b} d_{li}}{b-m} \qquad \text{if } d_{li} > 0 \tag{6}$$

$$WQ_l^4 = \frac{\sum_{i=m}^{b} d_{li}}{b-m} \qquad \text{if } d_{li} < 0 \tag{7}$$

representing the mean value of the detail coefficient for each of the four quadrants of the segment.

2.2. The CA section

The CA's inputs are a set of features vectors (one for each sequence), and a reference classification (e.g. obtained through statistical procedures) based on an array whose *i-th* element is the cluster number of the *i-th* sequence.

The distribution of these feature vectors is characterized via vector quantization (VQ) trained using a generalized K-means algorithm [20]. Each individual of the CA is a matrix, whose row represent the coordinates, in the feature space, of all possible centroids with the associated cluster. On the basis of the euclidean distance, therefore, each sequence will be associated to the closet centroid, and, correspondingly, to the related cluster.

Thus, the main task of the CA. is to find the coordinates and the minimum number of centroids able to identify the different clusters univocally, as schematically shown in Fig.2, for a 2D space. The fitness function corresponds to the percentage of sequences labelled by the CA as in the reference classification. The percentage of the sequences correctly labelled is weighted by the number of the centroids used for the classification, in order to promote the solution with less centroid.

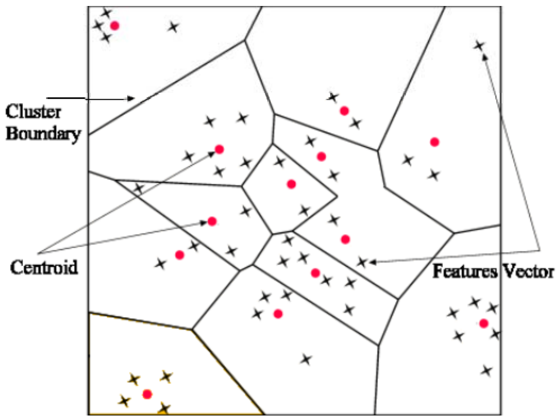

Fig. 2 – A 2D scheme of the features space.

3. PRELIMINARY EXPERIMENTAL RESULTS

The system was implemented in Matlab™. For the FE, the mother wavelet Bior5.5 was chosen due to the high correlation with the speed set. For the CA default configuration, according to literature [14]-[18], the following parameter setup was used: migration percentage 10, crossover percentage 90, elite count 1, GA cycles 20, accept percentage 20, population size 700 and influence type situational, topographical and historical.

The reference classification was achieved through a statistical procedure applied to on-field measurements on a large fleet of vehicles driving in Naples[4]. In a speed set of 114000 samples, 1523 sequences (i.e. a set of speed data with null first and last values) were found; for each sequence, wavelet analysis was carried out at the three most significant levels in order to extract 21 features. The data were The CA's solution was able to classify 81% of the sequences of the reference classification; for each of the clusters.

4. CONCLUSIONS

A research project on the measurement of emission factors in critical driving conditions arising from a cooperation among University of Sannio, Istituto Motori of CNR, and some companies of Italian Regione Campania was conceived.

Preliminary results of a CA-based approach to classification showed satisfying quality, although fitness values can be improved.

Next steps are mainly concerned to: (*i*) selection of features by a better discriminating power, and (*ii*) the integration of the proposed procedure in a larger DCY identification system.

ACKNOWLEDGMENTS

Experimental data were obtained under the framework of Atena Project whose support authors gratefully acknowledge. Authors thank also Giuseppe "Tiger" Lucariello for gathering experimental data.

REFERENCES

[1] A. Cappiello, I. Chabini, E. K. Nam, A. Luè, M. Abou Zeid, "A statistical model of vehicle emissions and fuel consumption", IEEE Proc. Int. Conf. on Intell. Transp. Sys., Singapore, September 3-6 2002, pp 801-809.

[2] K. Ahn. H. Rakha, M. van Aerde "Microscopic fuel consumption, and emission models." in Proc. of the 79th Annual Meeting of the Transportation Research Board, Washington D.C., January 1999.

[3] André M. "Real-world driving cycles for measuring cars pollutant emissions - Part A: The Artemis European driving cycles." INRETS report, Bron, France, n°LTE 0411, 97 p. 2004.

[4] L. Gortan, L. Mina, A. Fortunato, L. Borgarello L. Della Ragione, G. Meccariello, M. V. Prati, M. Rapone, "Preliminary results on emission and driving behaviour of ATENA fleet test project in Naples", Proceedings 5th international Conference on Internal combustion engines – ICE2001, Capri (Napoli) 2001, pp.52-60. SAE_NA Technical Paper 2001-01-083.

[5] L. Della Ragione, G. Meccariello, M. Rapone, V. Punzo, V. Torrieri, "Some comparisons between experimental emission data and simulation outputs of behavioral acceleration and PLS regression-based emission models", Proceedings 6th international Conference on Engines for automobile – ICE2003, Capri (Napoli) 2003, Paper SAE_NA 2003-01-57.

[6] M. Rapone, M. V. Prati, L. Della Ragione, G. Meccariello, M. A. Costagliola, "A novel Statistical model for the evaluation of vehivle emission factors. Application to a Euro III gasoline car fleet", Proc on ICE 2005 7th Int. Conf. on Eng. for Aut., Capri,Naples, Italy, September 11-16 2005.

[7] Hickman A J and McCrae I S (editor) (2003). Revised technical annex (February 2003) ARTEMIS. Assessment and reliability of transport emission models and inventory systems. Project funded by the European Commission within the 5th Framework Research Programme. DG TREN Contract No. 1999-RD.10429.ARTEMIS website - http://www.trl.co.uk/artemis/.

[8] EPA United States Environmental Protection Agency Air and Radiation EPA420-R-03-010 August 2003 User's Guide to MOBILE6.1 and MOBILE6.2 Mobile Source Emission Factor Model.

[9] Emfac2001/Emfac2002 Calculating emission inventories fro vehicles in California User's Guide (www.arb.ca.gov).

[10] Joumard R., Jost P., Hickman, J., (1995) Influence of Instantaneous Speed and acceleration on Hot passenger Car Emissions and Fuel Consumption SAE paper 950928.

[11] H.Guo, J.A.Crossman, Y.L. Murphey, M.Coleman, "Automotive signal diagnostics using wavelets and machine learning", *IEEE Trans.Veh.Technol.*,vol.49, Nov. 2000.

[12] Y. L. Murphey, H. Guo, "Automatic feature selection – a hybrid statistical approach", Proc. Int. Conf. Pattern Recogn., Barcelona, Spain, September 3-8, 2000.

[13] E. Jones, P. Runkle, "Genetic algorithm wavelet design for signal classification", *IEEE trans Pat. An. and Mac. Intell.*, Vol. 23, No. 8, August 2001.

[14] L. Cristaldi, M. Lazzaroni, A. Monti, F. Ponci, "A neurofuzzy application for AC motor drives monitoring system", *IEEE Trans. Instr. and Meas.*, Vol. 53, No. 4, August 2004.

[15] P. Arpaia, G. Lucariello, A Zanesco, "Multi-Agent remote predictive diagnosis of dangerous good transports", Proc. of IEEE Instr. and Meas. Tech. Conf., Ottawa, Ontario, Canada, 16-19 May 2005.

[16] P. Arpaia, G. Lucariello, A. Zanesco, "Automatic fault isolation by cultural algorithms with differential influence". Memorie per GMEE 2005, in Italian.

[17] M. Stenberg, R. G. Reynolds, "Using cultural algorithms to support re-engineering of rule-based expert systems in dynamic performance environments: a case study in fraud detection", *IEEE Trans. on Evolutionary Computation*, Vol.1, No. 4, November 1997, pp 225-243.

[18] R. L. Becerra, C. A. Coello Coello, "Culturizing Differential evolution for constrained optimization", Proc. of the Fifth Mexican Int. Conf. in Computer Science, ENC2004.

[19] G. Betta, C. Liguori, A. Pietrosanto, "The use of genetic algorithm for advanced instrument fault detection and isolation schemes", IEEE Proc. Instr. and Meas. Tech. Conf., Bruxelles, Belgium, June 4-6 1996, pp 1129-1134.

[20] V. Guralnik, G. Karypis, "A scalable algorithm for clustering protein sequences," in *Proc. Workshop Data Mining in Bioinformatics (BIOKDD)*, 2001, pp. 73-80.