Using Imprecise and Uncertain Information to Enhance the Diagnosis of a Railway Device

Zohra L. Cherfi, Latifa Oukhellou, Etienne Côme, Thierry Denœux, and Patrice Aknin

Abstract. This paper investigates the use of partially reliable information elicited from multiple experts to improve the diagnosis of a railway infrastructure device. The general statistical model used to perform the diagnosis task is based on a noiseless Independent Factor Analysis handled in a soft-supervised learning framework.

Keywords: Belief function theory, Soft-supervised learning, Independent Factor Analysis, EM algorithm, Fault diagnosis.

1 Introduction

When a pattern recognition approach is adopted to solve diagnosis problems, it involves using machine learning techniques to assign the measured signals to one of several predefined classes of defects. In most real world applications, a large amount of data is available but their labeling is generally a time-consuming and expensive task. However, it can be taken advantage of

Zohra L. Cherfi, Etienne Côme, and Patrice Aknin
UPE, IFSTTAR, GRETTIA, 2 rue de la Butte Verte, 93166 Noisy-le-Grand, France
e-mail: zohra.cherfi@ifsttar.fr, etienne.come@ifsttar.fr, patrice.aknin@ifsttar.fr

Latifa Oukhellou
LISSI, UPE Créteil, 61 av du Gal de Gaulle, 94100 Créteil, France
e-mail: latifa.oukhellou@ifsttar.fr

Thierry Denœux
HEUDIASYC, UTC, UMR CNRS 6599, B.P 20529, 60205 Compiègne, France
e-mail: thierry.denoeux@hds.utc.fr

S. Li (Eds.): Nonlinear Maths for Uncertainty and its Appli., AISC 100, pp. 213–220.
springerlink.com

expert knowledge to label the data. In this case, the class labels can be subject to imprecision and uncertainty. A solution to deal with imprecise and uncertain class labels have been proposed in [3] [6]. In this framework, this paper presents a fault diagnosis application using partially labeled data to learn a statistical model based on Independent Factor Analysis (IFA) [2]. Learning of this statistical model is usually performed in an unsupervised way. The idea investigated in this paper is to incorporate additional information on the class membership of some samples to estimate the parameters of the IFA model, using an extension of the EM algorithm [3] [6].

This paper is organized as follows. Background material on belief functions is first recalled in Sect. 2. Learning the IFA model from data with soft labels is then addressed in Sect. 3. Sect. 4 describes the application under study and introduces the diagnosis problem in greater detail. Experimental results are finally reported in Sect. 5, and Sect. 6 concludes the paper.

2 Background on Belief Functions

This section provides a brief account of the fundamental notions of the Dempster-Shafer theory of belief functions, also referred to as Evidence Theory [4, 9]. A particular interpretation of this theory has been proposed by Smets [11], under the name of the Transferable Belief Model (TBM).

2.1 Belief Representation

Let $\Omega = \{\omega_1, ..., \omega_n\}$ be a finite *frame of discernment*, defined as a set of exclusive and exhaustive hypotheses about some question Q of interest. Partial information about the answer to question Q can be represented by a *mass function* $m : 2^\Omega \rightarrow [0, 1]$ such that $\sum_{A \subseteq \Omega} m(A) = 1$. The quantity $m(A)$ represents a measure of the belief that is assigned to subset $A \subseteq \Omega$ given the available evidence and that cannot be committed to any strict subset of A. Every $A \subseteq \Omega$ such that $m(A) > 0$ is called a *focal set* of m. A mass function or a bba (for *basic belief assignment*) is said to be:

- *normalized* if $\emptyset$ is not a focal set (condition not imposed in the TBM);
- *dogmatic* if Ω is not a focal set;
- *vacuous* if Ω is the only focal set (it then represents total ignorance);
- *simple* if it has at most two focal sets and, if it has two, Ω is one of those;
- *categorical* if it is both simple and dogmatic.

A simple bba such that $m(A) = 1 - w$ for some $A \neq \Omega$ and $m(\Omega) = w$ can be noted A^w. Thus, the vacuous bba can be noted A^1 for any $A \subset \Omega$, and a categorical bba can be noted A^0 for some $A \neq \Omega$.

The information contained in a bba m can be equivalently represented thanks to the *plausibility* function $pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \forall A \subseteq \Omega$. The quantity $pl(A)$ is an upper bound on the degree of support that could be assigned to A if more specific information became available.

2.2 Information Combination

Let m_1 and m_2 be two bbas defined over a common frame of discernment Ω, they may be combined using a suitable operator. The most common ones are the conjunctive and disjunctive rules of combination defined, respectively, as:

$$(m_1 \textcircled{\cap} m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Omega. \tag{1}$$

$$(m_1 \textcircled{\cup} m_2)(A) = \sum_{B \cup C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Omega. \tag{2}$$

The mass assigned to the empty set may be interpreted as a *degree of conflict* between the two sources. An extension of the conjunctive rule proposed by Yager assumes that, in case of conflict, the result is not reliable but the solution must be in Ω. The mass on $\emptyset$ is thus redistributed to Ω which leads to a normalized bba [12].

The conjunctive and disjunctive rules of combination assume the independence of the data sources. In [5], Denœux introduced the cautious rule of combination $\textcircled{\wedge}$ to combine bbas provided by non independent sources. Although the cautious rule can be applied to any non dogmatic bba, it will be recalled here only in the case of *separable* bba, i.e., bbas that can be decomposed as the conjunctive combination of simple bbas [9] [10]. Let m_1 and m_2 be two such bbas. They can be written as $m_1 = \textcircled{\cap}_{A \subset \Omega} A^{w_1(A)}$ and $m_2 = \textcircled{\cap}_{A \subset \Omega} A^{w_2(A)}$, where $A^{w_1(A)}$ and $A^{w_2(A)}$ are simple bbas, $w_1(A) \in (0, 1]$ and $w_2(A) \in (0, 1]$ for all $A \subset \Omega$. Their combination using the cautious rule is defined as:

$$(m_1 \textcircled{\wedge} m_2)(A) = \textcircled{\cap}_{A \subset \Omega} A^{w_1(A) \wedge w_2(A)}, \tag{3}$$

where $\wedge$ denotes the minimum operator. This rule avoids to double-count common evidence when combining non distinct bbas (idempotence property) i.e., it verifies $m \textcircled{\wedge} m = m$ for all m.

3 Statistical Model and Learning Method

3.1 Independent Factor Analysis

IFA is based on a generative model that makes it possible to recover independent latent components (sources) from their observed linear mixtures [2]. In its noiseless formulation (used throughout this paper), the IFA model can be expressed as $\mathbf{y} = H\,\mathbf{z}$, where H is a nonsingular square matrix of size S, $\mathbf{y}$ is the observed random vector whose elements are the S mixtures and $\mathbf{z}$ the random vector whose elements are the S latent components. Each source density is a mixture of Gaussians (MOG), so that a wide class of densities can be approximated. The pdf of $\mathbf{z}$ is thus given by $f^{\mathcal{Z}}(\mathbf{z}) = \prod_{j=1}^{S} \sum_{k=1}^{K^j} \pi_k^j \varphi(z^j; \mu_k^j, \nu_k^j)$, where z^j denotes the j-th component of vector $\mathbf{z}$, $\varphi(.;\mu,\nu)$ denotes the pdf of a Gaussian random variable of mean μ and variance ν; π_k^j, μ_k^j and ν_k^j are the proportion, mean and variance of component k for source j, and K^j is the number of components for source j. In the classical unsupervised setting used in IFA, the problem is to estimate both the mixing matrix H and the MOG parameters from the observed variables $\mathbf{y}$ alone. Maximum likelihood estimation of the model parameters can be achieved by an alternating optimization strategy [1].

3.2 Soft-Supervised Learning in IFA

This section considers the learning of the IFA model in a soft-supervised learning context where partial knowledge of the cluster membership of some samples is available in the form of belief functions. In the general case, we will consider a learning set $\mathbf{M} = \{(\mathbf{y}_1, m_1^1, \ldots, m_1^S), \ldots, (\mathbf{y}_N, m_N^1, \ldots, m_N^S)\}$, where $m_i^1, \ldots, m_i^S$ is a set of bbas encoding uncertain knowledge on the cluster membership of sample $i = 1 \ldots N$ for each one of the S sources. Each bba m_i^j is defined on the frame of discernment $\mathcal{U}^j = \{c_1, \ldots, c_{K^j}\}$ composed of all possible clusters for source j. Let us denote by $\mathbf{x}_i = (\mathbf{y}_i, u_i^1, \ldots, u_i^S)$ the completed data where $\mathbf{y}_i \in \mathbb{R}^S$ are the observed variables and $u_i^j \in \mathcal{U}^j$, $\forall j \in \{1, \ldots, S\}$ are the cluster membership variables which are ill-known. In this model two independence assumptions are made. The random generation process induces the stochastic independence assumption between realizations:

$$f(\mathbf{X}; \boldsymbol{\Psi}) = \prod_{i=1}^{N} f(\mathbf{x}_i; \boldsymbol{\Psi}), \tag{4}$$

where $\boldsymbol{\Psi}$ is the IFA parameter vector, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ is the complete sample vector and $f(\mathbf{x}_i)$ the pdf of a complete observation according to the IFA model. Additionally, the imperfect perception of cluster memberships induces the following cognitive independence assumption (see [9, page 149]):

$$pl(\mathbf{X}) = \prod_{i=1}^{N} pl_i(\mathbf{x}_i) = \prod_{i=1}^{N}\prod_{j=1}^{S} pl_i^j(u_i^j), \tag{5}$$

where $pl(\mathbf{X})$ is the plausibility that the complete sample vector is equal to $\mathbf{X}$, $pl_i(\mathbf{x}_i)$ is the plausibility that the complete data for instance i is $\mathbf{x}_i$ and $pl_i^j(u_i^j)$ is the plausibility that the source j for example i was generated from component u_i^j. Under the assumptions (4) and (5) and following [6], the observed data log likelihood can be written as:

$$l(\boldsymbol{\Psi};\mathbf{M}) = -N\log(|\det(H)|) + \sum_{i=1}^{N}\sum_{j=1}^{S}\log\left(\sum_{k=1}^{K^j} pl_{ik}^j \pi_k^j \varphi((H^{-1}\mathbf{y}_i)^j;\mu_k^j,\nu_k^j)\right) \tag{6}$$

where $pl_{ik}^j = pl_i^j(c_k)$ is the plausibility that sample i belongs to cluster k of latent variable j. This criterion must be maximized with respect to $\boldsymbol{\Psi}$ to compute parameter estimates. An extension of the EM algorithm called E^{2}M for Evidential EM can be used to perform this task [6].

4 Diagnosis Approach

4.1 Problem Description

The track circuit is an essential component of the automatic train control system [8]. Its main function is to detect the presence or absence of vehicle traffic on a given section of the railway track. For this purpose, the railway track is divided into different sections (Fig. 1); each section is equipped with a specific track circuit consisting of: a transmitter connected to one of the two section ends, the two rails that can be considered as a transmission line, a receiver at the other end of the track section and trimming capacitors connected between the two rails at constant spacing to compensate the inductive behavior of the track. A train is detected when the wheels and axles short-circuit the track. It induces the loss of the track circuit signal and the drop of the received signal below a threshold indicates that the section is occupied. The different parts of the track circuit can be subject to malfunctions that must be detected as soon as possible to maintain the system at the required safety and availability levels. In the most extreme cases, an unfortunate attenuation of the transmitted signal may induce important signaling problems (a section can be considered as occupied even if it is not). The objective of diagnosis is to avoid such inconvenience on the basis of inspection signals analysis [8]. For this purpose, an inspection vehicle is able to deliver a measurement signal (denoted as Icc) linked to electrical characteristics of the system (Fig. 1). This paper describes the approach adopted for the diagnosis of track circuit from real inspection signals, it will focus on trimming capacitor faults.

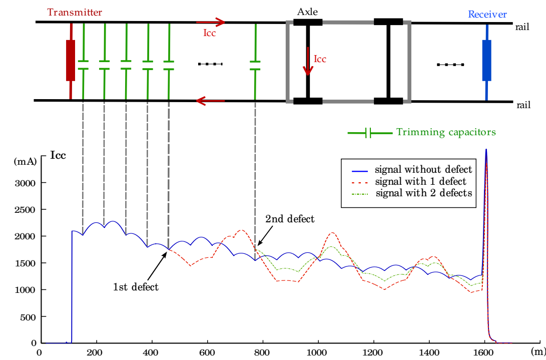

Fig. 1 Track circuit representation and examples of inspection signals (Icc) simulated along a 1500 m track circuit: one of them corresponds to a fault-free system, while the others correspond to a signal with one defective capacitor, and respectively a signal with two defective capacitors.

4.2 Diagnosis Methodology

A track circuit can be considered as a complex system made up of a series of S spatially related subsystems, where each subsystem correspond to a trimming capacitor. A defect on one subsystem can be represented either by its capacitance or by a discrete value if considering a finite number of operating modes. The approach adopted here consists in extracting features from the measurement signal, and building a generative model as shown in Fig. 2. In this model, the variables y_i^j are observed variables extracted by approximating each arch of the inspection signal (Icc) by a quadratic polynomial. The variables z_i^j are continuous latent variables corresponding to continuous values describing the subsystem defects (capacitances), while the discrete latent variables u_i^j correspond to the membership of the subsystem operating mode to one of the following three states: fault-free, medium defect, major defect. Assuming that a linear relationship exists between observed and latent variables and that each latent variable can be modeled semi-parametrically by a MOG, the involved generative model can be considered as an IFA model [2].

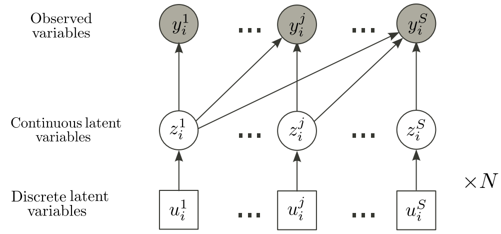

Fig. 2 Generative model for the diagnosis of track circuits represented by a graphical model

5 Results and Discussion

The diagnosis system was assessed using a real data set consisting of 422 inspection signals provided by the French National Railway Company (SNCF). A labeling campaign was organized with the aim of presenting separately the signals to four experts for labeling. Three classes were considered for the labeling operation, corresponding to the three operating modes of the subsystems. Experts were allowed to specify imprecise and uncertain labels that were represented as belief functions. The labels from individual experts were combined using each of the rules described in Sect. 2.2, and the IFA model was fit using the combined labels thanks to the E^2M algorithm [6].

Because of the lack of the ground truth about the real state of the components under diagnosis, we chose to take as reference another labeling of the same database obtained thanks to a third-party expertise. This particular labeling was achieved simultaneously by three observers and in favorable conditions to provide a reference tool. This reference labeling will be referred to as REF in the following. The classification results reported in Table 1 reveal a good classification performance despite some confusions between contiguous classes. The confusion matrices corresponding to individual experts provide some information on expert skills. Indeed, experts 1 and 4 seem to better detect major defects, while experts 2 and 3 are more accurate for the detection of medium defects. The combination of expert opinions makes it possible to improve the detection of both types of defects. The best results were achieved by the cautious rule, which suggests that the expert opinions cannot be regarded as independent. The confusion between contiguous classes can be explained by the fact that identification of medium defects is a difficult exercise due to the continuous nature of the real states.

Table 1 Confusion matrices between true classes ω_0, ω_1 and ω_2 (defined by the REF labeling) and their estimates d_0, d_1 and d_2 (computed on ten cross validation test sets)

	ω_0	ω_1	ω_2
d_0	98.8	33.1	2.1
d_1	0.9	51.1	6.9
d_2	0.2	15.8	90.9

(Expert 1)

	ω_0	ω_1	ω_2
d_0	98.9	34.7	3.0
d_1	0.8	58.8	12.2
d_2	0.3	6.5	84.7

(Expert 2)

	ω_0	ω_1	ω_2
d_0	98.7	22.1	2.1
d_1	1.1	63.6	13.8
d_2	0.2	14.3	84.1

(Expert 3)

	ω_0	ω_1	ω_2
d_0	98.8	34.6	3.3
d_1	1.0	49.6	5.8
d_2	0.2	15.8	90.9

(Expert 4)

	ω_0	ω_1	ω_2
d_0	98.9	30.7	2.9
d_1	0.9	58.0	7.7
d_2	0.2	11.3	89.4

(Disjonctive rule)

	ω_0	ω_1	ω_2
d_0	98.9	20.2	2.9
d_1	1.0	64.2	6.5
d_2	0.1	15.6	90.6

(Conjonctive rule)

	ω_0	ω_1	ω_2
d_0	98.9	20.4	2.6
d_1	1.0	65.3	4.9
d_2	0.1	14.2	92.4

(Cautious rule)

6 Conclusions

The particular application that was considered concerns the diagnosis of railway track circuits. Experiments were carried out with real signals labeled by four different human experts. Experts' uncertain knowledge about the state of each subsystem was encoded as belief functions, which were pooled using different combination rules. These combined opinions were shown to yield better classification results than those obtained from each individual expert, especially with the cautious rule of combination [5], which can be explained by the existence of common knowledge shared among the experts.

This work can be extended in several directions. The approach relies on expert knowledge elicitation in the belief function framework, an important problem that has not received much attention until now. More sophisticated combination schemes could also be considered: for instance, discount rates could be learned from the data to take into account the competence of each individual expert [7].

References

1. Amari, S., Cichocki, A., Yang, H.: A New Learning Algorithm for Blind Signal Separation. In: Proceedings of NIPS. MIT Press, Cambridge (1996)
2. Attias, H.: Independent factor analysis. Neural Computation 11(4), 803–851 (1999)
3. Côme, E., Oukhellou, L., Denoeux, T., Aknin, P.: Learning from partially supervised data using mixture models and belief functions. Pattern Recognit. 42(3), 334–348 (2009)
4. Dempster, A.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. 38, 325–339 (1967)
5. Denoeux, T.: Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence. Artificial Intelligence 172, 234–264 (2008)
6. Denœux, T.: Maximum likelihood from evidential data: An extension of the EM algorithm. In: Borgelt, C., González-Rodríguez, G., Trutschnig, W., Lubiano, M.A., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O., et al. (eds.) Combining Soft Computing and Statistical Methods in Data Analysis. AISC, vol. 77, pp. 181–188. Springer, Heidelberg (2010)
7. Elouedi, Z., Mellouli, K., Smets, P.: Assessing sensor reliability for multisensor data fusion within the Transferable Belief Model. IEEE Trans. on Syst. Man and Cybern. B. 34, 782–787 (2004)
8. Oukhellou, L., Debiolles, A., Denoeux, T., Aknin, P.: Fault diagnosis in railway track circuits using Dempster-Shafer classifier fusion. Eng. Appl. of Artif. Intell. 23(1), 117–128 (2010)
9. Shafer, G.: A mathematical theory of evidence. Princeton Univ. Press, Princeton (1976)
10. Smets, P.: The canonical decomposition of a weighted belief. In: Int. Joint Conf. on Artif. Intell. Morgan Kaufmann, San Francisco (1995)
11. Smets, P., Kennes, R.: The Transferable Belief Model. Artif. Intell. 66, 191–234 (1994)
12. Yager, R.: On the Dempster-Shafer framework and new combination rules. Information Sciences 41(2), 93–137 (1987)