Information Sciences 180 (2010) 5060–5076

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Piecewise-linear approximation of non-linear models based on probabilistically/possibilistically interpreted intervals' numbers (INs)

S.E. Papadakis *, Vassilis G. Kaburlasos

Department of Industrial Informatics, Technological Educational Institution of Kavala, 65404 Kavala, Greece

ARTICLE INFO

Article history:
Received 6 May 2009
Received in revised form 10 November 2009
Accepted 14 March 2010

Keywords:
Fuzzy inference systems (FIS)
Genetic optimization
Granular data
Intervals' number (IN)
Lattice theory
Linear approximation
Rules
Self-organizing map (SOM)
Similarity measure
Structure identification
TSK model

ABSTRACT

Linear models are preferable due to simplicity. Nevertheless, non-linear models often emerge in practice. A popular approach for modeling nonlinearities is by piecewise-linear approximation. Inspired from fuzzy inference systems (FISs) of Tagaki–Sugeno–Kang (TSK) type as well as from Kohonen's self-organizing map (KSOM) this work introduces a genetically optimized synergy based on intervals' numbers, or INs for short. The latter (INs) are interpreted here either probabilistically or possibilistically. The employment of mathematical lattice theory is instrumental. Advantages include accommodation of *granular* data, introduction of *tunable* nonlinearities, and induction of descriptive decision-making knowledge (rules) from the data. Both efficiency and effectiveness are demonstrated in three benchmark problems. The proposed computational method demonstrates invariably a better capacity for generalization; moreover, it learns orders-of-magnitude faster than alternative methods inducing clearly fewer rules.

1. Introduction

The need to induce, efficiently, an effective model (real function) $y : \mathsf{R}^N \rightarrow \mathsf{R}^M$ arises frequently in practical applications. In particular, linear models $y(\mathbf{x}) = c_0 + c_1x_1 + c_2x_2 + \cdots + c_Nx_N$ are preferable due to simplicity. However, most often, the dependence of a system output y on the input variables $x_1, \ldots, x_N$ is non-linear.

One way of modeling nonlinearities is by piecewise-linear approximation. For instance, in the context of fuzzy sets and systems, the *TSK (Tagaki–Sugeno–Kang) fuzzy model*, described by Sugeno and Kang [49], Sugeno and Tanaka [50], Sugeno and Yasukawa [51], Takagi and Sugeno [54], combines linguistic (fuzzy) interpretations of its numeric inputs with a (locally, within a cluster) linear computation of an output in order to achieve a non-linear input-to-output map. Recent TSK modeling applications have been reported also by Rezaee and Fazel Zarandi [44], Zhou and Gan [71], etc. For the reader's convenience, the operation of a TSK model is summarized in the Appendix A.

Critical for the computation of a TSK model is the computation of input data clusters. A popular clustering scheme is *Kohonen's self-organizing map (KSOM)* introduced by Kohonen [33], mainly for visualization of non-linear relations of multi-dimensional data. Er et al. [12] have confirmed the capacity of KSOM for rapid data processing. Pascual-Marqui et al. [42] have reported a soft (fuzzy) KSOM synergy with conventional fuzzy c-means, where the code vectors are distributed on a regular low-dimensional grid. Moreover, Vuorimaa [61] has introduced a fuzzy extension of KSOM for function $f : \mathsf{R}^N \rightarrow \mathsf{R}$

* Corresponding author. Fax: +30 2510462321.
E-mail addresses: spap@teikav.edu.gr (S.E. Papadakis), vgkabs@teikav.edu.gr (V.G. Kaburlasos).

0020-0255/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.03.023

approximation using triangular fuzzy membership functions, exclusively. Lately, Kaburlasos and Papadakis [23] have proposed *granular* (fuzzy) extensions of KSOM in classification applications.

This work introduces a synergy of TSK- with KSOM-inspired techniques towards an *efficient* as well as *effective* piecewise-linear approximation of non-linear models as explained below. The proposed synergy builds on an established mathematical result, namely the "resolution identity theorem", presented by Zadeh [68], which specifies that a fuzzy set can (equivalently) be represented either by its membership function or by its α-cuts.

Note that even though a fuzzy set can be defined on any universe of discourse, in practice, the *real numbers* universe of discourse R is preferred as pointed out by Kaburlasos and Kehagias [22]. More specifically, *fuzzy numbers* are typically employed, for instance in fuzzy inference systems (FISs). Recall that a fuzzy number is defined as a convex, normal fuzzy set, often with bounded support. A fuzzy number is defined on R with an *upper semicontinuous* membership function as described in Kaburlasos [19], Vroman et al. [60].

It turns out that a α-cut of a fuzzy number is an interval; hence, based on the aforementioned "resolution identity theorem", a fuzzy number can be represented by a set of intervals. In conclusion, Uehara and Fujise [56], Uehara and Hirota [57], Uehara et al. [58] have proposed a novel FIS design in practical applications based on α-cuts (intervals) of fuzzy numbers – Advantages include faster (parallel) data processing "level-by-level", "orders-of-magnitude" smaller computer memory requirements, etc. Senturk and Erginel [46] have employed α-cuts for enhancing traditional control strategies. Furthermore, Cornelis et al. [9], Nachtegael and Kerre [36] have considered α-cuts/intervals for fuzzy logic/morphology operations in theoretical studies involving ambiguity.

This work builds creatively on the "resolution identity theorem" by, first, considering the equivalent α-cuts (interval) representation for a fuzzy number and, second, by dropping the corresponding possibilistic interpretation. Hence, an *intervals' number (IN)* emerges as a mathematical object, which may admit either a *possibilistic* or a *probabilistic* interpretation as explained below. Advantages include an introduction of useful linear operations, tunable nonlinearities, a capacity to deal with granular data, etc. Instrumental for IN-based analysis and design is (mathematical) *lattice theory (LT)* because the set of (closed) intervals on the real line is partially (lattice)-ordered. For the reader's interest, the emergence of LT in information processing is outlined next.

Mathematical lattices have emerged in the first half of the nineteenth century as a spin off of work on formalizing propositional logic. During the next one hundred years LT was established, and compiled creatively by Garrett Birkhoff [4]. Currently, there is a number of research communities that employ LT in various information processing domains including, first, *Logic and Reasoning* for automated decision-making (see in Xu et al. [66]), second, *mathematical morphology* for signal/image processing (see in Ritter and Wilson [45]), third, *formal concept analysis* for knowledge-representation and information-retrieval (see in Ganter and Wille [15]), fourth, *computational intelligence* for clustering, classification, and regression applications (see in Kaburlasos [20]), etc.

There are two different approaches for employing LT in practice. The first approach, namely *order-based*, is based on semantics represented by the lattice (partial)-order as demonstrated also by Bloch et al. [5], Ganter and Wille [15], Lai and Xu [34]. The second approach, namely *algebra-based*, is based on the lattice (algebraic)-operations of meet (∧) and join (∨) as demonstrated also by Graña et al. [17], Ritter and Wilson [45], Soille [48], Valle and Sussner [59]. Various combinations of the aforementioned two approaches have also been reported, for instance in classification applications by da Silva and Sussner [10], Kaburlasos [20], Sussner and Esmi [52,53]. In this work, we describe a novel combination of the aforementioned two approaches.

Previous work by Kaburlasos [19,20], Kaburlasos and Kehagias [22], Kaburlasos and Papadakis [23,25], has employed the term *fuzzy interval number (FIN)* instead of the term *intervals' number (IN)*, because it stressed a fuzzy interpretation. Recently, Kaburlasos and Papadakis [24] have switched to the term IN, including also an improved mathematical notation. Likewise, the term "CALFIN", proposed previously for an algorithm which induces a "FIN" from a population of measurements, is eloquently replaced here by the term "CALCIN".

This paper presents significant enhancements over the preliminary work by Kaburlasos and Papadakis in [24] as follows. First, we introduce a novel similarity measure function ($\mu_\wedge$). Second, we detail structure/parameter identification algorithms based on $\mu_\wedge$ rather than on metric d_p, the latter was employed in [24]; here, we also compute the corresponding algorithm complexity. Third, we demonstrate an employment of a IN as either a probability- or a possibility-distribution. Fourth, we demonstrate two additional benchmark problems including improved experimental results; moreover, in all benchmark problems, we display rules induced. Fifth, we discuss novel theoretical perspectives. Sixth, we cite a large number of additional references including comparative discussions.

This paper is organized as follows. Section 2 summarizes the mathematical background. Section 3 presents a novel structure identification. Section 4 describes a novel parameter identification. Section 5 details, comparatively, experimental results. Section 6 concludes by summarizing our contribution including also future work. The Appendix A includes the proof of a proposition as well as two computational algorithms used in the experiments.

2. Mathematical background

This section summarizes useful mathematical results and tools introduced by Kaburlasos [20], Kaburlasos and Kehagias [21,22], Kaburlasos and Papadakis [23–25], Kaburlasos et al. [27]. Mathematical lattice theory here is instrumental.

Recall from Birkhoff [4] that given a set P, a binary relation ($\leqslant$) on P is called *partial order* if and only if it satisfies the following three conditions: "$x \leqslant x$" (*Reflexivity*), "$x \leqslant y$ and $y \leqslant x \Rightarrow x = y$" (*Antisymmetry*), and "$x \leqslant y$ and $y \leqslant z \Rightarrow x \leqslant z$"