Engineering Applications of Artificial Intelligence 35 (2014) 277–285

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

gsaINknn: A GSA optimized, lattice computing knn classifier

Yazdan Jamshidi [a], Vassilis G. Kaburlasos [b,*]

[a] Department of Computer Engineering, Kermanshah Science and Research Branch, Islamic Azad University, Kermanshah, Iran
[b] Human–Machines Interaction (HMI) Lab, Department of Computer & Informatics Engineering, Eastern Macedonia and Thrace Institute of Technology, Kavala 65404, Greece

ARTICLE INFO

Article history:
Received 15 March 2013
Received in revised form
19 April 2014
Accepted 22 June 2014
Available online 29 July 2014

Keywords:
Computing with words (CWW)
Classification
Gravitational search algorithm (GSA)
Intervals' number (IN)
k-Nearest neighbor (knn)
Lattice computing (LC)

ABSTRACT

This work proposes an effective synergy of the Intervals' Number k-nearest neighbor (INknn) classifier, that is a granular extension of the conventional knn classifier in the metric lattice of Intervals' Numbers (INs), with the gravitational search algorithm (GSA) for stochastic search and optimization. Hence, the gsaINknn classifier emerges whose effectiveness is demonstrated here on 12 benchmark classification datasets. The experimental results show that the gsaINknn classifier compares favorably with alternative classifiers from the literature. The far-reaching potential of the gsaINknn classifier in computing with words is also delineated.

1. Introduction

This work introduces a stochastically optimized, granular k-nearest neighbor (knn) classifier based on INs, where IN stands for Intervals' Number.

An IN is a mathematical object, which may represent either a fuzzy interval or a probability distribution (Papadakis and Kaburlasos, 2010). In any case, an IN can be interpreted as an information granule. INs have been studied in a series of publications. In particular, as explained in Kaburlasos et al. (2013a), it has been shown that the set F_1 of INs is a metric lattice with cardinality $\aleph_1$, where $\aleph_1$ is the cardinality of the set R of real numbers; moreover, F_1 is a cone in a linear space. Note that previous work (Kaburlasos, 2004; Kaburlasos, 2006) has employed the term FIN (i.e., fuzzy interval number) instead of the term IN because it stressed a fuzzy interpretation. Likewise, the term CALFIN, proposed previously for an algorithm which induces a FIN from a population of measurements, was later replaced by the term CALCIN (Papadakis and Kaburlasos, 2010). Recall that an IN computed by algorithm CALCIN retains all-order data statistics (Kaburlasos et al., 2013a). In the aforementioned context, the capacity as well as the rich potential of INs, especially in industrial applications, has been demonstrated (Kaburlasos and Kehagias, 2014; Kaburlasos and Pachidis, 2014; Papadakis and Kaburlasos, 2010).

INs have been used in an array of computational intelligence applications regarding clustering, classification and regression (Kaburlasos and Moussiades, 2014; Kaburlasos and Pachidis, 2014; Kaburlasos and Papadakis, 2006; Kaburlasos et al., 2012, 2013a; Papadakis and Kaburlasos, 2010; Papadakis et al., 2014). There is experimental evidence that a parametric, IN-based scheme can be optimized toward clearly improving performance. More specifically, optimization has been pursued by stochastic search techniques including genetic algorithms (Kaburlasos and Papadakis, 2006; Kaburlasos et al., 2012, 2013a; Papadakis and Kaburlasos, 2010) and particle swarm optimization (Papadakis et al., 2014) since, currently, there are no analytic optimization methods available in the lattice of INs.

Previous works have frequently employed an inclusion measure (σ) function as an instrument for decision making in the lattice of INs (Kaburlasos and Pachidis, 2014; Kaburlasos et al., 2012, 2013a; Papadakis et al., 2014). The interest of this work is in (fuzzy) nearest neighbor classification (Derrac et al., 2014). Note that, lately, a number of knn classifiers based on INs, namely INknn classifiers, have been introduced (Kaburlasos et al., 2014; Pachidis and Kaburlasos, 2012; Tsoukalas et al., 2013); nevertheless, none of the latter classifiers was optimized. On the grounds of compelling evidence, as explained above, this work proposes an optimized INknn classifier toward improving performance. We remark that various heuristic optimization methods have been proposed in machine learning including Simulated Annealing (SA) (Chang et al., 1994), Ant Colony (Chi and Yang, 2006), Particle Swarm Optimization (PSO) algorithms (Huang and Dun, 2008; Lotfi Shahreza et al., 2011), Differential Evolution (DE) (Liu and Sun, 2011; Si et al., 2012), Genetic Algorithms (GAs) (Polat and Yildirim, 2008), etc. Lately, Rashedi and colleagues have proposed the

* Corresponding author.
E-mail address: vgkabs@teikav.edu.gr (V.G. Kaburlasos).

http://dx.doi.org/10.1016/j.engappai.2014.06.018
0952-1976/

gravitational search algorithm (GSA) (Rashedi et al., 2009), that is a swarm based meta-heuristic search algorithm based on the Newtonian laws of gravity. The GSA has already been successfully applied to numerous problems (Bahrololoum et al., 2012; Li and Zhou, 2011; Rashedi et al., 2011; Rebollo-Ruiz and Graña, 2012, 2013; Rebollo et al., 2012; Sarafrazi and Nezamabadi-pour, 2013; Taghipour et al., 2010; Yin et al., 2011). This work proposes a synergy of GSA with the INknn classifier toward improving the capacity of the latter classifier. Hence, the (granular) gsaINknn classifier emerges whose capacity is demonstrated here comparatively on 12 benchmark datasets.

In a more general context, the proposed gsaINknn classifier is a scheme of the emerging lattice computing (LC) paradigm. Note that LC was originally defined as "the collection of Computational Intelligence tools and techniques that either make use of lattice operators inf and sup for the construction of the computational algorithms or exploit lattice theory for language representation and reasoning" (Graña, 2009). Recent work has extended the meaning of LC as "an evolving collection of tools and methodologies that process lattice ordered data per se including logic values, numbers, sets, symbols, graphs, etc." (Kaburlasos and Kehagias, 2014; Kaburlasos et al., 2013a, 2013b). The LC paradigm provides instruments for granular computing, where uncertainty/ambiguity is accommodated in partially/lattice-ordered information granules (Jamshidi and Nezamabadi-pour, 2013; Kaburlasos, 2010; Liu et al., 2013; Sussner and Esmi, 2011).

A number of LC models have already been proposed in the context of mathematical morphology. For instance, morphological neural networks (MNN) including both morphological perceptrons and fuzzy morphological associative memories (FMAMs) (Sussner and Esmi, 2009, 2011; Sussner and Valle, 2006; Valle and Grande Vicente, 2012) can be classified as LC models. In particular, Sussner and colleagues have employed a FMAM to implement a fuzzy inference system based on the complete lattice structure of the class of fuzzy sets (Sussner and Valle, 2006; Valle and Sussner, 2008, 2011). Furthermore, Graña and colleagues have applied LC techniques to image analysis applications of mathematical morphology (Graña et al., 2011, 2010, 2009). Of particular interest in LC is the notion of a *fuzzy lattice*, which has been proposed by Nanda toward fuzzifying a partial order relation (Nanda, 1989). Working independently Kaburlasos and colleagues, inspired from the adaptive resonance theory (ART) for neural computation (Carpenter et al., 1991, 1992), have proposed a number of *fuzzy lattice neural networks* for clustering and classification (Kaburlasos, 2006) operating on fuzzy lattice reasoning (FLR) principles. The FLR classifier was introduced in Kaburlasos et al. (2007) for inducing descriptive decision-making knowledge (rules) in a mathematical lattice data domain, including the space R^N as a special case; moreover, the FLR classifier has been successfully applied to a variety of problems such as ambient ozone estimation as well as air quality assessment (Athanasiadis and Kaburlasos, 2006). Recent trends in lattice computing appear in Graña (2012), Kaburlasos (2011), Kaburlasos and Ritter (2007).

The layout of this paper is as follows. Section 2 outlines the mathematical background. Section 3 presents the INknn classifier including an explanatory application example. Section 4 describes the GSA optimization algorithm. Section 5 details the gsaINknn classifier. Section 6 presents comparatively experimental results regarding 12 benchmark classification datasets. Finally, Section 7 concludes by both summarizing our contribution and delineating future work.

2. Mathematical background

This section outlines general lattice notions followed by a hierarchy of lattices ending up to Intervals' Numbers, or INs for short.

2.1. General lattices

A set P with a *partial order* (binary) relation $\sqsubseteq$ is called *partially ordered set* or *poset* for short, symbolically $(P,\sqsubseteq)$ (Birkhoff, 1967; Kaburlasos, 2006; Kaburlasos and Kehagias, 2014; Kaburlasos and Pachidis, 2014; Rutherford, 1965). A function $\varphi: P\to Q$ from a poset $(P,\sqsubseteq)$ to a poset $(Q,\sqsubseteq)$ is called *isomorphic* iff $x\sqsubseteq y \Leftrightarrow \varphi(x)\sqsubseteq\varphi(y)$. It is well known that the inverse $\sqsupseteq$, namely *dual* (order), of an order relation $\sqsubseteq$ is itself an order relation. A *lattice* $(\mathrm{L},\sqsubseteq)$ is a poset with the additional property that any two of its elements $a,b\in\mathrm{L}$ have both an *infimum* denoted by $a\sqcap b=\inf\{a,b\}$ and a *supremum* denoted by $a\sqcup b=\sup\{a,b\}$. The lattice operations $\sqcap$ and $\sqcup$ are called *meet* and *join*, respectively. A lattice $(\mathrm{L},\sqsubseteq)$ is called *complete* when each of its subsets has a supremum as well as an infimum in L. A non-void complete lattice has both a least element and a greatest element denoted by o and i, respectively. A lattice $(\mathrm{L},\sqsubseteq)$ is called *totally-ordered* iff for $a,b\in\mathrm{L}$ it is either $a\sqsupseteq b$ or $a\sqsubset b$. In this work we use "square symbols" such as $\sqcup$, $\sqcap$ and $\sqsubseteq$ with general lattice elements, "straight symbols" $\vee$, $\wedge$ and $\leq$ with real numbers, and symbols $\cup$, $\cap$ and $\subseteq$ with sets.

An *aggregate* lattice $(\mathrm{L},\sqsubseteq)$ is the Cartesian product of N *component* lattices $\mathrm{L}_1,\ldots,\mathrm{L}_N$; i.e. $(\mathrm{L},\sqsubseteq)=(\mathrm{L}_1,\sqsubseteq_1)\times\ldots\times(\mathrm{L}_N,\sqsubseteq_N)$. The product lattice L operations join and meet are defined as

$(a_1,\ldots,a_N)\sqcup(b_1,\ldots,b_N)=(a_1\sqcup_1 b_1,\ldots,a_N\sqcup_1 b_N)$ and
$(a_1,\ldots,a_N)\sqcap(b_1,\ldots,b_N)=(a_1\sqcap_N b_1,\ldots,a_N\sqcap_N b_N)$

A *valuation* on a lattice $(\mathrm{L},\sqsubseteq)$ is a real function v: $\mathrm{L}\to\mathrm{R}$ which satisfies $v(a)+v(b)=v(a\sqcap b)+v(a\sqcup b)$. A valuation v is called *positive* iff $a\sqsubset b$ implies $v(a)<v(b)$. A positive valuation function v: $\mathrm{L}\to\mathrm{R}$ implies a metric function d: $\mathrm{L}\times\mathrm{L}\to\mathrm{R}_0^+$ given by $d(x,y)=v(x\sqcup y)-v(x\sqcap y)$.

Generalized interval is an element of the product lattice $(\mathrm{L},\sqsupseteq)\times(\mathrm{L},\sqsubseteq)=(\mathrm{L}\times\mathrm{L},\sqsupseteq\times\sqsubseteq)$. The latter lattice may simply be denoted by $(\Delta,\sqsubseteq)$. A generalized interval is denoted by $[a,b]$. The ordering $(\sqsubseteq)$, join $(\sqcup)$ and meet $(\sqcap)$ operations in lattice $(\Delta,\sqsubseteq)$ are given as follows:

$[a,b]\sqsubseteq[c,d]\Leftrightarrow(c\sqsubseteq a \text{ and } b\sqsubseteq d)$, $[a,b]\sqcup[c,d]=[a\sqcap c, b\sqcup d]$, and
$[a,b]\sqcap[c,d]=[a\sqcup c, b\sqcap d]$

Here, we are interested in a *dual isomorphic* function θ: $\mathrm{L}\to\mathrm{L}$ on a general lattice $(\mathrm{L},\sqsubseteq)$ such that $x\sqsubset y \Leftrightarrow \varphi(x)\sqsupset\varphi(y)$. Based on both a positive valuation function v: $\mathrm{L}\to\mathrm{R}$ and a dual isomorphic function θ: $\mathrm{L}\to\mathrm{L}$ on a general lattice $(\mathrm{L},\sqsubseteq)$, a positive valuation function v_Δ is defined on lattice $(\Delta,\sqsubseteq)$ as follows: $v_\Delta([a,b])=v(\theta(a))+v(b)$.

2.2. A hierarchy of complete lattices

Next, we constructively develop a hierarchy of lattices from a *reference set* $\mathrm{L}\subseteq\overline{\mathrm{R}}$, where $\overline{\mathrm{R}}=\mathrm{R}\cup\{-\infty,+\infty\}$ is the totally-ordered set of *extended real numbers*. In particular, we choose L so that $(\mathrm{L},\leq)$ is a complete lattice. For example, L can be $\overline{\mathrm{R}}$ itself or it might be an interval $[a,b]\subset\overline{\mathrm{R}}$. In particular, for $\mathrm{L}=\overline{\mathrm{R}}$ it is $o=-\infty$ and $i=+\infty$, whereas for $\mathrm{L}=[a,b]$ it is $o=a$ and $i=b$.

Any strictly increasing real function v: $\mathrm{L}\to\mathrm{R}$ is a positive valuation on lattice $(\mathrm{L},\leq)$. Moreover, any strictly decreasing function θ: $\mathrm{L}\to\mathrm{L}$ is a dual isomorphic function on $(\mathrm{L},\leq)$. Note that choosing a suitable valuation function is problem dependent (Jamshidi Khezeli and Nezamabadi-pour, 2012; Kaburlasos et al., 2007; Liu et al., 2011).

Consider the lattice $(\Delta,\sqsubseteq)$ of generalized intervals stemming from lattice $(\mathrm{L},\leq)$. A metric distance function $d_\Delta:\Delta\times\Delta\to\mathrm{R}_0^+$ is defined on $(\Delta,\sqsubseteq)$ as follows:

$$\begin{aligned}d_\Delta([a,b],[c,d])&=v_\Delta([a,b]\sqcup[c,d])-v_\Delta([a,b]\sqcap[c,d])\\&=v_\Delta([a\wedge c,b\vee d])-v_\Delta([a\vee c,b\wedge d])\end{aligned}\quad(1)$$

We define the set of *conventional intervals* as $\mathrm{J(L)}=\{[a,b]: a,b\in\mathrm{L}$ and $a\sqsubseteq b\}$. Augmenting J(L) by the empty interval, denoted by O,