Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems ••• (••••) •••–•••
FUZZY sets and systems
www.elsevier.com/locate/fss

Penalty-based aggregation of multidimensional data

Marek Gagolewski [a,b,*]

[a] *Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland*
[b] *Faculty of Mathematics and Information Science, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland*

Received 4 January 2016; received in revised form 9 December 2016; accepted 10 December 2016

Abstract

Research in aggregation theory is nowadays still mostly focused on algorithms to summarize tuples consisting of observations in some real interval or of diverse general ordered structures. Of course, in practice of information processing many other data types between these two extreme cases are worth inspecting. This contribution deals with the aggregation of lists of data points in $\mathbb{R}^d$ for arbitrary $d \geq 1$. Even though particular functions aiming to summarize multidimensional data have been discussed by researchers in data analysis, computational statistics and geometry, there is clearly a need to provide a comprehensive and unified model in which their properties like equivariances to geometric transformations, internality, and monotonicity may be studied at an appropriate level of generality. The proposed penalty-based approach serves as a common framework for all idempotent information aggregation methods, including componentwise functions, pairwise distance minimizers, and data depth-based medians. It also allows for deriving many new practically useful tools.

Keywords: Multidimensional data aggregation; Penalty functions; Data depth; Centroid; Median

1. Introduction

Aggregation theory [5,9,32,38] focuses on a formal analysis of functions that, given a set of objects of the same kind, output a single item which is (in some sense) representative of all the inputs. Till very recently, functions like $\mathsf{F} : \mathbb{I}^n \to \mathbb{I}$ for some $\mathbb{I} = [a, b]$ and $n \geq 2$, fulfilling key application-specific properties like nondecreasingness, internality, conjunctivity, etc., were the most common objects of interest in this field. It is important to stress that particular aggregation functions were of course known long before aggregation theory became a genuine branch of applied mathematics and information science. In particular, *means* or *averaging functions* [5,11] – that is functions that are at least idempotent – include the famous arithmetic mean, median, quasi-arithmetic means, and OWA [74] operators, to name just a few. However, the emergence of this domain allowed to seek common patterns bracketing diverse ways to handle information overload as well as understand data aggregation processes much better. Apart from aggregation on bounded posets, e.g., [21,44,56], from quite recently, we finally start to observe a growing interest in

* Correspondence to: Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland.
E-mail address: gagolews@ibspan.waw.pl.

http://dx.doi.org/10.1016/j.fss.2016.12.009
0165-0114/© 2016 Elsevier B.V. All rights reserved.

aggregation *of* and *on* other practically useful structures, see, e.g., the papers [58,60] concerning the problem of combining rankings.

In this paper we are interested in functions that aim to aggregate a sequence of n numeric lists $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{R}^d$ for a fixed $d \geq 1$. Each such $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ can be written as:

$$\mathsf{F}\left(\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_d^{(1)} \end{bmatrix}, \ldots, \begin{bmatrix} x_1^{(n)} \\ x_2^{(n)} \\ \vdots \\ x_d^{(n)} \end{bmatrix}\right) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}. \tag{1}$$

Equivalently, we may conceive F as a function acting on a $d \times n$ matrix like:

$$\mathbf{X} = [\mathbf{x}^{(1)}\ \mathbf{x}^{(2)}\ \cdots\ \mathbf{x}^{(n)}] = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_d \end{bmatrix}.$$

Note that in data analysis, $\mathbf{x}^{(i)} \in \mathbb{R}^d = \mathbb{R}^{d\times 1}$ is typically called an observation – it designates an object or experimental unit (e.g., a person, autonomous vehicle, spatial location). On the other hand, $\mathbf{x}_j \in \mathbb{R}^{1\times n}$ denotes the j-th variable or feature (such as temperature, weight, or velocity).

The concept of a penalty function based on data in the real line, i.e., with $d = 1$, was first introduced by Yager [73], and then extended in numerous works, see, e.g., [8,15,16,75]; in particular, [13] gives the most recent summary together with a critical historical overview. Its aim is to measure the amount of "disagreement" between the inputs and the output being computed. Such a framework provides a very appealing way to define new *means*: it can be shown that minimizers of some penalty coincide with the class of idempotent functions [13].

Our aim here is to propose a penalty-based framework for idempotent functions that act on observations in $\mathbb{R}^d$ for an arbitrary $d \geq 1$, significantly extending our previous preliminary studies [32,33]. In the next section the notion of a penalty function generalizing the classical one is proposed. Basic desired properties of penalty-based mappings are discussed in Sect. 3. In Sects. 4, 5, and 6 we present three noteworthy classes of aggregation methods, respectively: componentwise extensions of unidimensional functions, those constructed upon pairwise distances between observations, and those defined by means of the notion of data depth. Section 7 concludes the paper and discusses some concepts which are frequently considered in classical aggregation theory, but in the extended setting are difficult to maintain. This includes, among others, the notion of monotonicity and orness measures.

2. Penalty-based framework

Given an arbitrary $x \in \mathbb{R}$, with $(d * x)$ we denote a d-tuple $(x, x, \ldots, x) \in \mathbb{R}^d$. Binary operations like $+, -, \cdot, /$, $\wedge$ (minimum), and $\vee$ (maximum) on vectors of equal lengths d are applied elementwise and thus output a vector of length d too. On the other hand, if one of the operands is a scalar, then it is extended to a vector of length d in such a way that, e.g., $\mathbf{x} + t = \mathbf{x} + (d * t)$. If $\mathbf{A} \in \mathbb{R}^{d\times n}$ is a matrix with d rows and n columns and $\mathbf{t} \in \mathbb{R}^d$, then by, e.g., $\mathbf{A} + \mathbf{t}$ we mean $\mathbf{A} + [\mathbf{t}\,\mathbf{t}\ \cdots\ \mathbf{t}]$, i.e., $\mathbf{t}$ is treated as a column vector. Moreover, $\mathbf{A} + t = \mathbf{A} + (d * t) = \mathbf{A} + [(d * t)\ \cdots\ (d * t)]$. Finally, for some $n \in \mathbb{N}$ we denote with $[n]$ the set $\{1, 2, \ldots, n\}$ and with $\|\cdot\|_p$ the L^p-norm on $\mathbb{R}^d$.

Having established the notation convention used throughout the paper, let us introduce the notion of a penalty function relative to given $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{R}^d$.

Definition 1. We call $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ a *penalty function*, whenever:

(a) $P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = 0$ if and only if $\mathbf{x}^{(i)} = \mathbf{y}$ for all $i \in [n]$;
(b) for every fixed $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{R}^d$ the minimum set of $P(\cdot; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$, i.e., $\left\{\mathbf{y} \in \mathbb{R}^d : P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \inf_{\mathbf{y}' \in \mathbb{R}^d} P(\mathbf{y}'; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})\right\}$, is nonempty, bounded, and convex.

Please cite this article in press as: M. Gagolewski, Penalty-based aggregation of multidimensional data, Fuzzy Sets Syst. (2016), http://dx.doi.org/10.1016/j.fss.2016.12.009