7

Fuzzy Filtering: A Mathematical Theory and Applications in Life Science

Mohit Kumar, Kerstin Thurow, Norbert Stoll, and Regina Stoll
University of Rostock
Germany

1. Introduction

A life science process is typically characterized by a large number of variables whose interrelations are uncertain and not completely known. The development of a computational paradigm, implementing an "intelligent" behavior in the sense of handling uncertainties related to the modeling of the interrelations among process variables, is an interesting research topic. A large number of studies apply computational intelligence techniques in the life science e.g.

- in modeling the environmental behavior of chemicals (Eldred & Jurs, 1999; Kaiser & Niculescu, 1999; Gini et al., 1999; L. Sztandera et al., 2003; Sztandera et al., 2003; Vracko, 1997; Benfenati & Gini, 1997; Gini, 2000; Mazzatorta et al., 2003),
- in medicine (Wilson & Russell, 2003b; Fukuda et al., 2001; Wilson & Russell, 2003a; Mandryk & Atkins, 2007; Lin et al., 2006; Rani et al., 2002; Adlassnig, 1986; Adlassnig et al., 1985; Bellazzi et al., 2001; 1998; Belmonte et al., 1994; Binaghi et al., 1993; Brai et al., 1994; Daniels et al., 1997; Fathitorbaghan & Meyer, 1994; Garibaldi & Ifeachor, 1999; Kuncheva & Steimann, 1999; Roy & Biswas, 1992; Steimann, 1996;Watanabe et al., 1994; Wong et al., 1990),
- in chemistry and drug design, see e.g. (Manallack & Livingstone, 1999; Winkler, 2004; Duch et al., 2007) and references therein.

The fuzzy systems based on fuzzy set theory (Zadeh, 1973; 1983) are considered suitable tools for dealing with the uncertainties. The use of fuzzy systems in data driven modeling is a topic that is widely studied by the researchers (Wang & Mendel, 1992; Nozaki et al., 1997; Shan & Fu, 1995; Nauck & Kruse, 1998; Jang, 1993; Thrift, 1991; Liska & Melsheimer, 1994; Herrera et al., 1994; González & Pérez, 1998; Babuška & Verbruggen, 1997; Babuška, 1998; Abonyi et al., 2002; Simon, 2000; 2002; Jang et al., 1997; Wang & Vrbanek, 2008; Lughofer, 2008; Kumar, Stoll & Stoll, 2009b; Lin et al., 2008; Kumar, Stoll & Stoll, 2009a) due to the successful applications of fuzzy techniques in data mining, prediction, control, classification, simulation, and pattern recognition.

It is assumed that input variables ($x_1, x_2, \ldots, x_n$) are related to the output variable y through a mapping:

$$y = f(x)$$

Source: Fuzzy Systems, Book edited by: Ahmad Taher Azar,
ISBN 978-953-7619-92-3, pp. 216, February 2010, INTECH, Croatia, downloaded from SCIYO.COM

where $x = [x_1\ x_2\ \ldots\ x_n] \in R^n$ is the input vector and the modeling aim is to identify the unknown function f . The fuzzy modeling is based on the assumption that there exists an ideal set of model parameters w^* such that model output $M(x;w^*)$ to input x is an approximation of the output value y. However, it may not be possible, for a given type and structure of the model M, to identify perfectly the inputs-output relationships. The part of the input-output mappings that can't be modeled, for a given type and structure of the model, is what we refer to as the uncertainty. Mathematically, we have

$$y = M(x; w^*) + n, \tag{1}$$

where n is termed as disturbance or noise in system identification literature. However, we refer n, in context to real-world modeling applications, to as uncertainty to emphasize that the uncertainties regarding optimal choices of the model and errors in output data resulted in the additive disturbance in (1). For an illustration, the authors in (Kumar et al., 2008), in context to subjective workload score modeling, explain the reasons giving rise to the uncertainty.

A robust (towards uncertainty n) identification of model parameters w^* using available inputs-output data pairs $\{ x(j), y(j) \}_{j=0,1,\ldots}$ is obviously a straightforward approach to handle the uncertainty. Several robust methods of fuzzy identification have been developed (Chen & Jain, 1994; Wang et al., 1997; Burger et al., 2002; Yu & Li, 2004; Johansen, 1996; Hong et al., 2004; Kim et al., 2006; Kumar et al., 2004b; 2003b; 2006c; 2004a; 2006a;b). It may be desired to estimate the parameters w^* in an on-line scenario using an adaptive filtering algorithm aiming at the filtering of uncertainty n from y. A classical application of adaptive filters is to remove noise and artifacts from the biomedical signals (Philips, 1996; Lee & Lee, 2005; Plataniotis et al., 1999; Mastorocostas et al., 2000; Li et al., 2008). The adaptive filtering algorithms applications are not only limited to the engineering problems but also e.g. to medicinal chemistry where it is required to predict the biological activity of a chemical compound before its synthesis in the lab (Kumar et al., 2007b). Once a compound is synthesized and tested experimentally for its activity, the experimental data can be used for an improvement of the prediction performance (i.e. online learning of the adaptive system). Adaptive filtering of uncertainties may be desired e.g. for an intelligent interpretation of medical data which are contaminated by the uncertainties arising from the individual variations due to a difference in age, gender and body conditions (Kumar et al., 2007).

2. The fuzzy filter

It is required to filter out the uncertainties from the data with applications to many real-world modeling problems (Kumar et al., 2007; Kumar et al., 2007; Kumar et al., 2007a;b; 2008; Kumar et al., 2009; Kumar et al., 2008). A filter, in the context of our study, simply maps an input vector x to the quantity $y - n$ (called filtered output $y_f = y - n$) and thus separates uncertainty n from the output value y.

2.1 A Takagi-Sugeno fuzzy filter

Consider a zero-order Takagi-Sugeno fuzzy model ($F_s : X \rightarrow Y$) that maps n–dimensional input space ($X = X_1 \times X_2 \times \ldots \times X_n$) to one dimensional real line. A rule of the model is represented as