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; Lughhofer, 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:

\[
y = f(x)
\]
where \( x = [x_1 \ x_2 \ ... \ x_n] \in \mathbb{R}^n \) 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

\[
y = M(x;w^*) + n, \quad (1)
\]

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 \( \{x(j),y(j)\}_{j=0,1,...} \) 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 \( F_i : X \rightarrow Y \) 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
If \( x_1 \) is \( A_1 \) and \( \cdots \) and \( x_n \) is \( A_n \) then \( y_f = s \).

Here \((x_1, \ldots, x_n)\) are the model input variables, \( y_f \) is the filtered output variable, \((A_1, \ldots, A_n)\) are the linguistic terms which are represented by fuzzy sets, and \( s \) is a real scalar. Given a universe of discourse \( X_j \), a fuzzy subset \( A_j \) of \( X_j \) is characterized by a mapping:

\[
\mu_{A_j} : X_j \rightarrow [0,1],
\]

where for \( x_j \in X_j \), \( \mu_{A_j}(x_j) \) can be interpreted as the degree or grade to which \( x_j \) belongs to \( A_j \). This mapping is called as membership function of the fuzzy set. Let us define, for \( j^{th} \) input, \( P_j \) non-empty fuzzy subsets of \( X_j \) (represented by \( A_{1j}, A_{2j}, \ldots, A_{P_j} \)). Let the \( i^{th} \) rule of the rule-base is represented as

\[
R_i : \text{If } x_1 \text{ is } A_{1i} \text{ and } \cdots \text{ and } x_n \text{ is } A_{ni} \text{ then } y_f = s_i,
\]

where \( A_{1i} \in \{A_{11}, \ldots, A_{P_1}\}, A_{2i} \in \{A_{12}, \ldots, A_{P_2}\} \) and so on. Now, the different choices of \( A_{1i}, A_{2i}, \ldots, A_{ni} \) leads to the \( K = \prod_{j=1}^{n} P_j \) number of fuzzy rules. For a given input \( x \), the degree of fulfillment of the \( i^{th} \) rule, by modelling the logic operator ‘and’ using product, is given by

\[
g_i(x) = \prod_{j=1}^{n} \mu_{A_{ji}}(x_j).
\]

The output of the fuzzy model to input vector \( x \in X \) is computed by taking the weighted average of the output provided by each rule:

\[
y_f = \frac{\sum_{i=1}^{K} s_i g_i(x)}{\sum_{i=1}^{K} g_i(x)} = \frac{\sum_{i=1}^{K} s_i \prod_{j=1}^{n} \mu_{A_{ji}}(x_j)}{\sum_{i=1}^{K} \prod_{j=1}^{n} \mu_{A_{ji}}(x_j)}.
\]

Let us define a real vector \( \theta \) such that the membership functions of any type (e.g. trapezoidal, triangular, etc) can be constructed from the elements of vector \( \theta \). To illustrate the construction of membership functions based on knot vector \( (\theta) \), consider the following examples:

### 2.1.1 Trapezoidal membership functions:

Let

\[
\theta = (a_1, t_1^1, \ldots, t_1^{2P_1-2}, b_1, \ldots, a_n, t_n^1, \ldots, t_n^{2P_n-2}, b_n)
\]

such that for \( i^{th} \) input \((x_i \in [a_i, b_i])\), \( a_i < t_1^1 < \cdots < t_i^{i2P_i-2} < b_i \) holds \( \forall \ i = 1, \ldots, n \). Now, \( P_i \) trapezoidal membership functions for \( i^{th} \) input \((\mu_{A_{1i}}, \mu_{A_{2i}}, \ldots, \mu_{A_{Pi}})\) can be defined as:
\[ \mu_{A_{i_1}}(x_i, \theta) = \begin{cases} 
1 & \text{if } x_i \in [a_i, t_1^i] \\
-\frac{x_i + t_1^2}{t_2^i - t_1^i} & \text{if } x_i \in [t_1^i, t_2^i] \\
0 & \text{otherwise} 
\end{cases} \]

\[ \mu_{A_{i_2}}(x_i, \theta) = \begin{cases} 
\frac{x_i - t_i^{2j-3}}{t_i^{2j-2} - t_i^{2j-3}} & \text{if } x_i \in [t_i^{2j-3}, t_i^{2j-2}] \\
1 & \text{if } x_i \in [t_i^{2j-2}, t_i^{2j-1}] \\
-\frac{x_i + t_i^{2j}}{t_i^{2j} - t_i^{2j-1}} & \text{if } x_i \in [t_i^{2j-1}, t_i^{2j}] \\
0 & \text{otherwise} 
\end{cases} \]

\[ \vdots \]

\[ \mu_{A_{i_n}}(x_i, \theta) = \begin{cases} 
\frac{x_i - t_i^{2P-3}}{t_i^{2P-2} - t_i^{2P-3}} & \text{if } x_i \in [t_i^{2P-3}, t_i^{2P-2}] \\
1 & \text{if } x_i \in [t_i^{2P-2}, b_i] \\
0 & \text{otherwise} 
\end{cases} \]

### 2.1.2 One-dimensional clustering criterion based membership functions:

Let

\[ \theta = (a_1, t_1^1, \ldots, t_1^{n-2}, b_1, \ldots, a_n, t_n^1, \ldots, t_n^{n-2}, b_n) \]

such that for \( i \)-th input, \( a_i < t_1^1 < \cdots < t_1^{n-2} < b_i \) holds for all \( i = 1, \ldots, n \). Now, consider the problem of assigning two different memberships (say \( \mu_{A_{i_1}} \) and \( \mu_{A_{i_2}} \)) to a point \( x_i \) such that \( a_i < x_i < t_1^i \), based on following clustering criterion:

\[ [\mu_{A_{i_1}}(x_i), \mu_{A_{i_2}}(x_i)] = \arg \min_{u_1, u_2} \left[ u_1^2 (x_i - a_i)^2 + u_2^2 (x_i - t_1^i)^2, u_1 + u_2 = 1 \right]. \]

This results in

\[ \mu_{A_{i_1}}(x_i) = \frac{(x_i - t_1^i)^2}{(x_i - a_i)^2 + (x_i - t_1^i)^2}, \quad \mu_{A_{i_2}}(x_i) = \frac{(x_i - a_i)^2}{(x_i - a_i)^2 + (x_i - t_1^i)^2}. \]

Thus, for \( i \)-th input, \( P_i \) membership functions can be defined as:

\[ \mu_{A_{i_1}}(x_i, \theta) = \begin{cases} 
1 & x_i \leq a_i \\
\frac{(x_i - t_1^i)^2}{(x_i - a_i)^2 + (x_i - t_1^i)^2} & a_i \leq x_i \leq t_1^i \\
0 & \text{otherwise} 
\end{cases} \]
\[ \mu_{A_{x_1}}(x_1, \theta) = \begin{cases} \frac{(x_1 - a_i)^2}{(x_1 - a_i)^2 + (x_1 - t_i^1)^2} & a_i \leq x_1 \leq t_i^1 \\ \frac{(x_1 - t_i^1)^2}{(x_1 - t_i^1)^2 + (x_1 - t_i^2)^2} & t_i^1 \leq x_1 \leq t_i^2 \\ 0 & \text{otherwise} \end{cases} \]

\[ \vdots \]

\[ \mu_{A_{x_n}}(x_n, \theta) = \begin{cases} \frac{(x_n - a_i)^2}{(x_n - a_i)^2 + (x_n - t_i^n)^2} & a_i \leq x_n \leq t_i^n \\ \frac{(x_n - t_i^n)^2}{(x_n - t_i^n)^2 + (x_n - t_i^{n+1})^2} & t_i^n \leq x_n \leq t_i^{n+1} \\ 0 & \text{otherwise} \end{cases} \]

For any choice of membership functions (which can be constructed from a vector \( \theta \)), (2) can be rewritten as function of \( \theta \):

\[ y_f = \sum_{i=1}^K \sum_{j=1}^n G_i(x_1, x_2, \ldots, x_n, \theta), G_i(x_1, x_2, \ldots, x_n, \theta) = \frac{\prod_{j=1}^n \mu_{A_{x_j}}(x_j, \theta)}{\sum_{i=1}^K \prod_{j=1}^n \mu_{A_{x_j}}(x_j, \theta)}. \]

Let us introduce the following notation: \( \alpha = [s_1 \cdots s_k] \in \mathbb{R}^K \), \( x = [x_1 \cdots x_n] \in \mathbb{R}^n \), \( G(x, \theta) = [G_1(x, \theta) \cdots G_k(x, \theta)] \in \mathbb{R}^K \). Now, (2) becomes

\[ y_f = G(x, \theta) \alpha. \]

In this expression, \( \theta \) is not allowed to be any arbitrary vector, since the elements of \( \theta \) must ensure

1. in case of trapezoidal membership functions,

\[ a_i < t_i^1 < \cdots < t_i^{2n-2} < b_i, \forall i = 1, \ldots, n, \quad (3) \]

2. in case of one-dimensional clustering criterion based membership functions

\[ a_i < t_i^1 < \cdots < t_i^{p-2} < b_i, \forall i = 1, \ldots, n, \quad (4) \]

to preserve the linguistic interpretation of fuzzy rule base (Lindskog, 1997). In other words, there must exists some \( \epsilon_i > 0 \) for all \( i = 1, \ldots, n \) such that for trapezoidal membership functions,

\[
\begin{align*}
t_i^1 - a_i & \geq \epsilon_i, \\
t_i^{j+1} - t_i^j & \geq \epsilon_i, \quad \text{for all } j = 1, 2, \ldots, (2P_i - 3) \\
b_i - t_i^{2P_i-2} & \geq \epsilon_i.
\end{align*}
\]

These inequalities and any other membership functions related constraints (designed for incorporating a priori knowledge) can be written in the form of a matrix inequality \( c \theta \geq h \).
A Sugeno type fuzzy filter can be represented as
\[ y_f = G^T(x, \theta)\alpha, \ c\theta \geq h. \] (5)

### 2.2 A clustering based fuzzy filter

The fuzzy filter of (Kumar et al., 2007; Kumar et al., 2007; Kumar et al., 2007a;b; 2008; Kumar et al., 2009; Kumar et al., 2008) has \( K \) number of fuzzy rules of following type:

*If \( x \) belongs to a cluster having centre \( c_i \) then \( y_f = s_i \)

*If \( x \) belongs to a cluster having centre \( c_k \) then \( y_f = s_k \)

where \( c_i \in \mathbb{R}^n \) is the centre of \( i \)th cluster, and the values \( s_1, \ldots, s_k \) are real numbers. Based on a clustering criterion, it was shown in e.g. (Kumar et al., 2008) that

\[
y_f = \sum_{i=1}^{K} s_i G_i(x, c_1, \ldots, c_k),
\]

\[
G_i(x, c_1, \ldots, c_k) = \frac{A_1(x, c_1, \ldots, c_k)}{\sum_{i=1}^{K} A_1(x, c_1, \ldots, c_k)}, A_i(x, c_1, \ldots, c_k) = \frac{A_{1i} + A_{2i}}{2}, m > 1,
\]

where \( A_{1i}, A_{2i} \) are given as

\[
A_{1i} = \begin{cases} 1 & x \in X \setminus \{c_j\}_{j=1}^{K}, \\ \frac{1}{(m-1)} \sum_{j=1}^{K} \frac{1}{\|x - c_j\|^2} & x = c_i, \\ 0 & x \in \{c_j\}_{j=1}^{K} \setminus \{c_i\} \end{cases}
\]

\[
A_{2i} = \exp(-\frac{\|x - c_i\|^2}{\delta_i}), \delta_i = \min_{j, j \neq i} \|c_j - c_i\|^2.
\]

With the notations:

\[
\alpha = [s_1 \ldots s_k] \in \mathbb{R}^k, \ \theta = [c_1^T \ldots c_k^T]^T \in \mathbb{R}^{kn}, G(x, \theta) = [G_1(x, \theta) \ldots G_k(x, \theta)] \in \mathbb{R}^k,
\]

the output of fuzzy filter for an input \( x \) can be expressed as

\[ y_f = G^T(x, \theta)\alpha. \] (6)

### 3. Estimation of fuzzy model parameters

The fuzzy filter parameters \( (\alpha, \theta) \) need to be estimated using given inputs-output data pairs \( \{x(j), y(j)\}_{j=0,1,\ldots,N} \). This section outlines some of our results on the topic.
Result 1 (The result of (Kumar et al., 2009b)) A class of algorithms for estimating the parameters of Takagi-Sugeno type fuzzy filter recursively using input-output data pairs \( \{ x(j), y(j) \}_{j=0,1,...} \) is given by the following recursions:

\[
\theta_j = \arg \min_{\theta} \left[ \Psi_j(\theta), c\theta \geq h \right]
\]

\[
\alpha_j = \alpha_{j-1} + \frac{P_j G(x(j), \theta_j)[y(j) - G^T(x(j), \theta_j)\alpha_{j-1}]}{1 + G^T(x(j), \theta_j)P_j G(x(j), \theta_j)}
\]

\[
\Psi_j(\theta) = \frac{|y(j) - G^T(x(j), \theta)\alpha_{j-1}|^2}{1 + G^T(x(j), \theta)P_j G(x(j), \theta)} + \mu_\alpha^{-1} \| \theta - \theta_{j-1} \|^2
\]

for all \( j = 0, 1, ... \) with \( \alpha_{-1} = 0 \), \( P_0 = \mu I \), and \( \theta_1 \) is an initial guess about antecedents. The positive constants \( (\mu, \mu_\alpha) \) are the learning rates for \( (\alpha, \theta) \) respectively. Here, \( \gamma \geq -1 \) is a scalar whose different choices solve the following different filtering problems:

- \( \gamma = -1 \) solves a \( H^\infty \)-optimal like filtering problem,
- \( -1 \leq \gamma < 0 \) solves a risk-averse like filtering problem,
- \( \gamma > 0 \) solves a risk-seeking like filtering problem.

The positive constants \( \mu_\theta \) in (9) is the learning rate for \( \theta \). The elements of vector \( \theta \), if assumed as random variables, may have different variances depending upon the distribution functions of different inputs. Therefore, estimating the elements of \( \theta \in R^L \) with different learning rates makes a sense. To do this, define a diagonal matrix \( \Sigma \) (with positive entries on its main diagonal):

\[
\Sigma = \begin{bmatrix}
\mu_{\theta(1)} & 0 & \cdots & 0 \\
0 & \mu_{\theta(2)} & \cdots & 0 \\
0 & 0 & \ddots & \ddots \\
0 & 0 & \cdots & \mu_{\theta(L)}
\end{bmatrix}
\]

and to reformulate (9) as

\[
\Psi_j(\theta) = \frac{|y(j) - G^T(x(j), \theta)\alpha_{j-1}|^2}{1 + G^T(x(j), \theta)P_j G(x(j), \theta)} + \mu_\theta^{-1} \| \Sigma^{-1/2}(\theta - \theta_{j-1}) \|^2.
\]

Result 2 (The result of (Kumar, Stoll & Stoll, 2009a)) The adaptive \( p \)-norm algorithms for estimating the parameters of Takagi-Sugeno type fuzzy filter recursively using input-output data pairs \( \{ x(j), y(j) \}_{j=0,1,...} \) take a general form of

\[
\theta_j = \arg \min_{\theta} \left[ E_j(\hat{\alpha}(\theta), \theta) + \mu_\alpha^{-1}d_\alpha(\theta, \theta_{j-1}); c\theta \geq h \right]
\]

\[
\alpha_j = f^{-1}\left( f(\alpha_{j-1}) + \mu_\phi \phi(y(j) - G^T(x(j), \theta_j)\alpha_{j-1})G(x(j), \theta_j) \right)
\]

Here,

\[
E_j(\alpha, \theta) = L_j(\alpha, \theta) + \mu_j^{-1}d_\alpha(\alpha, \alpha_{j-1}).
\]
\[ \hat{\alpha}(\theta) = f^{-1}\left( f(\alpha_{j-1}) + \mu_0 \phi(y(j) - G^T(x(j),\theta)\alpha_{j-1})G(x(j),\theta) \right), \]

\[ d_j(u,w) = \frac{1}{2} ||u||^2 - \frac{1}{2} ||w||^2 - (u-w)^T f(w), \]

where \((\mu_0,\mu_0)\) are the learning rates for \((\alpha, \theta)\) respectively, \(f\) (a \(p\) indexing for \(f\) is understood), as defined in \((\text{Gentile, 2003})\), is the bijective mapping \(f: \mathbb{R}^K \rightarrow \mathbb{R}^K\) such that

\[ f = [f_1 \cdots f_k]^T, f_j(w) = \frac{\text{sign}(w_j) \cdot |w_j|^{p-1}}{||w||_q^{p-2}}, \]

where \(w = [w_1 \cdots w_k]^T \in \mathbb{R}^k\), \(q\) is dual to \(p\) (i.e. \(1 \div p + 1 \div q = 1\)), and \(\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \
\[
\phi(e) = ae + be^3
\]
\[
P_e(y, \bar{y}) = \frac{a}{2} |y - \bar{y}|^2 + b \left[ \frac{\bar{y}^4}{4} - \frac{y^4}{4} - (\bar{y} - y)y^3 \right]
\]

- **algorithm A**:

\[
L_r(\alpha, \theta) = \cosh(y(j) - G^T(x(j), \theta)\alpha))
\]
\[
\phi(e) = \sinh(e)
\]
\[
P_e(y, \bar{y}) = \cosh(\bar{y}) - \cosh(y) - (\bar{y} - y)\sinh(y)
\]

The filtering algorithms, with a learning rate of

\[
\mu_j = \frac{2P_e\{y(j), G^T(x(j), \theta)\alpha_{j-1}\}}{den},
\]

\[
den = \phi\{y(j) - G^T(x(j), \theta)\alpha_{j-1}\}(p-1)[\phi(y(j)) - \phi(G^T(x(j), \theta)\alpha_{j-1})]\|G(x(j), \theta)\|^2,
\]

\[
P_e(y, \bar{y}) = \int\bar{y} (\phi(r) - \phi(y)) dr,
\]

achieves a stability and robustness against disturbances in some sense.

For a standard algorithm for computing \(\theta\) numerically based on (11), define

\[
k^d_{\theta, \theta_{j-1}} = \begin{cases} 
\frac{2d_{\theta, \theta_{j-1}}}{\|\theta - \theta_{j-1}\|^2}, & \text{if } \theta \neq \theta_{j-1} \\
1, & \text{if } \theta = \theta_{j-1}
\end{cases}
\]

to express (11) as

\[
\theta_j = \arg \min_\theta [E_j(\hat{\alpha}(\theta), \theta) + \frac{\mu_{\theta_j}^{-1}k^d_{\theta, \theta_{j-1}}}{2} \|\theta - \theta_{j-1}\|^2].
\]

Choosing a time-invariant learning rate for \(\theta\) in (14), i.e. \(\mu_{\theta_j} = \mu_{\theta}\), and estimating the elements of vector \(\theta\) with different learning rates as in (10), (14) finally becomes

\[
\theta_j = \arg \min_\theta [E_j(\hat{\alpha}(\theta), \theta) + \frac{k^d_{\theta, \theta_{j-1}}}{2} \Sigma^{-1/2}(\theta - \theta_{j-1})]\|\Sigma^{-1/2}(\theta - \theta_{j-1})\|^2].
\]

Define vectors \(r(\theta)\) and \(r_{\bar{y}}(\theta)\) as

\[
r(\theta) = \left[ \left( \frac{[y(j) - G^T(x(j), \theta)\alpha_{j-1}]^2}{1 + G^T(x(j), \theta)P_G(x(j), \theta)} \right)^{1/2} \right] \in \mathbb{R}^{L+1},
\]

\[
r_{\bar{y}}(\theta) = \left[ \right] \in \mathbb{R}^{L+1},
\]

www.intechopen.com
Algorithm 1 Fuzzy filtering

Require: Data pairs \( \{x(j), y(j)\}_{j=0,1,\ldots,N} \).
1: Choose the number of rules \( K \); initial guess about the antecedents \( \theta_{-1} \); positive learning rate \( \mu \) for \( \alpha \); positive learning rates \( \mu_{\theta(1)}, \mu_{\theta(2)}, \ldots, \mu_{\theta(K)} \) for different elements of \( \theta \); number of maximum epochs \( E_{\text{max}} \).
2: Set data index \( j = 0 \); epoch count \( EC = 0 \); initial guess about consequents \( a_{-1} = 0 \).
3: Choose a filtering criterion: either of result 1 or of result 2.
4: if filtering criterion of result 1 then
5: Choose a value of \( \gamma \geq -1 \).
6: else \{filtering criterion of result 2\}
7: Choose an algorithm out of the 5 different algorithms \((A_{1,p}, A_{2,p}, A_{3,p}, A_{4,p}, A_{5,p})\) and a value of \( p \) \((2 \leq p \leq \infty)\).
8: Define \( L_j(\alpha, \theta), \varphi(e), P_\theta(y, y) \) depending upon the choice of algorithm.
9: end if
10: while \( EC < E_{\text{max}} \) do
11: \quad while \( j \leq N \) do
12: \quad \quad if filtering criterion of result 1 then
13: \quad \quad \quad Define \( r(\theta) \) as (16). Let \( s^*(\theta) \) be the unique solution, obtained by the algorithm suggested in (Lawson & Hanson, 1995), of the following constrained linear least-squares problem:
14: \quad \quad \quad \begin{align*}
\min_S & \quad \|r(\theta) + r'(\theta)s\|^2; \quad cs \geq h - c\theta,
\end{align*}
where \( r'(\theta) \) is the Jacobian matrix of vector \( r \) with respect to \( \theta \), determined by the method of finite-differences. The Jacobian \( r'(\theta) \) is a full rank matrix, since the main diagonal of the diagonal matrix \( \Sigma \) has positive entries.
15: \quad \quad \quad Compute \( \theta_j \) based on (18) using a Gauss-Newton like algorithm:
16: \quad \quad \quad \begin{align*}
\theta_j = \theta_{j-1} + s^*(\theta_{j-1}).
\end{align*}
17: \quad \quad \quad Compute \( \alpha_j \) using (8).
18: \quad \quad \quad else \{filtering criterion of result 2\}
19: \quad \quad \quad Define \( r_q(\theta) \) as (17) and let \( s^*_q(\theta) \) be the unique solution of following constrained linear least-squares problem:
20: \quad \quad \quad \begin{align*}
\min_S & \quad \|r_q(\theta) + r'_q(\theta)s\|^2; \quad cs \geq h - c\theta,
\end{align*}
where \( r'_q(\theta) \) is the Jacobian matrix of vector \( r_q \) with respect to \( \theta \).
21: \quad \quad \quad Compute \( \theta_j \) based on (18) using a Gauss-Newton like algorithm:
22: \quad \quad \quad \begin{align*}
\theta_j = \theta_{j-1} + s^*_q(\theta_{j-1}).
\end{align*}
23: \quad \quad \quad Compute \( \alpha_j \) using (12) with the learning rate provided by (13).
24: \quad \quad \quad end if
25: \quad \quad \quad j \leftarrow j + 1
26: \quad \quad \quad end while
27: \quad \quad EC \leftarrow EC + 1; \quad \alpha_{-1} = \alpha_N; \quad \theta_{-1} = \theta_N; \quad j = 0.
28: \quad \quad end while
29: return identified fuzzy filter parameters \( \alpha^l = \alpha_N \) and \( \theta^l = \theta_N \).
so that (7) and (11) can be formulated as

\[
\theta_j = \begin{cases} 
\arg \min_\theta \left[ \| r(\theta) \|^2 ; c\theta \geq h \right], & \text{as per result} \\
\arg \min_\theta \left[ \| r_\alpha(\theta) \|^2 ; c\theta \geq h \right], & \text{as per result}
\end{cases}
\] (18)

Algorithm 1 presents an algorithm to estimate fuzzy filter parameters based on the filtering criteria of either result 1 or result 2. The constrained linear least-squares problem is solved by transforming first it to a least distance programming (Lawson & Hanson, 1995).

**Remark 1** Algorithm 1 estimates the parameters of the fuzzy filter of type (5). In the case of fuzzy filter of type (6), there are no matrix inequality constraints and thus linear least-squares problem will be solved at step 13 or 17 of algorithm 1.

4. Applications in life science

The efforts have been made by the authors to develop fuzzy filtering based methods for a proper handling of the uncertainties involved in applications related to the life science (Kumar et al., 2007; Kumar et al., 2008; Kumar et al., 2007; Kumar et al., 2009; Kumar et al., 2007a; Kumar et al., 2007; Kumar et al., 2008; 2007b). This section provides a brief summary of some of the studies.

4.1 Quantitative Structure-Activity Relationship (QSAR)

4.1.1 Background

The QSAR methods developed by Hansch and Fujita (Hansch & Fujita, 1964) identify relationship between chemical structure of compounds and their activity and have been applied to chemistry and drug design (Guo, 1995; Kaiser, 1999; Jackson, 1995). The QSAR modeling is based on the principle that molecular properties like lipophilicity, shape, electronic properties modulate the biological activity of the molecule. Mathematically, biological activity is a function of molecular properties descriptors:

\[
BA = f(d_1, d_2, \ldots),
\]

where BA is a biological response (e.g. IC\(_{50}\), ED\(_{50}\), LD\(_{50}\)) and \(d_1, d_2, \ldots\) are mathematical descriptors of molecular properties. During the last years, the applications of neural networks in chemistry and drug design has dramatically increased. A review of the field can be found e.g. in (Manallack & Livingstone, 1999; Winkler, 2004). While developing a QSAR model for the design and discovery of bioactive agents, we may come across the situation that descriptors don’t accurately capture the molecular properties relevant to the biological activity or the chosen model structure (i.e. number of adjustable model parameters) is not optimal. In such situations, there exist modeling errors. The common problems associated with QSAR modeling can be summarized as follows:
1. For the chosen structure of the model and descriptors, there may exist modeling errors. The commonly used nonlinear model training algorithms (e.g. gradient-descent based backpropagation techniques) are not robust toward modeling errors.

2. The model identification process may result in the overtraining. This leads to a loss of ability of the identified model to generalize. Although overtraining can be avoided by using validation data sets, but the computation effort to cross-validate identified models can result in large validation times for a large and diverse training data set.

4.1.2 A fuzzy filtering based method
An important issue in QSAR modeling is of robustness, i.e., model should not undergo overtraining and model performance should be least sensitive to the modeling errors associated with the chosen descriptors and structure of the model. The fuzzy filtering based method of (Kumar et al., 2007b) establishes a robust input-output mappings for QSAR studies based on fuzzy “if-then” rules. The identification of these mappings (i.e. the construction of fuzzy rules) is based on a robust criterion being referred to as “energy-gain bounding approach” (Kumar et al., 2006a). The method minimize the maximum possible value of energy-gain from modeling errors to the identification errors. The maximum value of energy-gain (that will be minimized) is calculated over all possible finite disturbances without making any statistical assumptions about the nature of signals. The authors in (Kumar et al., 2007b) compare their method with Bayesian regularized neural networks through the QSAR modeling examples of 1) carboquinones data set, 2) benzodiazepine data set, and 3) predicting the rate constant for hydroxyl radical tropospheric degradation of 460 heterogeneous organic compounds.

4.2 Fuzzy filtering for environmental behavior of chemicals
4.2.1 Toxicity modeling
A fundamental concern in the Quantitative Structure-Activity Relationship approach to toxicity evaluation is the generalization of the model over a wide range of compounds. The data driven modeling of toxicity, due to the complex and ill-defined nature of eco-toxicological systems, is an uncertain process. The development of a toxicity predicting model without considering uncertainties may produce a model with a low generalization performance. The work of (Kumar et al., 2007) presents a novel approach to toxicity modeling that handles the involved uncertainties using a fuzzy filter, and thus improves the generalization capability of the model. The method is illustrated by considering a data set built up by U.S. Environmental Protection Agency referring to acute toxicity 96-h LC₅₀ in the fathead minnow fish (Pimephales promelas) (Russom et al., 1997; Pintore et al., 2003; Mazzatorta et al., 2003; Gini et al., 2004). The data set contains 568 compounds representing several chemical classes and modes of action.

4.2.2 Bioconcentration factor modeling
This work of (Kumar et al., 2009) presents a fuzzy filtering based technique for rendering robustness to the modeling methods. A case study, dealing with the development of a model for predicting the bioconcentration factor (BCF) of chemicals, was considered. The conventional neural/fuzzy BCF models, due to the involved uncertainties, may have a poor generalization performance (i.e. poor prediction performance for new chemicals). The
approach of (Kumar et al., 2009) to improve the generalization performance of neural/fuzzy BCF models consists of
1. exploiting a fuzzy filter to filter out the uncertainties from the modeling problem,
2. utilizing the information about uncertainties, being provided by the fuzzy filter, for the identification of robust BCF models with an increased generalization performance.
The approach was illustrated with a data set of 511 chemicals (Dimitrov et al., 2005) taking different types of neural/fuzzy modeling techniques.

4.3 Mental stress assessment
The work presented in (Kumar et al., 2007) used fuzzy filtering for mental stress assessment via evaluating the heart rate signals. The approach consists of
1. online monitoring of heart rate signal,
2. signal processing (e.g. using the continuous wavelet transform to extract the local features of heart rate variability in time-frequency domain),
3. exploiting fuzzy clustering and fuzzy identification techniques to render robustness in heart rate variability analysis against uncertainties due to individual variations,
4. monitoring the functioning of autonomic nervous system under different stress conditions.
The experiments involved 38 physically fit subjects (26 male, 12 female, aged 18-29 years) in air traffic control task simulations.

4.4 Subjective workload score modeling
A fuzzy filtering based tool was developed in (Kumar et al., 2008) to predict the subjective workload score of the operators working in the chemistry laboratories with different levels of automation. The work proposed a fuzzy-based modeling technique that first filters out the uncertainties from the modeling problem, analyzes the uncertainties statistically using finite-mixture modeling, and, finally, utilizes the information about uncertainties for adapting the workload model to an individual’s physiological conditions. The method of (Kumar et al., 2008) was demonstrated with the real-world medical data of 11 subjects who conducted an enzymatic inhibition assay in the chemistry laboratories under different workload situations.

5. References


While several books are available today that address the mathematical and philosophical foundations of fuzzy logic, none, unfortunately, provides the practicing knowledge engineer, system analyst, and project manager with specific, practical information about fuzzy system modeling. Those few books that include applications and case studies concentrate almost exclusively on engineering problems: pendulum balancing, truck backeruppers, cement kilns, antilock braking systems, image pattern recognition, and digital signal processing. Yet the application of fuzzy logic to engineering problems represents only a fraction of its real potential. As a method of encoding and using human knowledge in a form that is very close to the way experts think about difficult, complex problems, fuzzy systems provide the facilities necessary to break through the computational bottlenecks associated with traditional decision support and expert systems. Additionally, fuzzy systems provide a rich and robust method of building systems that include multiple conflicting, cooperating, and collaborating experts (a capability that generally eludes not only symbolic expert system users but analysts who have turned to such related technologies as neural networks and genetic algorithms). Yet the application of fuzzy logic in the areas of decision support, medical systems, database analysis and mining has been largely ignored by both the commercial vendors of decision support products and the knowledge engineers who use them.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
