1. Introduction

Manufacturing is a knowledge-intensive activity. The knowledge underlying a specific manufacturing process or system is often extracted from a small set of experimental observations. To automate the knowledge extraction process various machine learning methods have been used (Pham & Afifi, 2005; Monostori, 2003). Even though such methods are used, a great deal of human intelligence (knowledge extractor’s judgment, preference) is required for getting good results (Ullah & Khalifa, 2006). As a result, a machine learning method that is able to utilize human cognition as straightforwardly as possible seems more realistic for extracting knowledge in manufacturing. In fact, human-assisted machine learning methods are in agreement with the modern concept of manufacturing automation—how to support humans with computers rather than how to replace humans by computers (Kals et al., 2004). Thus, for advanced manufacturing systems, the machine learning methods wherein humans and computers compliment each other and the course of knowledge extraction is determined by the human cognition rather than by a fully automated algorithmic approach is desirable.

Artificial intelligence community has also started to recognize the need for human-assisted machine learning methods (i.e., human comprehensible machine learning methods):

“Humans need to trust that intelligent systems are behaving correctly, and one way to achieve such trust is to enable people to understand the inputs, outputs, and algorithms used as well as any new knowledge acquired through learning. As the use of machine learning increases in critical operations it is being applied increasingly in domains where the learning system’s inputs and outputs must be understood, or even modified, by human operators....” (Dan Oblinger, AAAI Technical Report, WS-05-04, 2005.)
Now the question is how to develop human comprehensible machine learning methods for manufacturing? One of the possibilities is to integrate Inductive-Statistical Explanation introduced by Hempel (Hempel, 1965) and Fuzzy Set based computing introduced by Zadeh (Zadeh, 1965). Based on this contemplation this chapter is written. The structure of this chapter is as follows: Section 2 describes the Hempelian paradigm relevant to this chapter. Section 3 describes Zadehian paradigm relevant to this chapter. Section 4 describes the non-monotonic nature of real-life probability which is important for understanding the rationale behind the cross-fertilization between Hempelian and Zadehian paradigms. Section 5 describes the human comprehensible machine learning method for extraction knowledge extraction process. Section 6 describes how the proposed knowledge extraction process is applied for modeling and simulation of nonlinear behaviors in manufacturing. Section 7 describes how the proposed knowledge extraction process is applied to The presented knowledge extraction method can be used to establish the relationship between performance measures and control variables a machining operations using a small set of data. Finally, the concluding remarks are shown.

2. Hempelian Paradigm

Carl Gustav Hempel, a philosopher of natural science (Woodward, 2003), introduced two models of scientific explanation: 1) Deductive-Nomological (D-N) Explanation (Hempel & Oppenheim, 1948) and 2) Inductive-Statistical (I-S) Explanation (Hempel, 1965, 1968). In these models it is assumed that a scientific explanation deals with explanan (object that explains a problem) and explanandum (object that needs to be explained). D-N Explanation means that there is a logically provable universal law (logical positivism) and a “certain” conclusion can be made from the law if the underlying conditions are satisfied—the explanan must be true and the explanandum must be the logical consequence of it (explanan) for all circumstances. The logical setting of D-N Explanation is as follows:

\[
\begin{align*}
&\text{(scientific law)} \quad p \rightarrow q \\
&\text{(condition)} \quad p \text{ is satisfied} \\
&\text{(certain conclusion)} \quad q \text{ must be true}
\end{align*}
\]
For example,

\[
\begin{array}{ll}
\text{(scientific law)} & \text{bird} \rightarrow \text{fly} \\
\text{(condition)} & \text{It is a bird} \\
\text{(certain conclusion)} & \text{It certainly can fly} \\
\end{array}
\]

In (1), p (bird) is called explanan and q (fly) is called explanandum. On the other hand, I-S Explanation means that there is no law-like relationship between explanan (p, bird) and explanandum (q, fly) and if one can prove that the probability of explanandum given explanan is very high then the explanandum is “most probably” the best explanation for the explanan. Note that in I-S Explanation the conclusion is not certain.

The logical setting of I-S Explanation is as follows:

\[
\begin{array}{ll}
\text{(probability rule)} & \Pr(q \mid p) = r, \quad r = 1 \\
\text{(condition)} & p \\
\text{(most probable conclusion)} & q \\
\end{array}
\]

For example,

\[
\begin{array}{ll}
\Pr(\text{fly} \mid \text{bird}) = r, \quad r = \text{at least} \ 0.95 \\
\text{It is a bird} \\
\text{It most likely can fly} \\
\end{array}
\]

I-S Explanation is more suitable model of explanation (knowledge extraction) for manufacturing because in manufacturing the arguments are experimental data-driven rather than scientific law driven.

3. Zadehian Paradigm

Lotfi Ashker Zadeh, a logician of human cognition, introduced some formal settings for dealing with uncertainty in man-machine systems. The main theme of his philosophy is that the human cognition has two main facets: \textit{Partiality} and \textit{Granularity}. Partiality means tolerance to partial truth—truth value of a proposition or an event is not only true and false, but also partially true or false. Granularity means formation of granules (words or phrases) assigned
un-sharply to a set of values or attributes. To known more about Zadeh’s philosopy, refer to Zadeh, 1965; 1975; 1997; 2002; 2005a, 2005b. On of the mathematical entities underlying Granularity and Partiality is fuzzy number (Dubois & Prade, 1978). A fuzzy number $A$ is defined by a membership function $\mu_A$ (user-defined) from real line to the interval 0 to 1 and it must satisfy the conditions of normality, convexity, continuity and compactness, as follows:

$$\mu_A : R \rightarrow [0,1] \text{such that}$$

(normality) $\exists x \in R, \mu_A(x) = 1$

(convexity) $\mu_A(\lambda x + (1-\lambda)y) \geq \min \{\mu_A(x), \mu_A(y)\}$ for any $x, y \in R, \lambda \in [0,1]$ (3)

(continuity) upper semi-continuous

(compactness) $\text{Supp}(A) = \{x \in R | \mu_A(x) > 0\}$ is a compact set

In semantics sense a fuzzy number $A$ is a fuzzy subset of the real line whose highest membership values $\mu_A$ are clustered around a given real number (Dubois & Prade, 1978). Therefore, it is found useful in formally computing uncertain quantities like “probability is at least 0.95”. Consider, for example, two fuzzy numbers illustrated in Fig. 1. Both fuzzy numbers define the same quantity $A = “\text{probability is at least 0.95}”$ by using two different membership functions:

$$\mu_A(Pr) : [0,1] \rightarrow [0,1]$$

definition (a)

$$Pr \mapsto \max \left( \min \left( \frac{Pr - 0.9}{0.95 - 0.9}, 1 \right), 0 \right)$$

definition (b)

$$Pr \mapsto \max \left( \min \left( \frac{Pr - 0.9}{0.95 - 0.9}, 1 \right), 0 \right)$$

(3.1)
Figure 1. Defining probability is at least 0.95 using fuzzy numbers.

The implications of these two definitions will be explained in the next section.

4. Linguistic Likelihood

In I-S Explanation, one is required to use classical probability (Kolmogorov, 1933; Billingsley, 1995; Hajeck, 2003) defined by a space called probability space \((\Omega, F, Pr)\) wherein \(\Omega\) is a non-empty set called universal set; \(F\), know as filed, is a set of all subsets of \(\Omega\) that has \(\Omega\) as a member and that is closed under complementation (with respect to \(\Omega\)) and union; \(Pr\), known as probability function, is a function from \(F\) to the real numbers. \(Pr\) is monotonic and follows three axioms as follows:

- (Axiom of Non-negativity) \(Pr(X) \geq 0\), for all \(X \in F\),
- (Axiom of Normalization) \(Pr(\Omega) = 1\),
- (Axiom of Finite Additivity) \(Pr(X \cup Y) = Pr(X) + Pr(Y)\) for all \(X, Y \in F\) such that \(X \cap Y = \emptyset\).

To make sure that a value of \(Pr(X)\) follows the above mentioned axioms and other theorems and corollaries derived from the above axioms, the value of \(Pr(X)\) is determined by the following statistical procedure:

\[
Pr(X) = \lim_{N \to \infty} \frac{N_X}{N}
\]  

\((4)\)
In (4), $X$ is an event or proposition; $N$ is the number of trials and $N_X$ is the number of trials wherein $X$ is found to be true. In practice, it is almost impossible to make infinite number of trials and determine the precise value of $Pr(X)$. As a result, $Pr(X)$ is estimated from a relatively large number of trials (from the relative frequency $N_X/N$, $N$ being a relatively large number, i.e., from a relatively large set of data) or from a probability distribution assuming that $X$ follows a distribution (e.g., normal distribution, binomial distribution, etc.). Sometimes the situation is much complex—the underlying probability distribution is unknown or the probability has to be estimated from a relatively small set of data. This is the case in most of the manufacturing situations.

Thus, the probability is perfect, but we can’t elicit it perfectly” (O’Hagan & Oakley, 2004). In other words, all real-life probabilities are not monotonic (i.e., non-monotonic) or imprecise in nature. To deal with the non-monotonic nature of probability (i.e., imprecise probabilities) a model with upper and lower provision is used (Walley, 1991; Walley et al., 2004; de Cooman & Zaffalon, 2004; de Cooman et al., 2005; Lukasiewicz, 2005; Tonn, 2005) wherein a probability $Pr(X)$ is expressed by lower provision, $Pr_l(X)$, and upper provision, $Pr_u(X)$, i.e., $Pr(X) = [Pr_l(X), Pr_u(X)]$. For example, consider $Pr(\text{fly} | \text{bird}) = “\text{at least 0.95”}$. Here, the lower provision $Pr_l(X) = 0.95$ and upper provision is $Pr_u(X) = 1$.

If someone calculates the value of $Pr(X)$ from a limited number of observations, then (from the sense of imprecise probability) $Pr(X)$ should be treated in such a way as if it is a range rather than a single value. One of the ways to achieve this is to use a set of linguistic likelihoods defined by appropriate fuzzy numbers and translate $Pr(X)$ into the linguistic likelihood that contains $Pr(X)$ most (Ullah & Harib, 2005). The translated linguistic likelihood can then be used to find upper and lower provisions or other quantities that are important from the view point of imprecise probability (Zadeh, 2002; 2005). For example, consider that one calculates $Pr(X) = 0.85$ from a limited number of observations.

This $Pr(X) = 0.85$ can be translated into linguistic likelihood labeled most-likely if the linguistic likelihoods illustrated in Figure 2 are used because $\mu_{\text{most-likely}}(0.85)$ is greater than $\mu_{\text{likely}}(0.85)$ and $\mu_{\text{most-unlikely}}(0.85)$. In fact, $\mu_{\text{most-likely}}(0.85) = 0.833$, $\mu_{\text{likely}}(0.85) = 0.167$, and $\mu_{\text{most-unlikely}}(0.85) = 0$ because
\[
\mu_{\text{most-likely}}(Pr) : [0,1] \rightarrow [0,1], \quad Pr \mapsto \max\left(\min\left(\frac{Pr-0.6}{0.9-0.6}, 1\right), 0\right)
\]
\[
\mu_{\text{likely}}(Pr) : [0,1] \rightarrow [0,1], \quad Pr \mapsto \max\left(\min\left(\frac{Pr-0.1}{0.4-0.1}, 1\right), 0\right).
\]
\[
\mu_{\text{most-unlikely}}(Pr) : [0,1] \rightarrow [0,1], \quad Pr \mapsto \max\left(\min\left(\frac{0.4-Pr}{0.4-0.1}, 1\right), 0\right).
\]

Figure 2. Three linguistic likelihoods.

Therefore, \(Pr(X) = 0.85\) is no longer a crisp value “0.85“. It is a fuzzy number labeled “most-likely”. If needed, the lower and upper provisions can be calculated from “most-likely” using \(\alpha\)-cuts:

\[\text{most-likely}_\alpha = \{Pr | \mu_{\text{most-likely}}(Pr) \geq \alpha, \alpha \in (0,1)\}.\]

For example, if \(\alpha = 0.5\), then most-likely\(_{\alpha=0.5} = [0.75,1]\), i.e., \(Pr(X) = 0.75\) and \(\overline{Pr(X)} = 1\). Moreover, if needed, the expected value of most-likely

\[E(\text{most-likely}) = \frac{\int_0^1 [Pr \times \mu_{\text{most-likely}}(Pr)] dPr}{\int_0^1 \mu_{\text{most-likely}}(Pr) dPr} = 0.860\]

can be used to find out the average value of the imprecise probability \(Pr(X) = 0.85\).
Compared to upper and lower provisions, expected value is more robust. The explanation is as follows: Recall the linguistic likelihoods shown in Fig. 1. 
\[ E(\text{probability is at least 0.95}) = 0.96 \] for the first definition and \[ E(\text{probability is at least 0.95}) = 0.95 \] for the other definition, respectively. This implies that even though the definition of a linguistic likelihood varies from person to person, the expected value does not vary much (for the above two cases the expected values are 0.96 and 0.95, a different of 1%, only). Therefore, if the expected value of a linguistic likelihood is used in an inference mechanism, the inferred output will not vary much from person to person. In other words, an inference mechanism will become robust if the expected value of the linguistic likelihood of an imprecise probability is used.

The above explanation can be summarized into the following procedure:

- Determine \( Pr \) from a small set of data
- Define a set of linguistic likelihoods, \( \{L_1, \ldots, L_n\} \) using appropriate fuzzy number in the universe of discourse \([0,1]\). Here, \( L_i = \{(Pr, \mu_{Li}(Pr)) | Pr, \mu_{Li}(Pr) \in [0,1]\}, \forall i \in \{1, \ldots, n\} \). All \( \mu_{Li}(Pr) \) follow the characteristics of fuzzy number.
- Translate \( Pr \) into \( LL \in \{L_1, \ldots, L_n\} \) such that \( \mu_{LL}(Pr) > \mu_{Lj}(Pr), L_j \in \{L_1, \ldots, L_n\}-\{LL\}, \forall j =1, \ldots, n \). This means \( Pr \mid = LL \).
- Determine the expected value, \( E(LL) \), of \( LL \) as follow:

\[
E(LL) = \frac{\int (\mu_{LL}(Pr) \times Pr) dPr}{\int (\mu_{LL}(Pr)) dPr}.
\]

- Use \( E(LL) \) instead of \( Pr \) in further calculations. This means that \( Pr \mid = LL \mid = E(LL) \).

5. Knowledge Extraction Process

Based on the Hempel’s I-S Explanation and Zadeh’s fuzzy number based linguistic likelihoods the logical setting for extracting knowledge from a small set of numerical data is proposed, as follows:

\[
Pr(q \mid p) = \text{most - likely,} \quad E(\text{most - likely}) = 1
\]

\[
\text{If } p \text{ Then } \text{most - likely } q
\]

(5)
In (5), “most-likely” is a metaphor that expresses the imprecise likeliness of occurring q given p. According to I-S Explanation the expected value of “most-likely” should be near to 1. Now, if p is found to be true, the extracted knowledge (i.e., the “If...Then...” rule “If p Then most-likely q”) can be used to produce output. The output is a value in the range underlying most-likely q, i.e., a range little longer than q constrained by the expected value of “most-likely”. For example, consider the arbitrary case shown in Fig. 3. As seen from Fig. 3, Pr(q|p) = 0.8 | “most-likely”, according to the definition linguistic likelihoods illustrated in Fig. 2.

Therefore, the rule in (5) holds for the arbitrary case shown in Fig. 3. If the input is a point in the range “p”, then the output is in the range “q” for most of the cases and is out of the range “q” for a few cases. This means that most-likely q = q’ is a range little longer than q or q ⊆ q’. This leads to the following inference mechanism:

\[
\begin{align*}
\text{(Extracted)} & \quad \text{If } p \text{ Then most-likely } q \\
\text{(Given)} & \quad x \in p \\
\text{(y ∈ q’)} & \quad (q \subseteq q’) \\
\end{align*}
\]

(6)

To derive p and q from two fuzzy numbers A and B, α-cuts can be employed. As such, p = A_{α1} and q = B_{α2}. In this case, q’ can be derived in the following manner:

\[
\begin{align*}
\text{(Extracted)} & \quad \text{If } A_{α1} \text{ Then most-likely } B_{α2} \\
\text{(Given)} & \quad x \in A_{α1} \\
\text{(Defined)} & \quad B’_{α2} = B_{α2} \\
\text{Supp}(B’) & = (\min(U_a), \max(U_a)) \\
\alpha’ & = \alpha 2 \times E(\text{most - likely}) \\
\text{(y ∈ B’}_{α’} & \\
\end{align*}
\]

(7)

In (7), B’ is a fuzzy number whose support spreads all over the universe of discourse of B and its α-cut at α2 is equal to that of B. Figure 4 depicts the inference mechanism. A seen from Fig. 4, the inferred output y is any value in the range B’_{α’}, which is an α-cut of B’ at α’ = α 2 × E( most-likely). Since α’ < α2, B’_{α’} > B_{α2}. 

Figure 3. An Arbitrary relationship between \( p \) and \( q \)

Figure 4. The Proposed Inference Mechanism
6. Nonlinear Signal Modeling and Simulation

This section shows how the proposed knowledge extraction method is applied to capture the dynamics of a highly nonlinear behavior $Y(t)$ from its return map, i.e., a map of $(Y(t), Y(t+1))$. Particularly for capturing the dynamics of surface roughness the following form of rules are found useful (Ullah & Harib, 2004; 2006):

- If $Y(t)$ is extremely low Then $Y(t+1)$ is most-likely very low
- If $Y(t)$ is very low Then $Y(t+1)$ is most-likely very low
- If $Y(t)$ is extremely high Then $Y(t+1)$ is most-likely extremely high

Here the phrase “most-likely” is metaphor for the linguistic likelihood associated with imprecise probability $\Pr(Y(t+1)=\text{State}|Y(t)=\text{State})$ of transition from a state $\text{State}_i$ to the same state after a unit interval. Figures 5 and 6 show some of the results and user interfaces of the software tool developed by the authors recently (Ullah & Harib, 2006).

Figure 5 shows how a user defines the states of $Y(t)$ (under the message Linguistic Variable) and linguistic likelihoods (under the message Probability). In this stage the system automatically generates the return map from an input signal. The system then extracts the “if…then…” rules, as shown in Fig. 5.
These rules can be used recurrently to simulate a surface roughness profile similar to the original one. For example consider the simulated surface roughness profiles shown in Fig. 6 wherein two consecutive simulations are shown. These signals can be connected piece-wise to produce a signal similar to the original one. See Ullah & Harib, 2006 for other computational issues associated with this simulation techniques and implications of such knowledge based technique from the context of exchanging information of nonlinear behaviors from one manufacturing system to another.
7. Knowledge Extraction for Machining Operations

In machining of materials, machining conditions (feed (f), depth of cut (ap), width of cut (ae), cutting velocity (Vc), tool nose radius (re) and others process parameters) should be adjusted in such a way so that the required surface finish or productivity can be ensured. The presented knowledge extraction method can be used to establish the relationship between process Performance Measures (PMs) and process Control Variables (CVs). The relationship should provide a recommended range (RE) of a CV to ensure an acceptable range (AC) of a PM. For machining, surface finish (Ra, fractal dimension of roughness profile) and productivity (e.g., for turning productivity = Vc×f×ap) are two common machining PMs. On the other hand, CV can be defined using a function \( \phi(Vc,f,ap,ae,re) \). See Ullah & Harib 2005a; 2005b for more details. The logical setting proposed for extracting rules for PMs and CVs is as follows:

\[
\begin{align*}
\Pr(PM = AC_{a1} | CV = RE_{a2}) &= \text{most – likely} \\
\Pr(PM = AC_{a1} | CV = -RE_{a2}) &= \text{most – unlikely} \\
E(\text{most – likely}) &= 1 \\
E(\text{most – unlikely}) &= 0
\end{align*}
\]

Based on logical setting in (8), a JAVA™ based computing tool has been developed (Ullah & Harib, 2005b) to extract rules. For example, consider the case shown in Fig. 7. For this particular case, the user extracts “If…Then…” rules to establish the relationship between surface finish (Ra) and Vc, f, and re. The interest is to set the values of Vc, f, and re in such a way so that Ra remains relatively small. Therefore, the acceptable range of Ra is set by a trapezoidal fuzzy number \( AC = (0,1,1.5,2.5) \) as shown in Fig. 7. The user defines \( CV = Vc \times f \times re \) (in Fig. 7, Vc is shown by V and re is shown by r) based on the judgement that surface roughness is affected mainly by f, re, and Vc. The data of CV and PM plotted in Fig. 7 are corresponds to that of Ullah & Harib 2005b. As seen from Fig. 7, the desired rules according to the logical setting in (8) are found when recommended range of CV, i.e., RE, is set to be a fuzzy number \( RE = (90,150,140,90) \). The rules are:
If $V \times f \times r_e$ is $RE_{0.5}$ Then $Ra$ is Absolutely Likely $AC_{0.5}$

If $V \times f \times r_e$ is not $RE_{0.5}$ Then $Ra$ is Absolutely Unlikely $AC_{0.5}$

Here, $RE_{0.5} = [115,180]$. According to the logical setting in (8), “Absolutely Likely $AC_{0.5}$” = “Absolutely Likely [0.5,2]” = [0.4806,2.8944]. This means that if someone keeps $V \times f \times r_e$ in the range of [115,180], then $Ra$ will remain in the range [0.480,2.8944]. Thus, the computer comprehensible rule is as follows:

If $V \times f \times r_e$ is [115,180] Then $Ra$ is [0.4806,2.8944].

The above rule can be further modified so that the modified rule remains valid for all most all recommended input points and predicts output not in the range much longer than the predicted range. One of the possible modifications is as follows:

If $V \times f \times r_e$ is Moderate Then $Ra$ is Fine.

In the modified rule, “Moderate” is a triangular fuzzy number (120,150,180). This implies that $Supp(Moderate) = [120,180]$, i.e., a range that is narrowly included in [115,180]. On the other hand, “Fine” is a trapezoidal fuzzy number (0,0,1,3). This implies that $Supp(Fine) = [0,3]$, i.e., a range that narrowly includes [0.4806,2.8944].
Note that the linguistic likelihoods (Absolutely Likely...Absolutely Unlikely) are user-defined as shown in the bottom of Fig. 7.

8. Concluding Remarks

Zadeh’s partiality and granularity based computation has widely been applied in developing intelligent systems since its inception in 1965 whereas Hempel’s empirical positivism (i.e., I-S Explanation) remains relatively untouched. There is a study that shows that Hempel’s I-S Explanation underlies non-monotonic or default logic—a computational framework for making natural conclusion under evolving information (Tan, 1997). Hempel’s I-S Explanation inspired inference from medical data has gained some attention (Gandjour & Lauterbach, 2003). This chapter has added a new dimension to the application potential of Hempel’s I-S Explanation in intelligent systems development. Particularly it is
shown that cross-fertilization between Hempel’s I-S Explanation and Zadeh’s fuzzy number based imprecise probabilities provides the logical setting for human comprehensible machine learning methods that are able to extract human- and machine-friendly “If...Then...” rules from empirical facts (i.e., numerical data).

Two case studies are shown in this chapter to demonstrate the effectiveness of the method. In the first case study it is shown that the presented method is helpful in nonlinear signal modeling and simulation. Particularly, the case of knowledge based modeling and simulation of surface roughness (a very highly nonlinear behavior in manufacturing) is shown. This idea can be extended to model other nonlinear behaviors (productivity, cutting forces, cutting temperature, etc.) encountered in manufacturing. Since the nonlinear behaviors are stored by using a small set of “If...Then” rules, the method can be used to exchange the information of nonlinear behaviors from one manufacturing system to another.

The other case study deals with the extraction of “If...Then...” rules for machining. The goal is to get such rules that are able to predict the machining performance measures (e.g., surface roughness, productivity) for a given combination of cutting conditions (feed, depth of cut, cutting velocity, tool nose radius, etc.). Particularly, a rule is extracted to ensure fine surface finish by keeping the product of cutting velocity, feed rate and tool nose radius to a range. The issues of how to make the rules more general is also elaborated.

As the modern concept of manufacturing automation is “how to support humans with computers” rather than “how to replace humans by computers”, the presented knowledge extraction method will provide valuable hints for the manufacturing systems developers to develop more human- and computer-friendly computing tools.

9. References


the 4th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering (CIRP ICME 04), 30 June-2 July 2004, Sorrento (Naples), Italy.


The primary goal of this book is to cover the state-of-the-art development and future directions in modern manufacturing systems. This interdisciplinary and comprehensive volume, consisting of 30 chapters, covers a survey of trends in distributed manufacturing, modern manufacturing equipment, product design process, rapid prototyping, quality assurance, from technological and organisational point of view and aspects of supply chain management.

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