Fuzzy reasoning for assessing bulk tank milk quality

Taioun Kim
Industrial Engineering Kyungsung University
(twkim@ks.ac.kr)

Daeyou Jung
Industrial Engineering Kyungsung University
(dyjung@ks.ac.kr)

Bhushan M. Jayarao
Veterinary Science Penn State University
(bmj3@psu.edu)

Many dairy producers periodically receive information about their bulk tank milk with reference to bulk tank somatic cell counts, standard plate counts, and preliminary incubation counts. This information, when collected over a period of time, in combination with bulk tank mastitis culture reports can become a significant knowledge base. Several guidelines have been proposed to interpret farm bulk tank milk bacterial counts. However many of the suggested interpretive criteria lack validation, and provide little insight to the interrelationship between different groups of bacteria found in bulk tank milk. Also the linguistic terms describing bulk tank milk quality or herd management status are rather vague or fuzzy such as excellent, good or unsatisfactory. The objective of this paper was to develop a set of fuzzy descriptors to evaluate bulk tank milk quality and herd's milking practice based on bulk tank milk microbiology test results. Thus, fuzzy logic based reasoning methodologies were developed based on fuzzy inference engine. Input parameters were bulk tank somatic cell counts, standard plate counts, preliminary incubation counts, laboratory pasteurization counts, non agalactiae-Streptococci and Streptococci like organisms, and Staphylococcus aureus. Based on the input data, bulk tank milk quality was classified as excellent, good, milk cooling problem, cleaning problem, environmental mastitis, or mixed with mastitis and cleaning problems. The results from fuzzy reasoning would provide a reference regarding a good management practice for milk producers, dairy health consultants, and veterinarians.

Key words : Bulk tank milk, Mastitis, Fuzzy logic reasoning

Received: September 2004 Accepted: December 2004 Corresponding Author: Taioun Kim

1. Introduction

Bulk tank milk (BTM) from all dairy farms is periodically tested for antibiotic residues and for bacterial contamination. Many progressive milk cooperatives and processors periodically test for raw milk quality and also encourage milk producers to test their BTM for mastitis causing bacteria. Studies conducted over the last decade have shown that examination of BTM is of practical value in terms of time and cost of analysis for diagnosing multiple problems (current and potential) that might exist in a dairy herd related to milk quality and mastitis pathogens (Godkin and

Leslie, 1993). Recently, using BTM as a tool to ascertain BTM quality and troubleshoot herds with mastitis has received a lot of attention, especially from veterinarians and dairy health consultants who view milk quality and mastitis as an important aspect of their consultancy services for their clients (Mickelson et al., 1998).

Mastitis, a term used to define an inflammation of the mammary gland, can be a challenge to diagnose based on dairy production records and on-farm observations by the novice in the absence of culture results. Mastitis is usually classified as clinical or subclinical mastitis. Clinical mastitis is obvious and easier to identify and treat to the dairy farmers as the milk or udder is abnormal and perhaps systemic signs of disease are present. In contrast, subclinical mastitis is infections, which occur without obvious clinical signs such as abnormal milk, udder swelling or tenderness, or systemic signs such as fever, depression and drop in milk production. The latter is the most economically problematic from the point of view of the dairy industry as a whole. Also, mastitis is classified as contagious and environmental depending on the suspected site of origin and method of transmission. The diagnosis of clinical mastitis is straightforward because something is obviously wrong and it is easy to see. With subclinical mastitis, detection is most important, and it is through somatic cell counting. Thus somatic cell count (SCC) is a very useful measure to increase the awareness of subclinical mastitis and its effect on production and milk quality. The types of infection that can be routinely monitored with SCC are the contagious types and the environmental mastitis pathogens, in which these infections occur as subclinical pattern (Kirk, 2003).

The understanding of the bacteriology of raw milk, mastitis and farm management practices related to milking and milk hygiene has increased considerably, making it possible to formulate strategies to improve milk quality and reduce the incidence of mastitis in the dairy herd. With mastitis, earlier researches focused more statistical linear models to relate mastitis with SCC (Berning and Shook, 1992; Nielen et al., 1993; Pösö and Mäntysaari, 1996). Various approaches have been adopted in order to diagnose symptoms or predict complex systems in the area of reasoning. Recently, combining relevant machine learning techniques and artificial intelligence schemes were reported to enhance the prediction and classification results for those problems. Knowledge based system (Hogeveen, et al., 1995), artificial neural network schemes (Dybowski and Gant, 1995; Heald et al., 2000), and fuzzy logic (Salehi et al., 2000; Ureña et al., 2001; Klaus et al., 2003; Jahns et al., 2003) are some of those efforts to name a few.

Since Zadeh's seminal paper (1965) for vague and fuzzy problem domains, fuzzy mathematics, theoretical modifications and many applications can be found in Ross's (1995) and Zimmermann's (2001) works. Bridges et al. (1995) developed fuzzy descriptors for the time-varying data from weather file. The derived system was used both as a natural language generator and as a weather file selector for a crop simulation program. As an analytical tool for dairy herd improvement data, fuzzy logic was applied to evaluate milk yield and persistency (Lacroix, et al., 1998). A fuzzy indicator consisting of parity, state of lactation, and herd average production level was adopted to assess milk yield deviation from standard value. Salehi et al. (2000) proposed a neuro-fuzzifier that would mimic an expert's process of assigning dairy production records to fuzzy milk-yield sets and assessing their corresponding degrees of membership. Al-Faraj et al. (2001) developed a fuzzy logic crop water stress index using growth chamber data and tested on tall fescue canopies grown in the greenhouse. In addition to a few cases quoted above, fuzzy logic systems have been applied in many agricultural areas. For the mastitis and milk quality issues, many efforts have been tried utilizing artificial intelligence techniques including fuzzy logic.

As a tool for improving milk quality and herd udder health, BTM analysis is very important. But, most raw data gathered from dairy farm or found in real cases contain some degree of vagueness and fuzziness. This problem may be due to the complex nature of the real world. In case of statistical multivariate model or decision tree induction, the crisp values are used as input and inference is conducted based on the crisp value itself, which can not consider the fuzziness behind the input value. Thus, they represent only crisp relations but not fuzzy relations. In fuzzy relations, the strength of mapping from crisp input variable into the output space is expressed by membership function representing various "degrees" of strength of the relation. With these in mind, fuzzy logic based reasoning methodology will be well suited for this problem area.

The objective of this paper was to develop a set of fuzzy descriptors to evaluate BTM quality and herd's milking practice based on BTM microbiology test results. Thus, fuzzy logic based reasoning methodologies were developed based on fuzzy inference engine, then BTM quality was evaluated. The next section describes BTM analysis, the architecture of fuzzy reasoning model, membership function, and the process used to generate fuzzy reasoning. Implementation output from the fuzzy reasoning model, results of defuzzification, comparison between fuzzy output and expert's evaluation, discussion and conclusions were given in the following sections.

2. Methodology

2.1 Bulk tank milk analysis

Variables regarding the milk quality are composed of bulk tank somatic cell count (BTSCC), bacteriological tests for milk quality, contagious mastitis, and environmental mastitis pathogens. BTSCC is a valuable indicator of udder health and milk quality. Monitoring udder health status of a herd by BTSCC is very useful, especially in herds with contagious mastitis or environmental pathogens (Jayarao et al., 2001). Milk from uninfected quarters will generally have

BTSCC less than 200,000 cells/ml. Although no specific SCC minimum can be used for detection of an infection, the probability that an infection is present increases as the SCC increases. The procedures for conducting bacteriological tests for milk quality are described in the standard methods for examination of milk and milk products (Marshall, 1992). These procedures are industry accepted procedures. Bacteriological tests for milk quality include standard plate count (SPC), preliminary incubation count (PIC), and laboratory pasteurization count (LPC). SPC is the most commonly used test and is required by the FDA and the state regulatory agencies. The SPC gives an indication of the total aerobic bacteria present in milk. PIC is a test that received acceptance in recent years and is now widely used to estimate number of pyschrotrophic (cold-loving) bacteria. This test gives an indication about the level of on-farm sanitation, holding temperature of milk in the bulk tank, and in general is a reflection of milk production practices on the farm. LPC, also known as thermoduric count, is an estimate of the number of bacteria that can survive laboratory pasteurization at 62.8℃ (143°F) for 30 minutes. This process destroys most of the mastitis causing pathogens, selecting for those bacteria that can survive pasteurization temperatures (thermoduric bacteria).

Mastitis causing organisms in BTM can be classified into one of the two groups: contagious and environmental. The most common contagious organisms are *Staphylococcus aureus (S. aureus:* SA*)*, *Streptococcus agalactiae (Strep. Agalactiae:* SAG*)*, and Mycoplasma species. Among contagious bacteria, SA is most common in the herds and it is present in the BTM samples. The primary habitat of SA is the infected udder. SAG is a highly contagious pathogen, and very quickly spreads through the herd. Cow to cow transmission at the time of milking is responsible for the rapid spread of the organism. Among the environmental organisms, Coagulase negative Staphylococci (CNS), non agalactiae-Streptococci and Streptococci like organisms (NA-SSLO), Coliform cunt (CC) and non Coliform (NC) are analyzed from the BTM sampling.

The BTM data used in this research was collected from 126 dairy herds in 14 counties of Pennsylvania. Four BTM samples were collected at intervals of 15 days from each of the 126 participating dairy producers, which resulted in 504 BTM samples. The samples were analyzed for bulk tank somatic cell counts, bacteriological milk quality and mastitis pathogens. Bacteriological milk quality tests included SPC, LPC, PIC, and CC. Milk samples were examined for mastitis pathogens including SA, SAG, Mycoplasma spp., CNS, NA-SSLO, CC and NC. Information on farm management practices was collected from all of the 126 dairy herds. In addition herd level information and individual cow records were downloaded from the dairy herd improvement association, to ascertain the overall herd performance status. Table 1 shows suggested interpretive criteria for monitoring bulk tank milk and their threshold values respectively.

<Table 1> Suggested interpretive criteria for monitoring bulk tank milk and their threshold values†

Bulk tank	Jayarao et al., (2001)		Britt et al., (1997)		Bray and Shearer (1996)	
Bulk tank somatic cell count (BTSCC) (x 1000)	Low	<250	-	-	Normal	200
	Medium	250-500	-	-	Moderate	300-400
	High	>500	-	-	High	>500
Standard plate count (SPC) (x 10)	Low	<500	Excellent	<500	Normal	<1000
	Medium	500-1000	Acceptable	<1000	Moderate	2000-4000
	High	>1000	Concern	>2000	High	75000
Preliminary incubation count (PIC) (x 1000)	Low	<10	-	-	Normal	<10
	Medium	10-50	-	-	Moderate	20-40
	High	>50	-	-	High	750
Laboratory pasteurization count (LPC)	Low	<100	Excellent	<200	Normal	<1000
	Medium	100-200	Acceptable	400-700	Moderate	1500
	High	>200	Concern	>700	High	>1500
Staphylococcus aureus (SA)	Low	<1	Excellent	0	Normal	0
	Medium	100-500	Acceptable	0	Moderate	100-400
	High	>500	Concern	>50	High	>500
Streptococcus agalactiae (SAG)	Low	<1	-	-	Normal	
	Medium	1000-5000	-	-	Moderate	
	High	>6000	-	-	High	
Coagulase negative Staphylococci (CNS)	Low	<500	Excellent	< 300	Normal	500
	Medium	500-1000	Acceptable	<500	Moderate	600-1000
	High	>1000	Concern	>500	High	>1000
Non agalactiae-Streptococci and Streptococci like organisms (NA-SSLO)	Low	<500	Excellent	<700*	Normal	<500*
	Medium	>500-1000	Acceptable	>1200	Moderate	500-1000
	High	>1000	Concern	>1200	High	>1000
Coliform count (CC)	Low	<10	Excellent	<100	Normal	<500
	Medium	10-50	Acceptable	<400	Moderate	500-1000
	High	>50	Concern	>500	High	>1000
Non coliforms (NC)	Low	<500	--	--	--	-
	Medium	500-1,000	-	-	-	-
	High	>500	-	-	-	-

† Input variables are measured as crisp-set format as shown in the Table 1. Through the fuzzification process, the crisp input variables are transformed into fuzzy set.

2.2 Fuzzy logic based reasoning model

Fuzzy logic (FL) suits well for trading off between significance and precision-something that humans have been managing for a very long time. FL is a convenient way to map an input space to an output space. Sets in FL do not have sharp boundaries, but there is a degree of vagueness. Let μ, $(0 \leq \mu \leq 1)$ represent the membership degree of the set. In the classical set, $\mu_A(x) = 1$ or 0 whether x is a member set A or not. But in a fuzzy subset A, it has the value of $0 \leq \mu_A(x) \leq 1$.

Starting from BTM data described in the previous section, bacteriological test data of SCC, SPC, PIC, LPC, SA, SAG, CC, CNS, NA-SSLO are acquired. Through intensive statistical and empirical analysis, six parameters are chosen as final input parameters. SCC and SPC are basic indicators of milk quality. Higher PIC is a good indicator representing that milk may not be cooled properly or not be held at the right temperature. LPC shows the cleaning status of a herd. Higher CNS is an indicator that more attention should be given to udder preparation before milking. Between two contagious bacteria, SA is chosen as an input parameter as it is frequently found in the BTM sample. Thus, SCC, SPC, PIC, LPC, CNS and SA are the input parameters selected from the analysis.

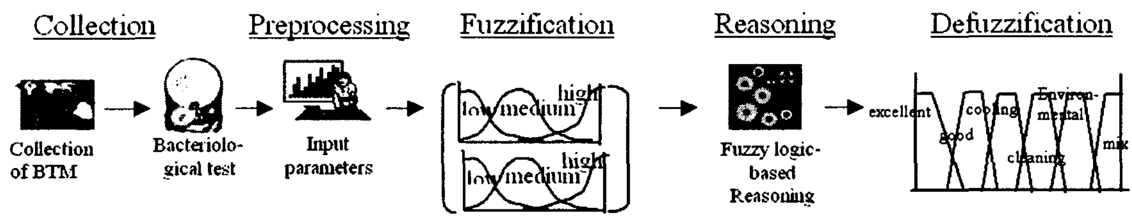
Selected input parameters from the bacteriological test data are fuzzified as linguistic terms like low, medium and high. The output from the fuzzy reasoning engine is defuzzified in order to match with the category of milk quality. The reasoning process based on fuzzy logic is shown in [Fig. 1].

The term 'fuzzification' is adopted meaning that the selected parameter is transformed into a value corresponding to the membership function. The membership function in the FL is assumed to vary between 0 and 1. The function can be an arbitrary curve whose shape can be defined representing the shape of the input parameter. Frequently used membership functions are triangular, trapezoidal, Gaussian, two-sided Gaussian, generalized bell, sigmoidal, and polynomial based curves. In this analysis, two-sided Gaussian and trapezoidal functions are adopted because they suit well with the input parameters and output pattern.

(1) Membership function

The selected input variables should be preprocessed to be used in the fuzzy logic based reasoning system. The mean BTSCC of the 126 dairy producers who participated in the study was 332 (1000 cells/ml). Based on the SCC data collected from dairy herds that participated in the Dairy Herd Improvement during 1996-1997, mean SCC in the US was 310 (1000 cells/ml). When the US was divided into four regions, northeast, southeast, midwest and west, herd SCC was 314, 370, 320 and 271 (1000 cells/ml) respectively. Pennsylvania ranked 20th with a state average of 331 (1000 cells/ml) in the US (Norman et al., 2000).

In order to generate qualitative descriptors, the input variable is analyzed to fit for the fuzzy attribute. Let a fuzzy set x be defined



[Fig. 1] BTM quality reasoning architecture based on fuzzy logic

$x_{(\text{membership variable, attribute})} = \{\mu_1/a_1,\ \mu_2/a_2,\ \mu_3/a_3,\ \cdots\}$

For example, the fuzzy attribute $x_{(\text{scc, low})}$ might be

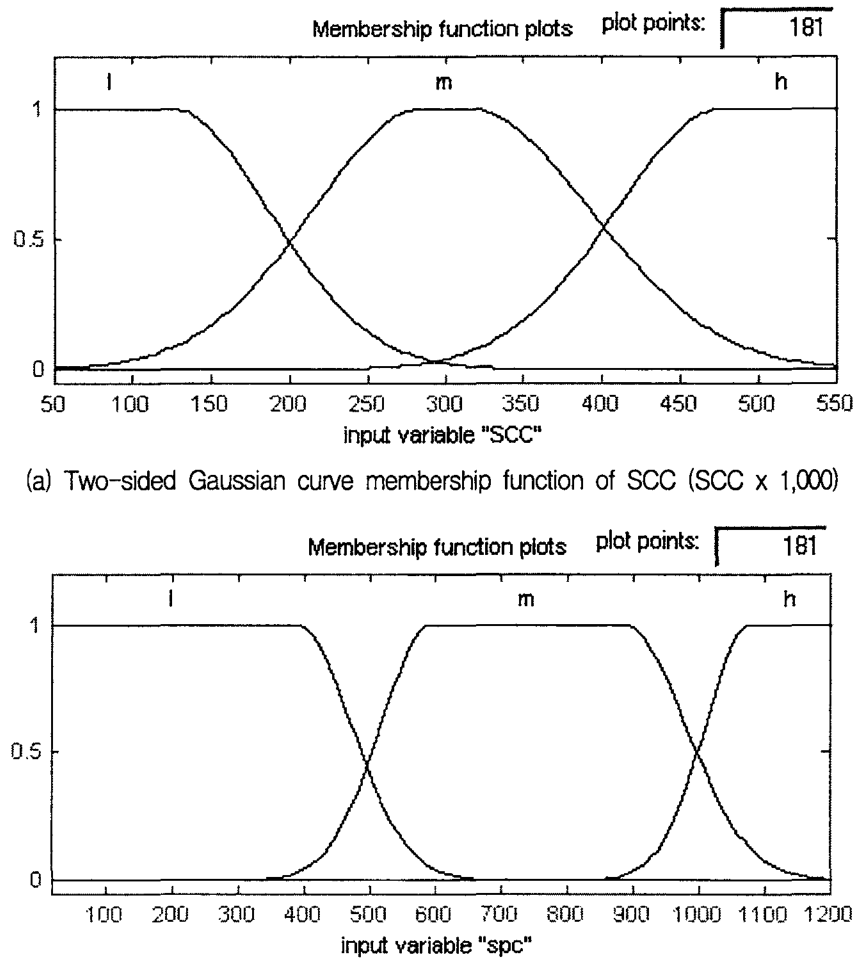
$x_{(\text{scc, low})} = \{1.0/100,\ 0.85/150,\ 0.5/200,\ 0.15/250\}$

Likewise, the medium and high attributes might be

$x_{(\text{scc, medium})} = \{0.15/150,\ 0.50/200,\ 0.85/250,\ 1.0/300,\ 0.85/350,\ 0.5/400,\ 0.15/450\}$

$x_{(\text{scc, high})} = \{0.15/350,\ 0.50/400,\ 0.85/450,\ 1.0/500\}$

Based on this, the membership function of SCC according to the two-sided Gaussian function is modeled as Fig. 2-(a). The threshold values converted into linguistic variables of low, medium



(a) Two-sided Gaussian curve membership function of SCC (SCC x 1,000)

(b) Two-sided Gaussian curve membership function of SPC (SPC x 10)

[Fig. 2] Membership functions for the input variables (SCC and SPC) and high profile of SCC are 200,000 and 400,000 respectively. Also, the membership function of SPC is shown in Fig. 2-(b) with threshold value of 5,000 and 10,000 respectively for low, medium and high profile.

In the original analysis data, maximum value appears very high because the number of bacteria increases sharply when a cow was infected by subclinical or contagious mastitis. Through the cumulative frequency distribution analysis for each herd, the highest value was adjusted to the adjusted maximum value as shown in Table 2. Also, the threshold values for low, medium and high are shown in Table 2, which are referenced from the analysis result in Table 1.

(2) Defuzzification

In order to obtain a single value for the output corresponding to an input $x = (x_1,\cdots,x_n)$, the membership function $\mu(y)$ needs to be transformed into a real number $D(C)$. This process is referred as defuzzification procedure. In this procedure, there are extreme value strategy and centroid strategy. The extreme value strategy uses extreme values of the membership function (generally the maximum) to define the crisp equivalence value. In this method, the used information is very limited, because it takes only extreme values. In the centroid strategy, more information is available. The "center of area" and "center of gravity (COG)" correspond to the centroid strategy. The COG method means weighted average and has a distinct geometrical meaning. In the statistics, this corresponds to the expected value of probability defined as

$$D(C) = \frac{\int y\mu(y)dy}{\int \mu(y)dy}$$

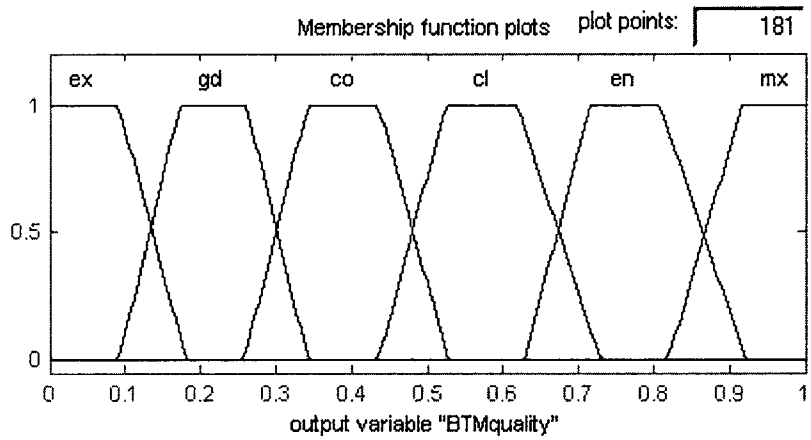
A chosen defuzzification procedure is used to produce an output y^* as

$$y^*(x_1,\cdots,x_n) = D(C)$$

The whole procedure is known as "combine-then-defuzzify" strategy. First, combine

<Table 2> Preprocessed data; original, adjusted and threshold value

	SCC	SPC	PIC	LPC	CNS	SA
	(x 1000)	(x 10)	(x 1000)			
Original maximum	738	3,572	140	6,400	15,175	275
Adjusted maximum maximum maximum maximum	550	1,200	30	400	1,500	120
Adjusted minimum	50	18	1	5	60	0
Threshold; Lower	200	500	10	100	500	100
Threshold; Upper	400	1,000	20	200	1,000	-



[Fig. 3] The output membership function for defuzzifying BTM quality

(From the Fig. each category of ex, gd, co, cl, en and mx corresponds to excellent, good, cooling, cleaning, environmental and mixed problems respectively.)

all the rules using fuzzy connectives to obtain an overall fuzzy set, and then defuzzify this fuzzy set by some chosen defuzzification procedure. As an output, BTM quality is classified as six categories; excellent condition, good condition, milk cooling problem, cleaning problem in milking system, environmental or subclinical mastitis problem, and mixed problem of both cleaning and mastitis combined. Fig. 3 shows membership function with trapezoidal shape which defuzzifies output value corresponding to the status of BTM quality.

(3) Reasoning rule

Fuzzy logic is the superset of standard Boolean logic. The truth of any statement in FL depends on the degree. Suppose that a fuzzy subset A of a set U is defined to be a function A: U $\rightarrow$ [0,1]. The operations of union, intersection, and compliment can be defined using membership function in the fuzzy set. Thus compound rules with multiple disjunctive and conjunctive antecedents can be decomposed into simple canonical forms.

$(A \vee B)(x) = \max\{A(x), B(x)\} = A(x) \vee B(x)$
$(A \wedge B)(x) = \min\{A(x), B(x)\} = A(x) \wedge B(x)$
$A'(x) = 1 - A(x)$

Consider FL system having n inputs $x_1 \in X_1, \ldots, x_n \in X_n$ in the input space $X_1, \cdots, X_n$ and output u. Let us suppose the system has r rules, where the jth rule has the form

R^j: "If x_1 is A_1^j, and $\cdots$ and x_n is A_n^j, then u is B^j," $j = 1,2,\cdots,r$

where $A_1, \cdots A_n$ and B are linguistic values defined by fuzzy sets on the ranges (universes of discourse) X and Y, respectively. The if-part of the rule "x_1 is A_1^j, and $\cdots$ and x_n is A_n^j" is called the premise or antecedent, while the then-part of the rule "u is B^j" is called conclusion or consequent. Let $M(x,u)$ be the binary fuzzy predicate "u is a reasonable control for x" and consider a predicate logic of the form

$$A^j(x) \rightarrow \exists u\ (M(x,u) \text{ and } B^j(u)),\ j=1,,2,\cdots r$$

This implies that for every x, if x satisfies the property A^j then there exists a value u which is "reasonable for x, and for which B^j hold. The conditional statement of the form "if x is A then u is B" can be translated into fuzzy binary relation

$$M(x,u) = A(x)\ \triangledown\ B(u)$$

where $\triangledown$ is a t-norm, such as min. The above equation is referred to as Mamdani's rule. The main idea of Mamdani's rule is to describe decision situation by means of linguistic variables and to use these variables an input to control rules.

The rule in the above example does not mean a casual relationship between x and u, nor an implication in the usual sense. It simply conveys the idea that for every x, if x is A^j, then there is output value u that is B^j, and which is reasonable value to use. Initially, a set of decision rule was created using inductive inference which can generate decision tree automatically based on the ID3 algorithm (Quinlan, 1986; Kim and Heald, 1999). Through the ID3 algorithm, the attribute with the highest information gain or greatest entropy reduction is chosen as the test attribute for the current node. This attribute minimizes the information needed to classify the samples in the resulting partitions and reflects the least randomness or impurity in these partitions. Initially more than 30 rules have been generated from the algorithm. Starting from the initial decision rule set, the reasoning rule base for the fuzzy logic was constructed by modifying and revising the initial rule after considering domain expert's knowledge and previous relevant literatures (Jones and Ward, 1990; Nielen et al., 1992; Heald et al., 2000; Jayarao et al., 2001). Domain experts are veterinarians, professionals in veterinary science and dairy science, experts in dairy herd improvement association, and consultants. Examples of reasoning rules representing each output category are illustrated as the following. In the rule structure, L means LOW, M means MEDIUM and H means HIGH, which corresponds to the range shown in Table 2 respectively.

- IF (SCC is L) AND (SPC is L) AND (PIC is L) AND {(LPC is L) OR (CNS is L)} AND (SA is L), THEN (BTM QUALITY is excellent).
- IF (SCC is M) AND (SPC is L) AND CNS is L AND (SA is L), THEN (BTM QUALITY is good).
- IF (SCC is M) AND (SPC is L) AND (PIC is NOT L), THEN (BTM QUALITY is cooling problem).
- IF (SCC is M) AND (SPC is M) AND (LPC is NOT H), THEN (BTM QUALITY is cleaning problem).

- IF (SCC is M) AND (SPC is M) AND (LPC is H), THEN (BTM QUALITY is environmental mastitis problem).
- IF (SCC is M) AND (SPC is H), THEN (BTM QUALITY is mixed problem of mastitis and cleaning).

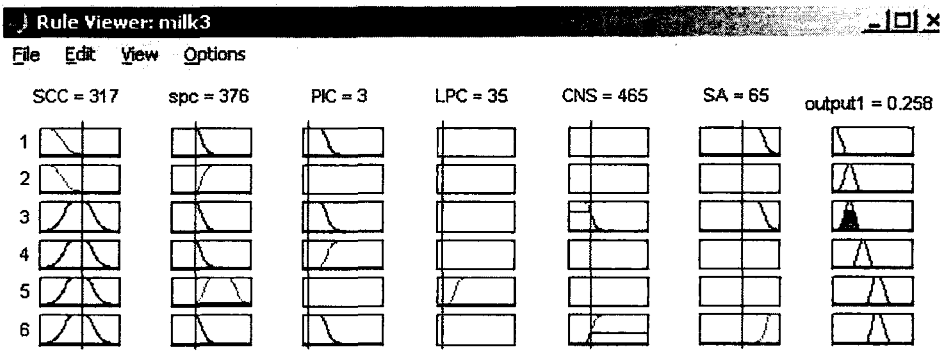
A few of the rules described above are graphically shown in the MATLAB as Fig. 4. As an illustrative example of fuzzy calculation, suppose set of input data corresponding to {SCC, SPC, PIC, LPC, CNS, SA} are given as {317, 376, 3, 35, 465, 65}. For this input, the third rule in Fig. 4 is fired among the reasoning rule base, which resulted in 0.258 as output value. From the defuzzication process, this value corresponds to the good ('gd') category.

After constructing the fuzzy rules, the initial model was evaluated using 50 random sample data from the original 504 test day data. For the final test of the model, 126 herd data which were made by averaging 504 test day data, were used. The 50 sample data adopted for the model evaluation would not influence the reasoning results because 126 data were acquired by averaging the test day data for each farm.

3. Results and discussion

3.1. Implementation

A fuzzy logic based reasoning system utilizing BTM data was modeled and tested for evaluating BTM quality in this study. To automate the modeling and implementation process from the set of BTM data, MATALB (2003) version 6.0 and Fuzzy Logic Toolbox were adopted. This software facilitates modeling of the membership function and creating of the reasoning rules. The reasoning methodology in this software follows the Mamdani's rule described in the previous section. The GUI function showing 3D view between two input variables and a decision variable helps un-



[Fig. 4] Graphical rule structure for reasoning BTM quality on the Fuzzy Logic Tool box

(For the input data of SCC=317,···, SA=65, the fuzzy logic output resulted in 0.258 by firing rule 3 in the rule base. This corresponds to good ('gd') category in Fig. 3.)

<Table 3> Confusion matrix between expert's evaluation and fuzzy results

Fuzzy result	mx					4	42
	en				1	19	4
	cl	1	1	1	5	5	
	co	1	4	8	2		1
	gd	4	19				
	ex†	4					
		ex	gd	co	cl	en	mx

Domain expert's evaluation

† Six categories correspond to excellent herd (ex), satisfactory good herd (gd), cooling problem (co), cleaning problem in milking system (cl), environmental or subclinical problem (en) and mixed problem of both cleaning and mastitis (mx).

derstanding of the relationship among variables. Using selected test data composed of SCC, SPC, PIC, in combination with other milk quality measures and contagious and environmental mastitis counts, the fuzzy logic system was evaluated to validate the overall performance of the model.

After validation, the model was implemented for the preprocessed BTM data as a batch mode. The results from fuzzy logic reasoning were compared with those from domain experts. Upon analysis of BTM data in the microbiological laboratory, veterinarians reviewed the results and evaluated the quality of BTM. Using confusion matrix, this evaluation results were compared with those from fuzzy logic reasoning, which was shown in Table 3. This defuzzification procedure was based on the functions shown in Fig. 3. As the BTM quality was grouped into six categories, the domain expert's evaluation and fuzzy results were grouped into six categories. Among all the surveyed 126 herds, correct reasoning ratio resulted in 77%. Considering six categories adopted in this research, the correct reasoning ratio appeared satisfactory. Suppose six categories are reduced into four categories by combining similar groups such as 'ex' with 'gd' and 'co' with 'cl.' The correct reasoning ratio increases to 83% which is also high prediction ratio.

3.2. Discussion

Various factors need to be examined to achieve sufficient information in the areas of the number of rules, the choice of parameters, fuzzy membership functions, and the logical connectives. In the choice of these nominal parameters, robustness properties should be considered. The description of reasoning knowledge in linguistic rules involves linguistic variables whose values should form fuzzy partitions of associated spaces. Thus, fuzzy partitioning of both input and output spaces enables some rule in the rule base fire with some positive degree.

More specific, the test results were examined deeply using sensitivity and specificity as given in Nielen et al. (1992), who has applied electronic conductivity for the mastitis detection. Table 4 shows the sensitivity and specificity for six category group calculated from <Table 3>.

<Table 4> Sensitivity and specificity of fuzzy logic for reasoning BTM quality

	Actual (domain expert's evaluation)												
Prediction (fuzzy logic reasoning)	Categories	Excellent (Ex) †		Good (Gd)		Cooling (Co)		Cleaning (Cl)		Environmental (En)		Mixed (Mx)	
		Ex	Other	Gd	Other	Co	Other	Cl	Other	En	Other	Mx	Other
	Group A	4	0	19	4	8	8	5	8	19	5	42	4
	Group B	6	116	5	98	1	109	3	110	9	93	5	75
	Total	10	116	24	102	9	117	8	118	28	98	47	79
	Sensitivity	.4		.79		.89		.63		.68		.89	
	Specificity		1.0		.96		.93		.93		.95		.95

† Six categories are the same as shown in Table 3. The term 'ex' means excellent herd, while the term 'other' corresponds to the herd group that was not classified as an excellent herd by the domain expert. Group A refers to dairy herds that were classified as excellent / good/ cooling/cleaning/ environmental / mixed. Group B consisted of dairy herds that were not classified into group A.

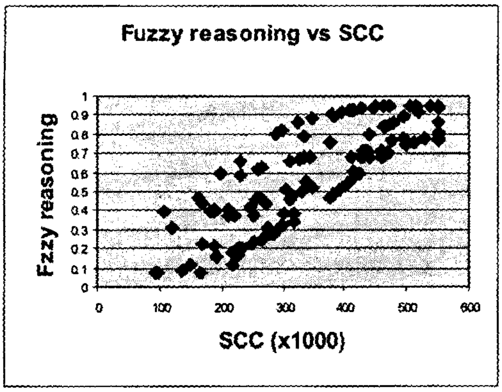
Sensitivity ranges from 0.4～0.894, and specificity ranges from 0.932～1.0. Considering there are six categories, and the size of category is small, the sensitivity value is satisfactory. Especially, specificity is very high, which means that there is much less possibility to predict wrong from another category herd.

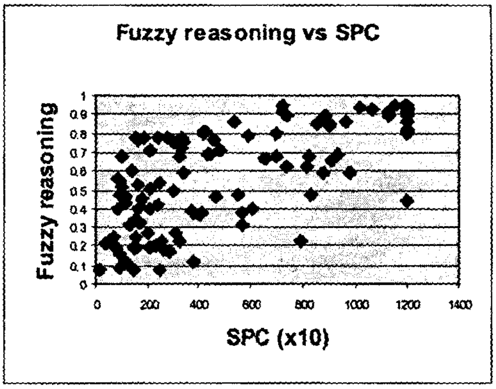
With the complexity of the given problem having six variables, five of them with three cases (low, medium, high) and one with two cases, a total of 486 ($3^5 \times 2$) rules are required to cover all the cases enumeratively. Previous researches show that SCC and SPC are important variables in evaluating mastitis in milk quality (Jones and Ward, 1990; Berning and Shook, 1992; Jayarao et al., 2001). In establishing the decision rules, as SCC and SPC are variables which can be fired earlier than the other variables, the number of rules can be reduced compared to the enumerative method. In the given problem, the sequence of input variables *does not influence the result.* Consider the decision rule for the BTM quality problem have n inputs x_1, x_2, ⋯, x_n and one output u with the following form.

IF x_1 is A_1 and ⋯ x_n is A_n THEN u is B_1.

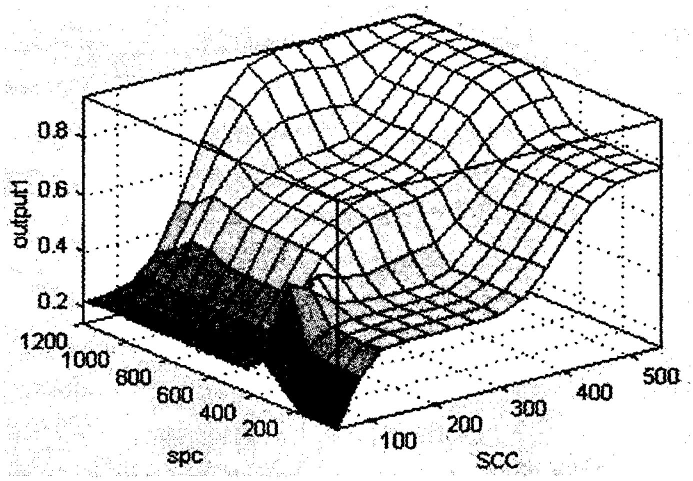
This can be rewritten as "If x is A then u is B_1" where A is the fuzzy Cartesian product of the A_i's characterized by $A(x_1, x_2, \cdots, x_n) = \wedge \{A_i(x_i), i=1,2,\cdots,n\}$. In this expression, the output from the Cartesian product is irrelevant of its sequence.

When *a priori* information are not available, all the possible combination of rules should be generated and evaluated. In the Al-Faraj et al (2001)'s case of crop water stress index, 150 rules were generated enumeratively and used from three variables with ten, five and three attributes respectively. When creating rule structure, a multiple-antecedent single-consequent rule can always be considered as a group of multiple-input single-output rules (Lee, 1990). It is possible to cast non-obvious rules into the form of above men-





[Fig. 5] Relationship between fuzzy reasoning vs SCC and SPC



[Fig. 6] 3D surface among SCC, SPC and reasoning output

tioned rule structure. The case of non-obvious rules are incomplete IF rules, mixed rules, fuzzy statement rules, comparative rules, unless rules, and quantifiers rules based on Mendel (2001).

In order to compare the fuzzy output with domain expert's evaluation, a mid-point value is assigned for the category's nominal value. When the results from fuzzy logic reasoning were compared with those from veterinarian's evaluation results, correlation coefficient R^2 resulted in 0.90 based on simple regression between the two outputs. The R^2 value in Al-Faraj et al. (2001)'s crop water stress index shows 0.70 for chamber tests and 0.42 or 0.51 for greenhouse tests. This implies that fuzzy logic based reasoning system for the BTM quality provides a good reference system possessing similar reasoning capabilities with human expert. Advantage of fuzzy reasoning system is that they present qualitative information based on linguistic terms which is similar with the language used by the veterinarians, advisors and producers. This means that they need not know the exact number of those variables. The system automatically transforms the numerical data into membership function and generates meaningful output. With the implementation issues, type of membership function or critical value can be easily modified interactively, thus changing the output value. This means that the system can be continuously adapted to fit for a specific domain or production area.

As SCC and SPC are key input parameters, the relationship between fuzzy reasoning output and SCC/SPC are analyzed how it is related to the reasoning output as shown in Fig. 5. For the SCC, the fuzzy reasoning results appear to be highly dependent on the value of SCC. In case of SPC, the reasoning output shows mixed results from excellent to environmental cases with the range of SPC <5000. For the higher values of SPC with SPC> 5000, most of the fuzzy output shows environmental mastitis or mixed problem of mastitis and cleaning issues. These results are similar with the evaluation output from domain experts such as veterinarians or field consultants (Jayarao et al., 2001). Fig. 6. illustrates 3D surface view showing the relationship between SCC, SPC and reasoning output.

Fuzzy logic based reasoning is a generic technology with many application fields and methodologies. With regard to application areas, crop management planning, dairy farm management, pruning system in vineyard, climate controller in potato bulk store, managing time-varing temperature data, analysis of milk yield, vision system for seeds germination and visual tomato quality grading can be mentioned to name only a few in the agricultural areas. From a methodological point of view, algorithmic applications, information processing, knowledge applications and hybrid application can be suggested. Especially recent development can primarily be found in the combinations of fuzzy technology with artificial neural networks, with genetic algorithm, with evolutionary strategies and with chaos theory.

Next research will focus on the implementation of the fuzzy logic based decision aid tool on the web environment. After the milk producers or extension specialists log on the web server, they can update their BTM test data on their own herd database. The web server shows them the evaluation results based on fuzzy logic at their computer through web database simultaneously. Through this process, the historical data will be accumulated, which adds the accuracy of the model's predictability and realizing the real time decision aid.

4. Conclusions

The demand of consumers for safe and high quality milk has placed a significant responsibility on dairy producers, retailers and manufacturers to produce and market high quality milk and milk products. The first step in the production of quality milk begins at the dairy farm; therefore, the responsibility lies with the dairy producer to produce raw milk under the strictest hygienic standards. All dairy producers recognize the fact that production of quality milk and lowered incidence of mastitis will result in increased returns for the milk produced.

Fuzzy logic based reasoning system has been proposed and tested for evaluating BTM quality. The quality of milk was classified as excellent, good quality, cooling problem, cleaning problem, subclinical mastitis problem, and mixed problem of mastitis and cleaning. A set of reasoning rules was constructed based on the inference mechanism. The reasoning accuracy from the confusion matrix comparison between fuzzy logic and domain expert showed 77%. When the six categories were reduced into four by combining similar categories, the reasoning accuracy increased into 83%. This implies that the fuzzy reasoning provides satisfactory output compared with the domain expert's evaluation. More specific, the test results were examined deeply using sensitivity and specificity. The values for each category showed higher except for one sensitivity with small group size. Overall, considering the number of group is six and each group has small size, the fuzzy logic reasoning showed good results.

Easily explained and understood fuzzy logic system could lead to faster adoption of these decision aids by dairy farmers, extension specialists, dairy health consultants, and veterinary health professionals in the future. Additionally, fuzzy logic based diagnostic rules can be implemented and automated on the web environment through the adoption of the extensible markup language and web service technology. Considering this, fuzzy logic based reasoning shows a good potential as an effective decision aid in the areas of dairy quality management and food safety, where quality and safety become more and more important issue.

Acknowledgements

This research has been partly supported by Brain Busan 21 (BB21) Program from Busan Metropolitan City, Korea. Also, this research has been partly supported by the Pennsylvania State Extension Program Priority Initiative mini grant under the Dairy Profitability Initiative, and USDA-Animal Health Formula Funds. The authors would acknowledge anonymous referees who have made a very valuable comments and recommendations.

References

[1] Al-Faraj, A., Meyer, G.E., Horst, G.L., "A crop water stress index for tall fescue irrigation decision-making - a fuzzy logic method," *Computers and Electronics in Agriculture,* Vol. 32,(2001), 69-84.

[2] Berning, L.M., Shook, G.E., "Prediction of mastitis using milk somatic clell count, N-Acetyl-B-D-Glucosaminidase and Lactose," *Journal of Dairy Science,* Vol. 75, (1992),

1840-1848.

[3] Bray, D. R., Shearer, J.K., "*Trouble-shooting a mastitis problem herd,*" Univ. of Florida cooperative extension circular 1162, Gainesville, Florida, 1996.

[4] Britt, J. S., Hartmann, F., Reinemann, D., "Use of microbiology and strategic site sampling at strategic times to solve high bacteria count problems in bulk tank milk, An actual case history," *Proc. 36th Natl. Mastitis Council*, (1997), 80-90.

[5] Bridges, S.M., Higginbotham, C., Mckinion, J.M., Hodges, J.E., "Fuzzy descriptors of time-varying data: theory and application," *AI appication*. Vol. 9, no. 2, (1995), 1-14.

[6] Dybowski, R., Gant, V., "Artificial neural networks in pathology and medical laboratories," *The Lancet*, Vol. 34. no 4, (1995), 1203-1207.

[7] Godkin M.A., Leslie, K. E., "Culture of bulk tank as a screening test: a brief review," *Canndian Veerinaryt Journal,* Vol. 34, (1993), 601-605.

[8] Heald, C.W., Kim, T., Sischo, W.M., Cooper, J.B., Wolfgang, D.R., "A Computerized mastitis decision aid for farm production records: A Neural network approach," *Journal of Dairy Science,*. Vol. 83, no. 4, (2000), 711-720.

[9] Hogeveen, H., Noordhuizen-Stassen, E.N., Tepp, D.M., Kremer, W.D.J., Van Vliet, J.H., Brand A., "A Knowledge-based system for diagnosis of mastitis problems at the herd level. 1 concept," *Journal of Dairy Science*, Vol. 78, (1995), 1430-1440.

[10] Jahns, G., Nielsen, H. M., Paul, W., "Measuring image analysis attributes and modeling fuzzy consumer aspects for tomato quality grading," *Computers and Electronics in Agriculture,* Vol. 31, (2001), 17-29.

[11] Jayarao, B. M., Pillai, S. R., Wolfgang, D. R., Griswold, D. R., Hutchinson, L. J., "Herd level information and BTM analysis: tools for improving milk quality and heard udder health," *The Bovine Practitioner* Vol. 35, (2001), 23-35.

[12] Jones, G.F., Ward, G.E., "Evaluation of a scheme for predicting the gram-staining reaction of organisms causing bovine mastitis," *Journal of American Veterinary Medicine Association,"* Vol. 196, no. 4, (1990), 597-599.

[13] Kim, T., Heald, C.W., "Inducing inference rules for the classification of bovine mastitis," *Computers and Electronics in Agriculture,* Vol. 23, (1999), 27-42.

[14] Kirk, J.H., "Subclinical mastitis and somatic cell counts," *Working paper, University of Califomai, Davis. Veterinary medical teaching and research ccenter.* Tulare, CA, (2003).

[15] Klaus, G., Nagy, L., Farkas, I., "Improved climate control for potato stores by fuzzy controllers," *Computers and Electronics in Agriculture,* Vol. 40, (2003), 127-140.

[16] Lacroix, R., Huijbers, J., Tiemessen, R., Lefebvre, D., Marchand, D., Wade, K. M., "Fuzzy set-based analytical tools for dairy herd improvement," *Applied. Engineering in Agriculture,* Vol. 1 4, no. 1, (1998), 79-85.

[17] Lee, C., "Fuzzy logic in control systems: Fuzzy logic controller," Part II. *IEEE Trans. On SMC,* Vol. 20, (1990), 419-435.

[18] Marshall, R. T., *Standard methods for the examination of dairy products.* 16th edition. American public health association. Washington, DC, 1992.

[19] Matlab, 2003, http://www. mathworks.com.

[20] Mendel, J. M., Uncertain rule-based fuzzy

logic systems. Prentice Hall PTR, Upper Saddle River, NJ, 2001.

[21] Mickelson A., Hansen, L., Morris, N., "The impact of environmental mastitis on milk quality in the Pacific northwest," *Proc. 37th Natl Mastitis Council*, (1998), 26-34.

[22] Nielen, M., Deluyker, H., Schukken, Y.H., Brand, A., "Electrical conductivity of milk: measurement, modifiers and meta analysis of mastitis detection performance," *Journal of Dairy Science*, Vol. 75, (1992), 606-614.

[23] Nielen, M., Schukken, Y.H., Van der Broke, J., Brand, A., "Relations between on-ine electrical conductivity and daily milk production on a low somatic cell count farm," *Journal of Dairy Science,* Vol. 76, (1993), 2589-2596.

[24] Norman, H.D., Miller, R.H., Wright, J.R., "Herd and state means for somatic cell count from dairy herd improvement," *Joyrnal Dairy Science* Vol. 83, (2000), 2782-2788.

[25] Pösö, J., Mäntysaari, E.A., "Relationships between clinical mastitis, somatic cell score, and production for the first three lactations of Finnish ayrshrine," *Journal of Dairy Science,* Vol. 79, (1996), 1284-1291.

[26] Quinlan, J.R., "Induction of decision trees," Machine Learning. Vol. 1, (1986), 81-106.

[27] Ross, T.J., "Fuzzy logic with engineering applications," McGraw-Hill, Inc., New York, NY, 1995.

[28] Salehi, F., Lacroix, R., Wade, K.M., "Development of neuro-uzzifiers for qualitative analysis of milk yield," Computers and Electronics in Agriculture, Vol. 28, (2000), 171-186.

[29] Ureña, R., Rodriguez, F., Berenguel, M., "A Machine vision system for seeds germination quality evaluation using fuzzy logic," Computers and Electronics in Agriculture, Vol. 32, (2001), 1-20.

[30] Zadeh, L. "A Fuzzy sets," Information and Control, Vol. 8, (1965), 338-353.

[31] Zimmermann, H. J., Fuzzy set theory - and its applications, fourth ed., Kluwer Academic Publishers, Boston, MA, 2001.

요약

Bulk tank milk의 품질평가를 위한 퍼지기반 추론

김태운* · 정대유* · Bhushan M. Jayarao***

우유생산 농가에서는 그들 젖소의 우유를 저장하는 탱크 (bulk tank milk: BTM)로부터 채취된 샘플로부터 분석된 우유에 대한 품질관련 항목들, 즉 체세포 수 (somatic cell count: SCC), 표준 plate count (standard plate count: SPC), 사전 incubation count (preliminary incubation count: PIC) 등에 관한 정보를 정기적으로 제공 받는다. 이러한 정보는 일정기간 쌓이게 되면 우유의 품질을 유지하고 목장을 관리할 수 있는 중요한 지식 베이스가 될 수 있다. 그러나 우유품질이나 목장의 관리상태를 평가하는 기준은 모호하고 퍼지한 용어로 많이 표현되고 있다. 즉 우유 품질을 최상급, 상급, 중간, 불량으로 표시하거나 목장의 관리상태를 아주 양호, 양호, 미흡 등으로 표시한다. 이러한 서술방식은 퍼지이론에서의 모호한 상태를 표현하는 기준과 많이 부합되고 있다. 본 연구의 목적은 BTM으로부터 추출한 샘플로부터 미생물학적 분석을 통해서 나온 결과를 이용해서 BTM 의 품질과 목장의 관리상태에 대하여 추론하는 것을 목표로 하고 있다. 따라서 퍼지추론엔진에 기초하여 퍼지로직 기반의 추론방법을 개발하고 실제 데이터를 이용해서 평가하였다. 입력 데이터로는 Bulk Tank SCC, SPC, PIC, laboratory pasteurization count (LPC), non agalactiae Streptococci, Streptococci like organisms, *Staphylococcus aureus*등이다. 이러한 입력자료에 근거하여 BTM 의 품질상태를 아주 양호, 양호, cooling 문제, 청결문제, 환경적 mastitis, 환경적/청결 복합문제로 분류하고, 낙농가로부터 채취한 실제 데이터를 이용하여 추론하였다. 본 퍼지 추론 결과는 낙농생산자, 컨설턴트, 수의사 등 관련 종사자들에게 의사결정을 위한 참고자료로서 활용이 가능하다.

* 경성대학교 산업공학과
** 펜실베니아 주립대, 수의과학과