

훼손된 시계열 데이터 분석을 위한 퍼지 시스템 융합 연구

김동원

인하공업전문대학 디지털전자과

Fused Fuzzy Logic System for Corrupted Time Series Data Analysis

Dong W. Kim

Dept. of Digital Electronics, Inha Technical College

요약 본 논문에서는 노이즈에 의해 훼손된 시계열 데이터의 모델링에 대하여 다룬다. 모델링 기법으로, 논싱글톤 퍼지 시스템을 사용한다. 논싱글톤 퍼지 시스템의 주요특징은 미지의 비선형시스템의 입력이 퍼지값으로 모델링 된다는데 있다. 그러므로 퍼지시스템에 인가되는 학습데이터나 입력데이터 등이 노이즈나 외부 환경에 의해 변형된 경우에 매우 유용하게 적용될 수 있다. 성능비교를 위해 벤치마크 데이터로 잘 알려진 Mackey-Glass 데이터를 사용한다. 이들 데이터 모델링을 통하여 결과를 비교, 분석하여 논싱글톤 퍼지시스템이 잡음에 대하여 보다 강인하고 효율적임을 본 논문에서 보인다.

주제어 : 시계열, 데이터, 퍼지 시스템, 융합, 분석

Abstract This paper is concerned with the modeling and identification of time series data corrupted by noise. As modeling techniques, nonsingleton fuzzy logic system (NFLS) is employed for the modeling of corrupted time series. Main characteristic of the NFLS is a fuzzy system whose inputs are modeled as fuzzy number. So the NFLS is especially useful in cases where the available training data or the input data to the fuzzy logic system are corrupted by noise. Simulation results of the Mackey-Glass time series data will be demonstrated to show the performance of the modeling methods. As a result, NFLS does a much better job of modeling noisy time series data than does a traditional Mamdani FLS.

Key Words : Time series, data, fuzzy system, fusion, analysis

1. 서론

시계열 데이터는 어떤 특정대상의 시간적 변동을 지속적으로 관찰하여 얻은 데이터의 계열이다. 우리 주변에서 많은 시계열 현상들을 접하게 되는데, 가격-소비 등의 경제현상, 기온-강수량 등의 자연현상, 주식시장, 인구증가, 날씨 등을 대표적인 예로 들 수 있다. 이러한 시

계열 데이터의 가장 기본적인 특징은 관측된 데이터들끼리의 종속성이다. 일반적으로 말하면, 측정된 데이터 계열의 과거값은 미래값에 영향을 미친다. 시계열 데이터를 동정하고 모델링하는 이유는 이와 같은 비선형 관계를 표현되는 종속성을 분석하고, 변화를 일으키는 배경을 이해하며, 미래값의 변동을 예측하는데 있다[1]. 그러나 시계열 데이터의 고유한 특성으로 인해 시계열데이터

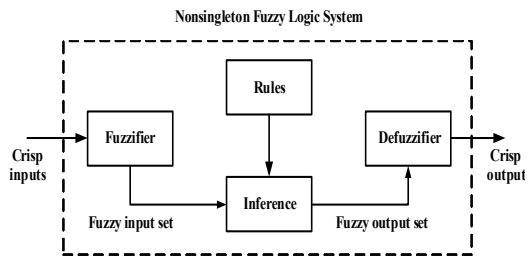
*교신저자 : 김동원(dwnkim@inhac.ac.kr)

접수일 2018년 01월 25일 수정일 2018년 03월 05일 심사완료일 2018년 03월 10일

분석은 어려운 작업으로 분류된다. 또한 데이터 측정에서 발생하는 불명확성은 피할 수 없는 한계이다. 학습데이터가 측정노이즈에 의해 훼손되었을 때 불명확성은 반듯이 발생된다. 최근 들어 이러한 불명확성을 다루기 위해 논싱글톤 퍼지로지 시스템이 활발히 연구되고 이용되고 있다[1-5]. 논싱글톤 퍼지로지 시스템은 입력값이 퍼지값으로 모델링되는 특징을 가진 퍼지 시스템이다. 이러한 특징으로 인해 입력데이터가 노이즈에 의해 변형된 경우에 매우 효율적이며 다양하게 응용되었다 [6-10]. 그러나 여전히 그 성능을 평가받지 못한 곳이 있으며 특히 노이즈가 가미된 시계열 데이터의 동정과 관련된 응용성에 관한 연구가 상당히 요구된다. 본 논문에서는 논싱글톤 퍼지시스템을 이용하여 노이즈에 의해 훼손된 시계열 데이터의 모델링 [10-15]을 수행하고 이를 분석한다. 또한 성능을 확인하기 위해 일반적인 싱글톤 퍼지시스템과 성능을 비교 분석하고 논싱글톤 퍼지시스템이 잡음에 보다 강인하고 효율적임을 보인다.

2. 논싱글톤 퍼지시스템

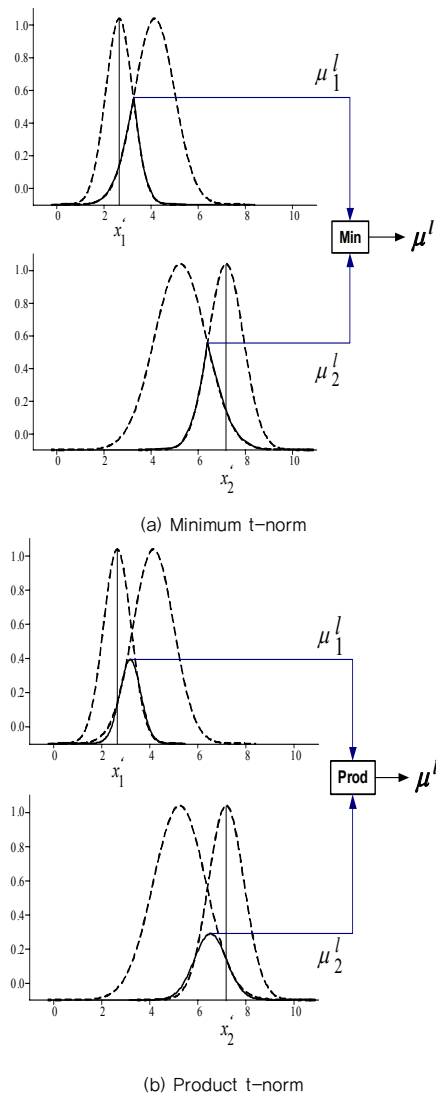
훼손된 시계열데이터에 사용된 논싱글톤 퍼지시스템의 구조를 간략히 살펴보면 [Fig. 1.]과 같다. 구체적인 내용은 [2,3]에서 자세히 확인할 수 있다.



[Fig. 1] Structure of nonsingleton fuzzy system

[Fig. 1.]에서 보듯이 크리스프한 입력값은 일단 퍼지 입력값으로 퍼지화 되고 퍼지추론기에서 활성화 된다. 논싱글톤 퍼지시스템은 일반적인 퍼지시스템과 같이 퍼지집합, if-then 규칙, 퍼지추론의 개념에 기반하고 있으며, 그 구조는 동일하다. 하지만 유일한 차이점은 퍼지화기에 있다. 싱글톤 퍼지시스템과는 다르게 입력값은 퍼지값으로 변화되며 이 퍼지값이 퍼지추론기의 새로운 입력값으로 사용된다.

퍼지값은 멤버쉽함수, μ_x , 로 볼 수 있으며 이것은 [0,1]사이의 값으로 맵핑된다. 따라서, $x \in U$ 일때 $\mu_x : U \rightarrow [0, 1]$ 이다. 그러므로 논싱글톤 퍼지화기는 x_i 를 원소로 가진 퍼지집합 X 로 x 값을 매핑시킨다. μ_x 는 $x_i = x$ 일때 가장 큰 퍼지값을 가지며, 두 값이 점점 멀어질때 점점 작은 퍼지값을 가지게 된다. 다시 말해, 측정값 x_i 는 논싱글톤 퍼지시스템에서 퍼지값으로 맵핑된다. 여기서, 논싱글톤 퍼지시스템의 입력과 퍼지규칙의 전반부 멤버쉽함수사이의 t-norm 연산을 그림으로 나타내면 [Fig. 2.]와 같다.

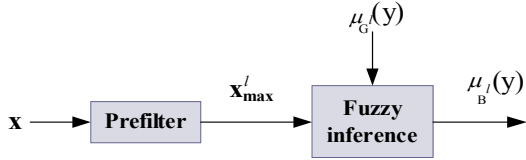


[Fig. 2] t-norm operation in nonsingleton fuzzy system

M개의 규칙을 가진 퍼지시스템을 고려해보면 다음과 같다. 여기서, l 번째 규칙은 R^l 이며, 규칙의 전반부와 후반부 멤버십 함수는 각각 n 개와 1개이다. 식 (1)과 같은 퍼지시스템을 논싱글톤 퍼지시스템으로 표현하면 [Fig. 3].과 같다. 입력 X 는 전처리과정을 거쳐 X_{\max}^l 로 변환되고 퍼지추론의 새로운 입력값이 된다. 이러한 전처리 과정을 거침으로써 초기데이터 측정시의 불확실성을 잘 처리할 수 있다. [Fig. 3].에서 나타난 출력값 $\mu_{B^l}(y)$ 는 다음과 같다. 여기서, T 와 $*$ 는 각각 minimum (product) t-norm을, 식 (2)와 (4)에서 각괄호는 각 규칙의 적합도를 나타낸다.

$$R^l : \text{IF } x_1 \text{ is } F_1^l \text{ and } x_2 \text{ is } F_2^l \text{ and } \dots x_n \text{ is } F_n^l$$

Then y is $G^l \quad l=1, \dots, M$ (1)



[Fig. 3] Prefiltering of the nonsingleton fuzzy system

$$\mu_{B^l}(y) = \mu_{G^l}(y) * \{ [\sup_{x_1 \in X_1} \mu_{X_1}(x_1) * \mu_{F_1^l}(x_1)] * \dots * [\sup_{x_n \in X_n} \mu_{X_n}(x_n) * \mu_{F_n^l}(x_n)] \}, \quad y \in Y \quad (2)$$

$$\mu_{Q_k^l}(x_k) \equiv \mu_{X_k}(x_k) * \mu_{F_k^l}(x_k) \quad (3)$$

$$\mu_{B^l}(y) = \mu_{G^l}(y) * [T_{k=1}^n \mu_{Q_k^l}(x_{k,\max}^l)] \quad (4)$$

모든 멤버십 함수는 가우시안형태를 사용하고, product t-norm을 사용했을때, k 번째 입력값과 전반부 멤버십함수는 각각 식 (5)와 (6)과 같으며, 식 (7)~(11)에 의해 각 규칙의 파라미터들이 학습된다.

$$\mu_{X_k}(x_k) = \exp\left(-\frac{1}{2} \left[\frac{(x_k - m_{X_k})^2}{\sigma_{X_k}^2} \right]\right) \quad (5)$$

$$\mu_{F_k^l}(x_k) = \exp\left(-\frac{1}{2} \left[\frac{(x_k - m_{F_k^l})^2}{\sigma_{F_k^l}^2} \right]\right) \quad (6)$$

$$y(\mathbf{x}^{(i)}) = f_{ns}(\mathbf{x}^{(i)}) = \frac{\sum_{l=1}^M \bar{y}^l \prod_{k=1}^n \mu_{Q_k^l}(x_{k,\max}^{l,(i)})}{\sum_{l=1}^M \prod_{k=1}^n \mu_{Q_k^l}(x_{k,\max}^{l,(i)})}$$

$$= \frac{\sum_{l=1}^M \bar{y}^l \prod_{k=1}^n \exp\left(-\frac{1}{2} \left[\frac{(x_k^{(i)} - m_{F_k^l})^2}{\sigma_X^2 + \sigma_{F_k^l}^2} \right]\right)}{\sum_{l=1}^M \prod_{k=1}^n \exp\left(-\frac{1}{2} \left[\frac{(x_k^{(i)} - m_{F_k^l})^2}{\sigma_X^2 + \sigma_{F_k^l}^2} \right]\right)} \quad (7)$$

$$m_{F_k^l}(i+1) = m_{F_k^l}(i) - \alpha_m [f_{ns}(\mathbf{x}^{(i)}) - y^{(i)}] \times$$

$$[\bar{y}^{(i)} - f_{ns}(\mathbf{x}^{(i)})] \times \left[\frac{x_k^{(i)} - m_{F_k^l}(i)}{\sigma_X^2(i) + \sigma_{F_k^l}^2(i)} \right] \phi_l(\mathbf{x}^{(i)}) \quad (8)$$

$$\bar{y}(i+1) = \bar{y}(i) - \alpha_{\bar{y}} [f_{ns}(\mathbf{x}^{(i)}) - y^{(i)}] \phi_l(\mathbf{x}^{(i)}) \quad (9)$$

$$\sigma_{F_k^l}(i+1) = \sigma_{F_k^l}(i) - \alpha_{\sigma} [f_{ns}(\mathbf{x}^{(i)}) - y^{(i)}] \times$$

$$[\bar{y}^{(i)} - f_{ns}(\mathbf{x}^{(i)})] \times \sigma_{F_k^l}(i) \left[\frac{x_k^{(i)} - m_{F_k^l}(i)}{\sigma_X^2(i) + \sigma_{F_k^l}^2(i)} \right]^2 \phi_l(\mathbf{x}^{(i)}) \quad (10)$$

$$\sigma_X(i+1) = \sigma_X(i) - \alpha_X [f_{ns}(\mathbf{x}^{(i)}) - y^{(i)}] \times$$

$$[\bar{y}^{(i)} - f_{ns}(\mathbf{x}^{(i)})] \times \sigma_X(i) \left[\frac{x_k^{(i)} - m_{F_k^l}(i)}{\sigma_X^2(i) + \sigma_{F_k^l}^2(i)} \right]^2 \phi_l(\mathbf{x}^{(i)}) \quad (11)$$

여기서, $y(X)$, \bar{y}^l , ϕ_l 은 각각 논싱글톤 퍼지시스템의 출력값, 후반부 멤버십함수의 중심, 퍼지기저함수이며, $m_{F_k^l}$, $\sigma_{F_k^l}$, σ_X 는 각각 k 번째 전반부 멤버십함수의 평균값과 표준편차, 그리고 입력변수의 표준편차이다. α_m , α_{σ} , α_X 는 각각 $m_{F_k^l}$, $\sigma_{F_k^l}$, σ_X 의 학습률이다.

3. 시뮬레이션 및 결과고찰

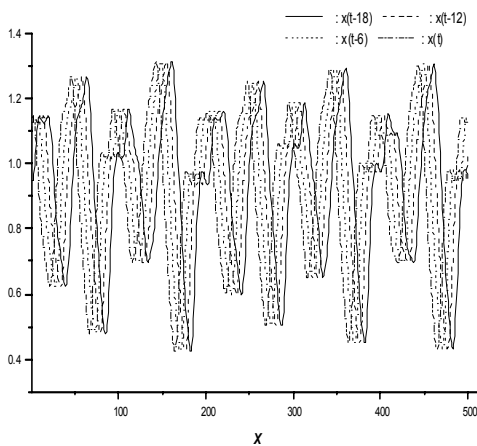
논싱글톤 퍼지시스템의 성능을 평가하기 위해 잘 알려진 벤치마크 문제인 Mackey-Glass 시계열 데이터를 이용한다. 이 데이터는 미래값을 예측하는 것으로 수많은 연구자들에 의해 사용되고 있으며 현재도 많은 연구가 진행되고 있다. Mackey-Glass 미분방정식으로 정의되고 식 (12)와 같다. 데이터 입출력형태는 4입력 1출력이며 식 (13)과 같고, 성능지수는 RMSE로 식 (14)와 같다.

$$\frac{dx(t)}{dt} = \frac{0.2x(t-d)}{1+x^{10}(t-d)} - 0.1x(t) \quad (12)$$

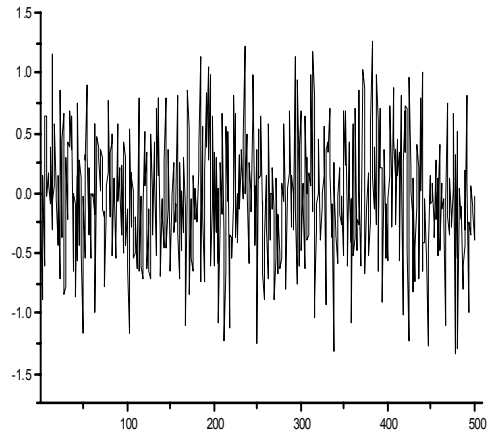
$$[x(t-18), x(t-12), x(t-6), x(t) : x(t+6)] \quad (13)$$

$$PI = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (14)$$

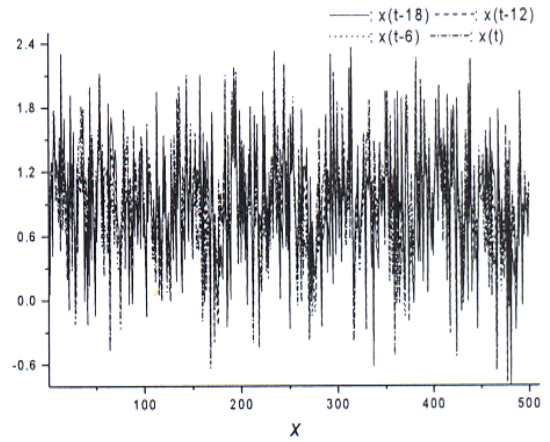
[Fig. 4.5.6].은 각각 노이즈가 없는 원신호로 식 (13)에서 보여준 4입력을 나타내며, 원신호를 훼손시킬 노이즈 신호와 노이즈에 의해 훼손된 Mackey-Glass 시계열 데이터를 보여준다. 여기서 [Fig. 6]의 데이터는 퍼지시스템의 새로운 입력으로 사용된다. [Fig. 7]에서 새로운 입력으로 실행된 싱글톤 퍼지시스템과 논싱글톤 퍼지시스템의 결과를 보여준다. 여기서 각 입력변수는 2개의 멤버쉽함수가 할당되었으며 학습률은 0.1485이다. 그림에서 보듯, 논싱글톤은 노이즈에 대한 강한 적응력을 확인할 수 있다. 이때 RMSE는 0.096이다.



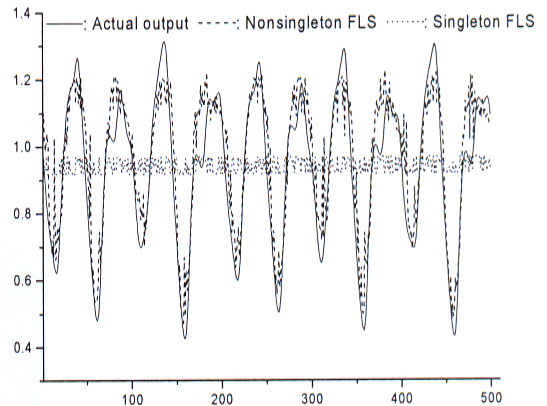
[Fig. 4] Noise-free Mackey-Glass time series data



[Fig. 5] Noise signal



[Fig. 6] Corrupted Mackey-Glass time series data



[Fig. 7] Actual output and nonsingleton, singleton FLS' output

4. 결론 및 향후과제

본 논문에서는 논싱글톤 퍼지시스템을 이용하여 훼손된 시계열데이터의 동정과 모델링을 수행하였으며, 이의 성능평가를 위해 일반적인 싱글톤 퍼지시스템과 결과를 비교하였다. 시뮬레이션에선 잘 알려진 성능평가 데이터인 Mackey -Glass 시계열 데이터를 이용하였으며, 불확실성이 내포된 잡음을 처리하는데 있어 매우 유용함을 보였다. 향후과제로 좀더 다양한 비선형 시스템에 논싱글톤 퍼지시스템을 적용하여 결과를 비교분석하고, 구조적인 최적화에 좀더 많은 연구가 필요하다.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant No. NRF-2017R1D1A1B03031467).

REFERENCES

[1] M.C. Mackey, and L. Glass, "Oscillation and chaos in physiological control systems," *Sci.*, vol. 197, pp.287-289, 1977.

[2] D. Kim, S.H. Huh, and G.T. Park, "Modeling Corrupted Time Series Data via Nonsingleton Fuzzy Logic System," *Lect. Notes Comput. Sc.*, vol. 3316, pp.1298-1303, 2004.

[3] G.C. Mouzouris and J.M. Mendel, "Nonsingleton Fuzzy Logic Systems: Theory and Application," *IEEE Trans. Fuzzy Syst.* vol. 5, no. 1, pp.56-71, 1997.

[4] M. Balazinski, E. Czogala, and T. Sadowski, "Control of metal-cutting process using neural fuzzy controller," in *2nd IEEE Int. Conf. Fuzzy Syst., San Francisco, CA*, Mar. vol. 1, pp.161-166, 1993.

[5] Mendel J, "Type-2 fuzzy sets and systems: An overview". *IEEE Comput. Intell. Mag.* 3(1): pp.20-29 2007.

[6] Mendel J, John R, Liu F. "Interval type-2 fuzzy logic systems made simple". *IEEE Trans. Fuzzy Syst.* 14(6): pp.808-821. 2006.

[7] Castillo O, Melin P. "A review on the design and optimization of interval type-2 fuzzy controllers". *Appl. Soft Comput.* 12(4): pp.1267-1278, 2012.

[8] Melin P, Mendoza O, Castillo O, "Face Recognition With an Improved Interval Type-2 Fuzzy Logic Sugeno Integral

and Modular Neural Networks". *IEEE Transactions on Systems, Man, and Cybernetics, Part A* 41(5): pp.1001-1012, 2011.

[9] Castillo O, Melin P, Pedrycz W. "Design of interval type-2 fuzzy models through optimal granularity allocation". *Appl. Soft Comput.* 11(8): pp.5590-5601, 2011.

[10] Melin P, Mendoza O, Castillo O. "An improved method for edge detection based on interval type-2 fuzzy logic". *Expert Syst. Appl.* 37(12): pp.8527-8535, 2010.

[11] Tan W W and Wu D. "Design of type-reduction strategies for type-2 fuzzy logic systems using genetic algorithms". *Stud. Comput. Intell.* 66: pp.169-187, 2007.

[12] Liu F. "An efficient centroid type-reduction strategy for general type-2 fuzzy logic system". *Inf. Sci.* 178(9): pp.2224-2236, 2008.

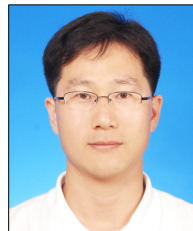
[13] Hameed I, "Simplified architecture of a type-2 fuzzy controller using four embedded type-1 fuzzy controllers and its application to a greenhouse climate control system", *Proc. Inst. Mech. Eng. Part I: J. Syst. Control Eng.* 223(5), pp.619-631. 2009.

[14] Hidalgo D, Melin P, Castillo O, "An optimization method for designing type-2 fuzzy inference systems based on the footprint of uncertainty using genetic algorithms", *Expert Syst. Appl.* 39(4), pp.4590-4598, 2012.

[15] Kim Y, An J, Kim J, "A fuzzy obstacle avoidance controller using a lookup-table sharing method and its applications for mobile robots", *Int. J. Adv. Robot. Syst.* 8(5), pp.39-48, 2011.

김 동 원(Dong Won Kim)

[정회원]



- 2007년 2월 : 고려대학교 전기공학과 (공학박사)
- 2008년 1월 ~ 2008년 12월 : UC Berkeley(Prof. Zadeh 그룹) EECS Post-Doc.
- 2009년 1월 ~ 2009년 8월 : UC Davis AME Post-Doc.
- 2009년 9월 ~ 현재 : 인하공업전문대학 디지털전자과 교수

<관심분야>

제어시스템, 로봇틱스, 지능연산, 퍼지시스템, 뉴럴네트워크, 딥러닝 시스템, 휴머노이드 로봇, 지능연산 융합 로봇시스템 등