

# 사회네트워크에서 잠재된 신뢰관계망 추론을 위한 ANFIS 모형

송희석

(한남대학교 경상대학 경영정보학과 부교수, hssong@hnu.kr)

## Abstract

We are sometimes interacting with people who we know nothing and facing with the difficult task of making decisions involving risk in social network. To reduce risk, the topic of building Web of trust is receiving considerable attention in social network. The easiest approach to build Web of trust will be to ask users to represent level of trust explicitly toward another users. However, there exists sparsity issue in Web of trust which is represented explicitly by users as well as it is difficult to urge users to express their level of trustworthiness. We propose a fuzzy-based inference model for Web of trust using user behavior information in social network. According to the experiment result which is applied in Epinions.com, the proposed model show improved connectivity in resulting Web of trust as well as reduced prediction error of trustworthiness compared to existing computational model.

**Key Words** : Trust, Sparsity, Web of trust, Fuzzy System, ANFIS

## 1. 서론

사회네트워크를 통한 정보와 경험의 공유가 급증하고 있는 가운데, 정보의 과다로 인한 정보 탐색의 지연, 거짓정보 또는 타인에 의해 의도된 정보를 이용하여 의사결정을 함으로써 발생하는 위험부담과 같은 다양한 역기능이 동시에 급증하고 있어 이에 대한 대안 마련이 절실히 요구되고 있다(Abdul-Rahman and Hailes, 2000). 사회네트워크에서 사용자간 안전한 정보와 경험의 교류를 위한 대안은 사용자간 신뢰도를 활용하

여 정보를 필터링하는 것으로, 이를 위해 사용자간 신뢰관계망을 구축하는 연구가 최근 주목을 받고 있다(Kim et al., 2008). 여기서 신뢰(Trust)라 함은 특정 사용자가 “다른 특정 사용자” 또는 “이전의 지식과 경험”과 같은 임의의 대상에 대해 가지는 주관적인 믿음의 정도를 의미하며 (Golbeck, 2005; Sabater and Sierra, 2005), 동일한 대상에 대해 개인별로 믿음의 정도가 다를 수 있다는 점에서 기존의 평판(Reputation)과는 다른 개념이다. 또한 사용자들간의 신뢰관계를 이용하여 네트워크를 구성할 수 있는데 이러한 네트워크를 신뢰망(Web of Trust) 또는 신뢰관계망이라고 부른다.

사회네트워크에서 사용자간 신뢰망을 구축하기 위한 가장 일반적인 방법은 사용자들에게 타인에 대한 신뢰정도를 명시적으로 표현하도록 요구하는 방법이다. 그러나 사용자로 하여금 다른 사용자에 대한 신뢰의 정도를 명시적으로 표현하도록 강제하는 것이 매우 어렵기 때문에 명시적인 표현을 이용해 구축된 신뢰관계망에는 희박성(Sparsity) 문제가 존재한다(Kim et al., 2008). 또한 사용자간 신뢰의 수준은 시간이 지남에 따라 달라질 수 있기 때문에 이를 지속적으로 업데이트하도록 하는 데도 한계가 존재한다. 이에 본 연구에서는 사용자 행위정보에 기반 하여 신뢰관계망을 추론하는 방법을 제안한다. 소셜네트워크에서 사용자 행위정보란 특정 사용자가 특정분야의 글을 작성하거나, 다른 사용자의 글에 답글을 달거나, 다른 사용자의 글을 평가하는 등의 모든 사용자 행위를 포함한다. 소셜네트워크는 서비스 특성 상 다양한 사용자의 작성 글과 답글 또는 평가 등의 상호작용 행위가 포함되기 때문에 이러한 정보를 바탕으로 신뢰관계망을 추론할 수 있다면 희박성(Sparsity) 문제의 해결이 일정수준 가능해 진다. 본 연구에서는 사용자의 작성글과 답글 또는 평가 등의 상호작용 행위정보를 모두 이용하여 신뢰관계망을 추론하는 모형을 제안한다.

사용자 행위정보에 기반 하여 신뢰관계망을 추론하기 위한 아이디어는 다음과 같다. 먼저 특정 사용자는 그의 관심분야와 유사한 분야에 있어서 전문성을 갖춘 사용자를 신뢰할 것이다. 이때, 특정분야의 작성글을 많이 작성했거나 작성글에 대한 높은 평가를 받은 사용자를 해당분야의 전문성이 높은 것으로 생각할 수 있다. 예를 들면, 사용자 A가 액션영화에 대해 관심이 많다면, 액션영화에 대해 작성글을 많이 작성했거나 작성글에 대해 높은 평가를 받은 사용자 B를 다른 사용자보다 신뢰할 것이다. 따라서 원천사용자가 타겟사용자를 얼마나 신뢰하는지를 추론하기 위한 아이디어는 타겟사용자의 카테고리별 전문성 뿐 아니라 원천사용자의 카테고리별 관

심도를 종합하여 판단하는 것이며, 이러한 아이디어에 의해 신뢰망을 추론하게 되면 신뢰망의 희박성문제도 해결될 수 있다.

한편, 신뢰도란 개념은 수치로 정확히 표현하기 어려울 뿐 아니라 불확실성이 내재된 개념이다(Zadeh, 1965; Chang et al., 2005). 이러한 속성은 높다, 낮다 등 인간이 사용하는 퍼지용어로 기술되는 것이 바람직하기 때문에 퍼지기법을 통한 표현과 추론이 적합한 분야이다. 퍼지기법은 다른 기법에 비해 다양한 사용자의 의견이나 다른 형태의 측정값들을 종합하여 평가를 내릴 때 더욱 효과적인 기법으로 알려져 있다(Akhter et al., 2005; Song et al., 2005). 이에 본 연구에서는 퍼지신경망을 적용하여 신뢰관계망을 추론하고 기존의 신뢰추론 기법(Kim et al., 2008)과 비교하여 신뢰추론의 정확성을 평가하고자 한다. 퍼지신경망 모형은 Jang(1993), Nauck and Kruse(1999), Palit and Popovic (2000), Wang (1994) 등에 의해 다양한 형태로 제안되어 왔다. 이 중 탁월한 함수근사능력을 갖추어 최근 세계 열 예측 및 시스템컨트롤 분야에 폭넓게 적용되고 있는 모형이 Jang (1993)에 의해 제안된 ANFIS (Adaptive Network-based Fuzzy Inference System)모형이다. 본 연구에서는 ANFIS모형을 기반으로 사용자 행위정보를 활용한 신뢰관계망 예측 모형을 개발하고, 이를 소셜네트워크 중 인지도가 높은 Epinions.com 사이트에 적용하여 그 성능을 기존의 비퍼지화 신뢰관계망 추론 기법과 비교하고자 한다. 퍼지신경망 모형을 적용하여 신뢰관계망을 추론하면 예측의 정확도도 높아질 뿐 아니라, 추론과정에서 퍼지 if-then 규칙을 생성하기 때문에 비 퍼지화 기법에서는 제시될 수 없는 대상도메인에 대한 지식을 동시에 제시해 준다는 장점을 가진다 (Babuka and Verbruggen, 2003).

## II. 선행연구

신뢰(Trust)는 그 특성 상 불확실하고, 주관적이며, 시간에 따라 변화될 수 있는 개념이기 때문에 수치로 표현하기 보다는 퍼지용어로 표현하는 것이 적합하다는 점에 착안하여 최근 퍼지를 기반으로 신뢰를 추론하고자 하는 연구들이 등장하고 있다. 신뢰추론에 퍼지를 적용하고자 한 기존연구들은 대부분 P2P시스템에 대한 연구들이다. Song et al. (2005)은 경매P2P 시스템에서 구매자와 판매자의 로컬신뢰도를 계산하고, 동료들의 로컬신뢰도와 거래량 및 거래일자들 기반으로 다시 글로벌평판을 추론하는 2단계의 퍼지추론시스템인 FuzzyTrust 시스템을 제안하였다. 구매자의 로컬신뢰도를 계산하기 위해서는 지불방법, 지불시기정보를 설명변수로 이용하였고, 판매자의 로컬신뢰도를 계산하기 위해서는 제품품질, 납기시간을 설명변수로 이용하는 퍼지추론시스템을 제안하였다. 그들은 퍼지추론시스템이 불확실한 개념을 표현하고 처리하는데 효과적이었을 뿐 아니라, 악의적인 판매자를 식별하는데 우수한 모형임을 보여주었다. 또한 퍼지시스템은 동료들로부터 수집된 불확실한 정보들을 종합하는데 효과적인 기법임을 주장하였다.

Aringhieri et al. (2006)은 P2P시스템에서 각 노드의 평판을 계산하기 위해 퍼지시스템을 활용하였다. 이들은 특정노드에 관한 다양한 평판을 수집하기 위해 P2P노드들에게 신호를 전송하고, 수집된 다양한 의견들을 퍼지기법을 적용하여 종합하는 두 단계의 방법을 통해 평판도를 계산하였다. Griffiths et al. (2006)는 P2P시스템에서 에이전트들이 협업을 하는 과정에서 불확실성으로 인해 실패의 위험이 존재하며 이러한 위험을 최소화하기 위해 신뢰추론이 필요함을 주장하였고, 퍼지로직을 이용한 신뢰추론 모형을 제시한바 있다. Lin et al. (2009)는 동료들 간의

신뢰관계는 불확실성이 내재되어 있기 때문에 퍼지로직의 적용이 적합한 분야라고 설명하고, 계층적 퍼지추론시스템을 적용하여 신뢰를 추론하였다. 그들은 기존의 퍼지추론시스템은 속성의 수(설명변수의 수)와 속성 값의 수준이 많아지면 기하급수적으로 규칙의 수가 증가하게 되고 이로 인해 계산량의 증가와 규칙에 대한 이해가 어려워짐을 주장하였고, 규칙의 수를 줄이기 위해 계층적 퍼지추론시스템을 사용하여 신뢰관계를 추론하는 모형을 제시하였다. 실험결과 그들이 제시한 계층적 퍼지추론방식이 기존의 퍼지추론방식에 비해 퍼지규칙의 수를 현저히 줄이는 기법임을 보여주었다. Lesani and Montazeri (2009)는 신뢰를 계산할 때 수치모형과 같은 비퍼지화 모형을 적용하면 상충(Contradict)되는 정보를 결합하는데 있어서 부적합함을 보여주었고, 퍼지신뢰도를 결합하는 과정에서 퍼지추론기법을 사용하였고 퍼지기법이 비퍼지화기법보다는 추론성능이 탁월함을 보여주었다.

이와 같이 퍼지를 이용한 신뢰예측 방법이 여러 학자들에 의해 시도되고 있지만, 대부분의 연구는 P2P시스템에 적용된 연구이며, 이를 사회네트워크에 적용한 연구는 찾아보기 어렵다. 또한 사용자간 상호작용행위에 대한 모니터링이 어려운 P2P시스템과는 달리 사회네트워크에서는 사용자간 다양한 상호작용행위 정보가 저장되고 웹페이지에 노출되어 있음에도 불구하고 이러한 정보를 기반으로 퍼지추론시스템을 설계하고자 한 시도는 아직 찾아볼 수 없다. 사용자 행위정보를 기반으로 신뢰관계망을 추론할 경우, 신뢰관계망의 희박성문제도 해결할 수 있다는 장점을 가진다. 이외에도 기존연구들은 퍼지규칙의 최적 구조설계와 모수조정을 위한 학습과정을 거치지 않은, 단순 퍼지추론시스템 수준의 연구이다. 따라서 퍼지신경망 모형화를 통해 최적 네트워크 구조설계와 모수조정이 이루어진다면 예측 정확도 향상의 기회가 한층 증가할 것으로

보인다.

### III. 신뢰추론을 위한 구성개념 정의

사용자 행위정보에 기반 하여 신뢰관계망을 추론하기 위해 우선 실험대상 사이트인 Epinions.com 사이트에서 발생하는 사용자 행위 관계를 규명하고, 이를 바탕으로 신뢰망 추론에 필요한 구성개념을 정의한다. Epinions.com 사이트는 자동차, 책, 영화, 음악, 컴퓨터, 전자제품 등 여러 분야로 구분되어 있으며, 각 분야마다 다시 세부적인 서브카테고리로 나누어져 해당 카테고리내의 제품 또는 콘텐츠에 대해 리뷰를 작성하도록 되어있다. 임의의 사용자는 작성된 리뷰를 평가할 수 있다. 원천사용자가 타겟사용자를 얼마나 신뢰하는지를 추론하기 위한 아이디어는 원천사용자의 카테고리별 관심분야 벡터를 구성하고, 타겟사용자의 카테고리별 전문성 벡터를 추출하여 두 벡터로부터 신뢰도에 대한 추정값을 일차적으로 유도한 후 이들을 퍼지신경망 기반 신뢰추론모형의 입력변수로 활용하는 것이다. 사용자의 카테고리별 관심도와 카테고리별 전문성은 다음과 같이 구해진다. 사용자별 관심도와 전문성을 측정하기 위한 식은 Rigg (2001) 에 의해 처음 제안되고, Kim et al. (2008)에 의해서도 채택된 모델로 활용하기로 한다. 먼저 사용자의 카테고리별 관심도는 각 카테고리별로 얼마나 많은 리뷰를 작성하였고 얼마나 많은 평가를 수행하였는지를 종합하여 측정할 수 있다. 일반적으로 리뷰보다 평가가 더 많기 때문에 각각을 별도로 사용자 관심도를 측정하는데 반영하기로 한다. 아래식은 사용자의 카테고리별 관심도를 나타낸다. U가 사용자의 수이고, C

가 카테고리의 수라고 할 때, 사용자의 카테고리별 관심도 행렬 A는 행의 크기가 U이고 열의 크기가 C인 행렬이 된다.

$$A_{ij} = \frac{1}{2} \times \left( \frac{a_{ij}^r}{\text{MAX}_{j \in \text{all category}}(a_{ij}^r)} + \frac{a_{ij}^w}{\text{MAX}_{j \in \text{all category}}(a_{ij}^w)} \right)$$

$A_{ij}$ 는 사용자  $i$ 의 카테고리  $j$ 에 대한 관심도 ( $0 \leq A_{ij} \leq 1$ )  
 $a_{ij}^r$ 는 사용자가 카테고리  $j$ 에 관해 평가한 리뷰의 총수  
 $a_{ij}^w$ 는 사용자가 카테고리  $j$ 에 관해 작성한 리뷰의 총수

사용자의 카테고리별 전문성은 카테고리별 리뷰에 대한 평판을 토대로 측정할 수 있다. 즉 특정카테고리에서 작성한 리뷰들이 좋은 평가를 많이 받게 되면 해당사용자의 해당분야 전문성이 높은 것으로 간주할 수 있다. 아래식은 사용자의 카테고리별 전문성을 측정하기 위한 식이다.

$$E_{ij} = \left( 1 - \frac{1}{n_{u_i^w} + 1} \right) \times \frac{\sum_{j \in R(u_i^w)} \bar{r}_j}{n_{u_i^w}}$$

$E_{ij}$  사용자  $i$ 의 카테고리  $j$ 에 관한 전문성  
 $n_{u_i^w}$  리뷰작성자  $u_i^w$ 가  $j$ 카테고리에서 작성한 리뷰의 총수  
 $R(u_i^w)$  리뷰작성자  $u_i^w$ 가  $j$ 카테고리에서 작성한 리뷰집합  
 $\bar{r}_j$  리뷰  $j$ 의 품질

식에서 사용자의 카테고리별 전문성은 특정사용자가 해당카테고리 내에서 작성한 리뷰의 품질을 평균하여 측정한다. 리뷰품질을 산술평균함으로써 특정 리뷰작성자의 전문성을 측정하게 되면, 리뷰를 단 한 번만 작성한 작성자나 여러 번에 걸쳐 리뷰를 작성한 사용자가 동일하게 평가될 수 있다. 일반적으로 많은 수의 리뷰를 작성하고, 리뷰의 평균품질도 높은 사용자는 리뷰의 평균품질이 높긴 하지만, 적은 수의 리뷰를 작성한 사용자보다 해당분야의 전문성을 보다 높게 평가하는 것이 바람직하다. 따라서 식의 첫 번째 항목인 는 리뷰 작성횟수가 적은 사용자의 전문성을 할인하기 위해 추

가되었다. 한편 식에 나타난 리뷰의 품질은 해당리뷰에 대한 평가치를 평가자의 평판도에 따라 가중 평균함으로써 구해질 수 있다. 아래식은 리뷰의 품질을 측정하는 식이다. 즉 보다 신뢰할만한 평가자가 평가한 평가치에 보다 가중치를 높이 주어 리뷰의 품질을 측정한다.

$$\bar{r}_j = \frac{\sum_{i \in U^r(r_j)} \bar{u}_i^r \rho_{ij}}{\sum_{i \in U^r(r_j)} \bar{u}_i^r}$$

$\bar{r}_j$ 는 리뷰  $j$ 의 품질 ( $0 \leq \bar{r}_j \leq 1$ )  
 $U^r(r_j)$ 는 리뷰  $j$ 를 평가한 평가자의 집합  
 $\bar{u}_i^r$ 는 리뷰평가자  $u_i^r$ 의 평판  
 $\rho_{ij}$ 는 평가자  $u_i^r$ 이 리뷰  $j$ 부여한 평가치

마지막으로 리뷰평가자의 평판은 아래식과 같이 측정할 수 있다. 동일한 리뷰에 대해 대부분의 평가자는 유사한 평가치를 부여했음에도 불구하고 특정 평가자만 편차가 큰 평가치를 부여한 경우, 이 평가자는 신뢰하기가 힘들기 때문에 평판도를 낮게 부여하는 것이 바람직하다. 따라서 리뷰평가자의 평판은 리뷰의 평균품질로부터 동떨어진 정도로 평가할 수 있다. 즉 리뷰에 대한 특정 평가자의 평가치가 리뷰  $j$ 의 품질과 유사 할수록 해당 평가자의 평판은 높게 나타나야한다. 식의 첫 번째 항목은 평가 횟수가 적은 사용자의 평판을 할인하기 위해 추가되었다.

$$\bar{u}_i^r = \left(1 - \frac{1}{n_{u_i^r} + 1}\right) \times \frac{\sum_{j \in R(u_i^r)} 1 - |\bar{r}_j - \rho_{ij}|}{n_{u_i^r}}$$

$\bar{u}_i^r$  리뷰평가자  $u_i^r$ 의 평판 ( $0 \leq \bar{u}_i^r \leq 1$ )  
 $n_{u_i^r}$  리뷰평가자  $u_i^r$ 이  $j$ 카테고리에서 평가한 리뷰의 총수  
 $R(u_i^r)$  리뷰평가자  $u_i^r$ 이  $j$ 카테고리에서 평가한 리뷰 집합

원천사용자가 타겟사용자를 얼마나 신뢰하는지를 추론하기 위한 방법은 원천사용자의 카테고리별 관심분야 벡터 와, 타겟사용자의 카테고리별 전문성 벡터를 종합하여 일차 신뢰도 추정치를 계산한 후 이를 입력 변수로 활용하여 ANFIS 신뢰추론모형을 구성하는 것이다. 즉 특정 타겟사용자( $j$ )의 전문성이 높은 카테고리 분야에 대해, 원천사용자( $i$ )의 해당 카테고리 관심도 또한 높게 나타날 때 원천사용자의 타겟사용자에 대한 신뢰도가 높게 평가되어야 할 것이다. 따라서 두 벡터를 다음과 같은 세 가지 방법으로 종합함으로써 일차적인 신뢰추정식을 구성하기로 한다. 첫 번째 식은 타겟사용자  $j$ 의 카테고리별 전문성을 원천사용자  $i$ 의 관심도 값으로 가중 평균함으로써 신뢰추정식을 구성한 것이다. 이 신뢰추정식은 타겟사용자의 카테고리별 전문성이 낮더라도 원천사용자의 카테고리별 관심도가 높으면 신뢰 추정값이 높게 나타나는 문제가 있을 수 있다. 두 번째 식은 원천사용자  $i$ 의 카테고리별 관심도와 타겟사용자  $j$ 의 카테고리별 전문성간의 유사도를 유클리디안 거리의 관점에서 측정하고 이를 기반으로 신뢰추정식을 구성한 것이다. 즉 타겟사용자의 전문성이 높을 뿐 아니라 전문성과 관심도 값 사이의 거리유사도가 높다면 신뢰추정치도 높을 것이라는 판단에서 유도된 추정식이다. 마지막 식(최대신뢰 카테고리 기반 신뢰추정치)은 원천사용자가 높은 관심을 보이는 어떠한 카테고리에 있어서 타겟사용자의 전문성이 높다면 다른 카테고리의 관심도 및 전문성과 무관하게 타겟사용자를 신뢰할 수 있다는 점에서 추가로 구성된 추정식이다.

전문성을 관심도로 가중평균한 신뢰추정치 :

$$S_{ij}^A = \frac{\sum_k a_{ik} e_{jk}}{\sum_k a_{ik}}$$

거리유사도기반신뢰추정치 :

$$S_{ij}^B = \sum_k \frac{\max(|a_{ik} - e_{jk}|, |a_{ik} - e_{jk}|)}{\max(|a_{ik} - e_{jk}|)} e_{jk}$$

최대신뢰 카테고리 기반 신뢰추정치 :

$$S_{ij}^C = \max_{k \in \text{모든카테고리}} \frac{\max(|a_{ik} - e_{jk}|, |a_{ik} - e_{jk}|)}{\max(|a_{ik} - e_{jk}|)} e_{jk}$$

$A_i$ 는 원천사용자  $i$ 의 카테고리별 관심도 벡터,

$$A_i = [a_{i1} a_{i2} \dots a_{ik}]$$

$E_j$ 는 타겟사용자  $j$ 의 카테고리별 전문성 벡터,

$$E_j = [e_{j1} e_{j2} \dots e_{jk}]$$

#### IV. ANFIS기반 신뢰망 추론모형

일반적으로 퍼지신경망을 구현하기 위해서는 네트워크 구조설계가 필요하다(Bersini and Bontempi, 1997). 네트워크 구조설계라 함은 추론에 필요한 퍼지규칙의 구조와 형태를 결정하고 퍼지추론 방법을 선택하는 것을 포함한다. 퍼지규칙은 퍼지용어로 기술되기 때문에 각 퍼지용어에 대한 소속함수(Membership Function)의 형태뿐 아니라 입력변수들로 구성된 다차원의 입력공간을 몇 개의 세부공간으로 분할할 것인지도 결정하여야 한다. 퍼지공간에 대한 분할이 이루어지면 각 퍼지공간에 대응하는 퍼지 규칙이 필요하게 되며, 이렇게 정해진 퍼지 규칙의 개수와 규칙의 형태는 추론성과 규칙의 해석용이성을 결정하는 중요한 요인이 된다. 초기 퍼지규칙이 구성되면 이를 이용하는 퍼지추론 방법이 결정되어야 한다. 이는 AND, OR 등의 연산과 같은 퍼지연산방법과 디퍼지화 방법을 포함한다. 추론모형의 네트워크 구조설계가 완료되면 각 네트워크 노드에서 사용되는 파라미터를 추론오차가 최소화되도록 조정하는 학습과정을 거친다. 각 네트워크 노드에서 사용되는 파라미터는 퍼지규칙의 조건부

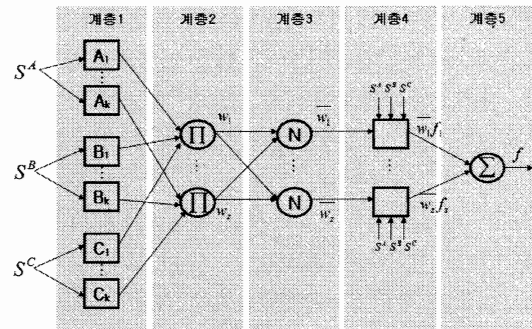
에 기술된 각 퍼지용어의 소속함수를 결정하기 위한 파라미터와 규칙의 결과부에 나타나는 선형함수의 파라미터를 의미한다. 일반적으로 ANFIS에서는 경사하강법(Gradient Descent)과 혼합학습법(Hybrid Learning)을 학습알고리즘으로 사용하고 있다. 퍼지추론 과정은 입력변수로 구성된 다차원의 입력공간을 세부영역으로 구분하여 각 세부영역에서 추론의 결과 값을 산출하도록 하는 “분할 후 정복(Divide and Conquer)” 방식으로 이루어진다. 즉, 퍼지규칙의 조건부는 입력공간을 세부영역으로 분할하게 되며, 퍼지규칙의 결론부는 각 영역에서의 추론결과를 기술하게 된다. 따라서 퍼지규칙의 구조를 설계한다는 것은 입력공간을 어떻게 분할하는가와 밀접히 관계되어 있다. ANFIS에서 입력변수들로 구성된 다차원공간을 분할하는 방법으로 그리드분할(Grid Partitioning)과 차감클러스터링(Subtractive Clustering)방법을 주로 사용한다(송희석외, 2009).

그리드 분할(Grid partitioning)은 입력공간을 중복이 생기지 않도록 격자구조로 공간을 분할하는 것이다. 일반적으로 그리드 분할법을 적용하면 분할된 퍼지영역이 동질적으로 존재하기 때문에 퍼지규칙의 해석용이성(Interpretability)이 높아진다(Bersini and Bontempi, 1997). 이 방법은 입력변수의 수가 작을 때(입력공간의 차원이 낮을 때) 주로 사용되는 방법이다. 예를 들면, 입력변수가 10개이고, 각 입력변수가 2개의 소속함수로 구분이 되는 경우, 개의 세부영역이 생기게 되며, 각 영역마다 1개의 규칙씩 총 1024개의 퍼지규칙이 만들어지게 되어 매우 복잡한 시스템이 된다. 이와 같이 입력변수가 많거나 소속함수의 수가 많은 경우, 차감 클러스터링(Subtractive Clustering)기법을 사용하는 것이 효과적이다. 이 방법은 다차원의 입력데이터를 분석

하여 이를 군집화 함으로써 입력공간을 몇 개의 세분된 퍼지공간으로 구분하는 방법이다. 차감 클러스터링은 퍼지 군집의 개수를 지정하지 않더라도 입력데이터의 공간을 적절한 갯수로 분할한다(Chiu, 1994). 차감 클러스터링에서는 군집의 반경을 지정 하게 되는데 이는 데이터 공간을 단위 하이퍼큐브로 간주할 때 군집 중심으로 부터의 영향 범위를 나타내며 0에서 1사이의 값을 가진다. 너무 작은 군집반경을 지정하면 작은 군집이 많이 생기게 되고 따라서 퍼지규칙의 수가 많아진다. 반대로 군집반경을 크게 설정하면 군집의 수가 작아져서 규칙의 수도 작아지게 된다. 입력공간의 분할이 이루어지고 나면 소속함수의 형태를 정의해야 한다. 소속함수의 형태로 주로 사용되는 것은 삼각형(Triangular)함수, 사다리꼴(Trapezoidal)함수, 가우시안(Gaussian)함수, 종형(bell)함수, 시그모이달(sigmoidal)함수 등이 있다. 이중 가우시안함수와 종형함수는 부드러운 곡선 형태를 취하며 작은 수의 파라미터와 간략한 표현법에 의해 퍼지집합을 묘사하기 때문에 가장 폭넓게 사용되고 있다. 한편 시그모이달함수는 오른쪽과 왼쪽이 개방된 형태의 함수로써 “매우크다”, “매우높다” 등의 극단적인 퍼지집합을 표현하는데 자주 사용된다 (Jin et al., 1999; Jin, 2000; Malhotra et al., 2002).

그림1은 신뢰망추론을 위해 구성된 ANFIS 네트워크구조로써 세 개의 유사도 기반 신뢰추정치를 입력변수로 취하고, Takagi-Sugeno규칙을 이용하고 있다. ANFIS 기반 신뢰망 추론모형은 다섯계층의 전방향 연결 신경망 구조로 되어 있다. ANFIS에서 각 계층별 노드는 서로 다른 노드함수를 가지며, 이 노드함수는 학습과정을 통해 조절이 되는 파라미터를 가진다. 그러나 노드를 연결하는 링크는 노드 간 흐름의 방향을 나타낼 뿐 어떠한 가중치나 파

라미터도 가지지 않는다. 그림에서 사각형으로 표시된 노드는 파라미터를 포함하는 적응형 노드를 의미하며, 원으로 표시된 노드는 파라미터 없이 주어진 연산만을 수행하는 일반 노드를 나타낸다.



<그림 1> 신뢰망 추론모형 네트워크 구조

ANFIS의 학습과정은 주어진 입력과 출력데이터를 사용하여 출력오차를 최소화하도록 규칙의 조건부 및 결론부 파라미터를 조정해 나가는 것이다. 학습 알고리즘은 역전파알고리즘(Back Propagation Algorithm)과 혼합알고리즘(Hybrid Algorithm)이 있는데 본 연구에서는 혼합알고리즘을 사용한다. 혼합알고리즘은 전방향 학습과 역방향 학습으로 구성되는데, 전방향 학습에서는 조건부 파라미터를 고정시키고 결론부 파라미터를 최적화하기 위해 최소자승법(Least-squares Method)을 적용한다. 최적의 결론부 파라미터가 발견되면 역방향 학습이 즉시 수행된다. 역방향 학습에서는 소속함수를 정의하는 규칙의 조건부에 있는 파라미터를 조정하기 위해 경사하강(Gradient Descent) 기법을 채택하고 있다.

## V. 실험

본 연구에서는 사회네트워크로 널리 알려져 있는 Epinions (www.epinions.com) 사이트를 대상으로 제안한 신뢰망 추론모형의 정확성을 기존모형과 대비하여 평가하였다.

Epinions 사이트는 다양한 카테고리분야의 제품 및 콘텐츠에 대해 리뷰를 작성하게 되어 있고, 각 리뷰에 대해 사용자들의 평가가 이루어진다. 또한 이 사이트는 명시적인 신뢰관계를 표현할 수 있기 때문에 다양한 사용자 행위 데이터뿐 아니라 명시적인 신뢰관계 정보 데이터에 기반하여 신뢰관계망의 정확성을 평가할 수 있기 때문에 본 연구의 목적에 매우 부합되는 사이트라 할 수 있다. 이 사이트의 카테고리 중 비디오와 DVD 카테고리를 선택하여 이 카테고리 내의 12개 서브 카테고리의 사용자를 대상으로 신뢰관계망을 추론하기로 한다. 데이터 수집 방법은 자동수집로봇 프로그램을 이용하여 먼저 이 카테고리분야에 리뷰와 평가 내역들을 모두 수집하고, 각 리뷰와 평가의 주체가 되는 사용자 정보를 수집하게 된다. 또한 이들 사용자간의 명시적인 신뢰관계망 정보도 수집하게 된다. 자료수집결과 최종적으로 44,196명의 사용자를 수집할 수 있었고, 이들 간 상호작용이 최소 한 건 이상 존재하는 링크의 개수는 총 644,992개가 존재하였다. 또한 사용자간에 명시적으로 표현된 신뢰관계(링크)의 수는 총 74,198건 이었다.

### 5.1 예측성능평가

신뢰추론의 정확성평가는 본 연구모형에서 계산된 사용자들 간의 신뢰추론 결과를 명시적으로 표현된 신뢰값과 비교함으로써 이루어진다. 이를 위해 신뢰관계가 명시적으로 표현된 링크 74,198건 중 50,000건을, 사용자간 상호작용이 한 건 이상 존재하지만 신뢰관계는 알 수 없는 링크 644,992건 중 50,000건을 각각 임의 추출하여 총 100,000개의 링크에 대한 성능평가용 데이터 집합을 구성하였다. 이 데이터를 50% 씩 구분하여 훈련용 데이터 집합과 검증용 데이터 집합으로 구분하였다. 실험은 매트

랩 7.6.0 버전의 퍼지 툴박스를 사용하여 진행되었다. 퍼지규칙의 출력함수는 1차 선형 함수를 사용하였으며, 입력변수의 소속함수는 가우시안 함수를 활용하였다. 모수추정을 위한 학습방법으로 혼합학습(Hybrid learning)법을 적용하였으며, 충분히 많은 데이터를 학습에 이용하였기 때문에 반복횟수는 3으로 지정하였다. 실제 모든 모형에서 3회 반복 이전부터 최소 오차가 시작되어 이후 이 오차가 유지되는 것을 확인할 수 있었다. 모형의 평가 기준으로 RMSE(Root Mean Square Error)를 사용하였다. 실험을 위한 모형은 계산모형에 의한 신뢰추정치인  $S^A$ ,  $S^B$ ,  $S^C$  각 각에 대해 RMSE를 평가하고, 이들을 입력변수로 하는 ANFIS모형을 수행하여 각 모형의 RMSE를 평가하였다. ANFIS모형은  $S^A$ ,  $S^B$ ,  $S^C$  각 각을 입력변수로 사용하는 모형과  $S^A$ ,  $S^B$ ,  $S^C$  를 동시에 입력변수로 사용하는 모형으로 구분하여 실험을 진행하였다.

표1은 실제 신뢰관계가 명시적으로 표현되어 있고, 직접적인 상호작용이 있는 링크만을 대상으로 신뢰 추론정확도를 계산한 결과이다. 이 실험에서는 추가로 인공신경망에 의한 추론오차도 동시에 평가하였다. 사용된 인공신경망은 은닉층이 20개 노드로 구성된 구조이고, Levenberg-Marquardt 알고리즘을 이용하여 검증용 데이터의 오차제곱합이 최소가 되는 지점까지 반복하여 학습을 수행하였다. 실험결과 노이즈의 영향이 적은 신뢰관계링크에서는 수리모형에 비해 학습형 모형의 RMSE가 현저하게 낮게 나타났다. 이 중 훈련데이터의 RMSE는  $S^A$ ,  $S^B$ ,  $S^C$  를 동시에 입력변수로 사용하는 ANFIS모형에서 가장 낮았고 검증데이터의 RMSE는 인공신경망 모형에서 가장 낮게 나타났다. 대체로 인공신경망과 ANFIS모형의 추론성능은 유사하게 나타났으나, ANFIS모형의 경우 퍼지규칙을 생성하여



좁으로써 도메인에 대한 이해가 용이하다는 점에서 장점을 가진다.

구분	입력	분할방법	RMSE	
			훈련데이터	검증데이터
Kim et al. (2008)	S <sup>A</sup>	-	0.3915	0.3833
추가된 신뢰추정치	S <sup>B</sup>	-	0.3978	0.4295
	S <sup>C</sup>	-	0.5188	0.5339
ANN모형에 의한 신뢰추정결과	S <sup>A</sup> S <sup>B</sup> S <sup>C</sup>	-	0.0298	0.0447
ANFIS모형에 의한 신뢰추정결과	S <sup>A</sup>	GP(3)	0.0297	0.0450
		SC(0.5)	0.0298	0.0451
	S <sup>B</sup>	GP(3)	0.0298	0.0450
		SC(0.5)	0.0298	0.0450
	S <sup>C</sup>	GP(3)	0.0299	0.0452
		SC(0.5)	0.0299	0.0452
	S <sup>A</sup> S <sup>B</sup> S <sup>C</sup>	SC(0.5)	0.0297	0.0450
		GP(4 4 4)	0.0294	0.0450

<표 1> 모형별 RMSE 평가결과

## 5.2 데이터 희박성 이슈

Epinion사이트의 DVD 및 비디오 카테고리의 전체 사용자 44,196명에 대해 상호간에 가능한 총 관계(링크)의 수는 1,953,242,220 개가 존재한다. 이 중 신뢰관계가 명시적으로 표현된 링크의 수는 74,198개이며, 나머지 링크에 대해서는 신뢰관계를 파악할 수 없다. 그러나 본 연구에서 제시한 방법에 따르면 리뷰작성 또는 리뷰에 대한 평가가 한 건 이상 존재하는 원천사용자와 리뷰글을 최소 한건 이상 작성하여 전문성에 대한 평가가 가능한 타겟 사용자 간의 모든 가능한 링크에 대해 신뢰관계 추론이 가능해진다. 현재 수집된 전체 사용자 44,196명은 최소 한 건 이상 리뷰작성 또는 리뷰에 대한 평가를 수행한 사용자이고, 카테고리별 전문성에 대한 평가가 가능한 사용자의 수는 24,723명이므로 신뢰관

계 추론이 가능한 링크의 수는 총 1,092,657,708개가 된다. 신뢰관계망의 연결성 수준을 전체 링크수 대비 신뢰관계에 대한 파악이 가능한 링크 수로 정의한다면, 명시적인 신뢰표현에 의한 신뢰관계의 연결성 수준은 전체의 0.0038% 임에 비해, 본 연구에서 제시한 추론방법을 적용할 경우 신뢰관계망의 연결성 수준은 55.940%로 급증하게 된다.

## 5.3 퍼지규칙을 이용한 도메인 해석

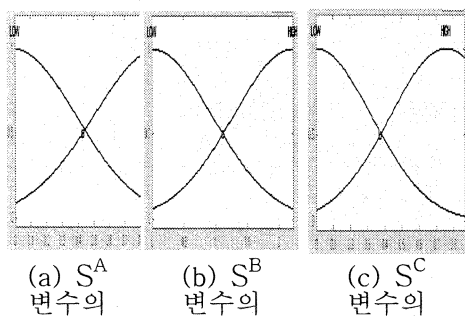
여기서는 S<sup>A</sup>, S<sup>B</sup>, S<sup>C</sup>를 동시에 입력변수로 사용하고, 그리드분할을 이용하여 입력공간을 분할하는 ANFIS모형을 사용하여 생성된 퍼지규칙을 분석해 보고, 이를 통해 도메인에 대한 이해를 높이고자 한다. 일반적으로 퍼지규칙의 해석용이성은 모형의 정확성과 상충관계에 있다. 이에 정확성을 다소 희생하더라도 퍼지규칙의 해석이 용이한 모형을 선정하여 퍼지규칙을 생성하기로 한다. 그림2는 각 입력변수를 두 개의 퍼지집합으로 구분하는 그리드 분할 모형에 의해 생성된 퍼지규칙이다. 이모형을 훈련데이터에 적용한 RMSE는 0.4460 이었고, 검증데이터에 적용한 RMSE는 0.4876이었다. ANFIS모형에서 생성된 규칙은 퍼지규칙이므로 연속형 값을 가지며, 특정 입력이 주어지면 이 입력과 매치되는 규칙들의 적합도를 종합하여 최종결론을 내리게 된다. ANFIS의 퍼지규칙은 로컬매핑(개별 전문가에 의해 기술된) 규칙이기 때문에 개별규칙만으로 결과를 직관적으로 이해하기는 어렵다. 조건부가 유사한 여러 규칙들의 결론부를 가중 평균하는 디퍼지화를 통해 최종결과가 도출됨을 이해해야 한다. 퍼지 규칙의 이러한 로컬매핑 특성은 현재 훈련 중인 학습데이터의 출력 오차를 최소화 할 뿐 아니라 과거에 학습된 데이터의 출력오차도 최소로 유지하도록 함으로써 정확성 향상에

기여하게 된다. 그림2에서 규칙4는  $S^A$  값이 낮고,  $S^B$ ,  $S^C$  가 높은 값을 가질 때 최종 신뢰추정 값은 상대적으로 높아지는 경우가 많음을 의미한다. 또한 규칙2,3으로부터  $S^A$  값이 낮고,  $S^B$ ,  $S^C$  중 하나가 낮은 값을 가질 때 최종 신뢰추정 값도 상대적으로 낮아진다는 것을 알 수 있다. 규칙1과 같이  $S^A$ ,  $S^B$ ,  $S^C$  값이 동시에 낮음에도 불구하고 신뢰추정값이 중간 수준(0.3352)으로 나타난 것은 노이즈의 영향으로 보이며, 이는 신뢰관계를 알 수 없는 링크 중에서도 신뢰관계가 존재하는 링크가 많이 포함되어 있음을 의미한다.

1. IF $S^A$ =LOW AND $S^B$ =LOW AND $S^C$ =LOW THEN 0.3352
2. IF $S^A$ =LOW AND $S^B$ =LOW AND $S^C$ =HIGH THEN -0.7278
3. IF $S^A$ =LOW AND $S^B$ =HIGH AND $S^C$ =LOW THEN -0.6349
4. IF $S^A$ =LOW AND $S^B$ =HIGH AND $S^C$ =HIGH THEN 2.8050
5. IF $S^A$ =HIGH AND $S^B$ =LOW AND $S^C$ =LOW THEN 0.7770
6. IF $S^A$ =HIGH AND $S^B$ =LOW AND $S^C$ =HIGH THEN 0.9336
7. IF $S^A$ =HIGH AND $S^B$ =HIGH AND $S^C$ =LOW THEN 0.8428
8. IF $S^A$ =HIGH AND $S^B$ =HIGH AND $S^C$ =HIGH THEN 0.7622

<그림2> ANFIS GP(2 2 2)모형의 퍼지 규칙

그림 3은 각 입력변수별 소속함수의 형태와 위치를 나타내고 있다. 그리드분할 모형의 경우, 차감클러스터링 모형에 비해 각 퍼지집합의 소속함수가 비교적 뚜렷이 구분되도록 조정되는 경향이 있어 퍼지집합에 퍼지용어의 할당이 용이하다. 그림에서도 소속함수의 위치가 상대적으로 명확히 구분되고 있음을 알 수 있다.



소속함수    소속함수    소속함수  
<그림3> 입력변수별 소속함수

#### IV. 결론

본 연구에서는 사회네트워크에서 사용자 행위 정보를 바탕으로 신뢰관계망을 추론하기 위한 퍼지신경망 모형인 ANFIS모형을 제안하였다. 사용자에게 의해 명시적으로 표현된 신뢰관계망에는 희박성 문제가 존재하게 되며, 본 연구에서 제안한 모형으로 신뢰관계망을 추론할 경우 연결성 수준이 제고되어 희박성 문제가 상당 수 개선되는 장점을 가진다. 또한 예측 정확성에 있어서도 기존의 수리모형대비 높은 정확도를 나타내고 있음을 확인할 수 있었다. 퍼지신경망은 학습의 결과로 최적의 모수를 결정할 뿐만 아니라 일반 신경망 모형에서와 달리 퍼지규칙을 생성한다. 퍼지규칙을 활용하면 도메인에 대한 이해가 가능할 뿐 아니라 예측 결과를 해석하는데도 사용될 수 있다. ANFIS모형은 인간 전문가와 상호작용하면서 규칙을 정제해 나갈 수 있다. 특히 인간전문가가 사전지식을 이용하여 초기 퍼지규칙을 만들고 난 후 모형을 학습하면 최적에 수렴하는 시간을 단축할 뿐 아니라, 전역최적치 도달가능성이 높아진다. 그러나 ANFIS모형은 입력변수의 수가 많아질수록 입력공간 차원의 증가로 인해 규칙의 수가 증가하여 계산시간의 증가와 함께 규칙에 대한 해석을 복잡하게 만드는 요인이 된다. 또한 ANFIS모형은 입력공간의 분할, 소속함수의 형태 결정, 규칙의 구조 결정, 추론 연산방법의 결정, 학습 방법과 계수의 결정 등 다양한 요소의 조합을 통해 구축된다. 따라서 실제 문제에 적용하기 위해서는 ANFIS 모형의 특성과 해당 도메인의 데이터 특성을 잘 파악하고 있는 전문가의 도움이 필요할 수도 있다.

향후 연구로써 본 연구에서 제안된 모형을 다양한 사회네트워크에 적용하여 일반성을 확보 것이 필요하다. 또한 본 연구의 결과 신뢰관계

망의 희박성 문제를 상당수 해결할 수 있지만 여전히 사회네트워크에서 사용자 행위가 파악되지 않는 사용자에 대해서는 신뢰관계의 추론이 어렵다. 이에 명시적으로 표현된 신뢰값과 과거 사용자행위에 기반한 신뢰예측 뿐만 아니라 입소문에 의한 신뢰예측(신뢰전파)을 결합한 예측방법의 제시가 필요할 것이다. 사용자는 제한된 숫자의 다른 사용자와 상호작용 관계를 가지는 경향이 있기 때문에 명시적인 신뢰관계 표현이나 과거 사용자행위에 기반한 신뢰예측을 통해 추론된 신뢰망은 작은 신뢰망이 고립되어 존재하는 형태가 될 가능성이 크다. 이에 신뢰전파를 이용하여 신뢰망간 결합을 시도할 수 있다. 이때 예측 방법에 따라 동일한 링크에 대해 상충된 예측값이 나올 수 있으며 이를 해결하는 로직의 개발이 필요할 것을 보인다.

### 참 고 문 헌

송희석, 김재경, “ANFIS에서 생성된 규칙의 해석용이성 평가”, 지능정보연구, 제15권, 제4호, 2009, 123-140.

Abdul-Rahman, A. and S. Hailes, “Supporting trust in virtual communities,” Proceeding of HICSS’00, 2000, 6007-6008.

Akhter, F., D. Hobbs, and Z. Maamar, “A fuzzy logic-based system for assessing the level of business-to-consumer (B2C) trust in electronic commerce”, Expert Systems with Applications, Vol.28, 2005, 623-628.

Atsalakis, G. S. and K. P. Valavanis, “Forecasting stock market short-term trends using a neuro-fuzzy based methodology”, Expert Systems with Applications, Vol. 36 (2009), 10696-10707.

Babuka, R. and H. Verbruggen, “Neuro-fuzzy methods for nonlinear

system identification”, Annual Reviews in Control, Vol. 27 (2003), 73-85.

Bersini, H. and G. Bontempi, “Now comes the time to defuzzify neuro-fuzzy models”, Fuzzy Sets and Systems, Vol. 90 (1997), 161-169.

Chang, E., PK. Thomson, TS. Dillon, and FK. Hussain, The Fuzzy and Dynamic Nature of Trust, 16th International Conference on Database and Expert Systems Applications (DEXA) in conjunction with International Conference on Trust, Privacy and Security in Digital Business, 2005.

Chiu, S., “Fuzzy Model Identification Based on Cluster Estimation”, Journal of Intelligent & Fuzzy Systems, Vol. 2, No. 3 (1994).

Efendigil, T., S. Onut and C. Kahraman, “A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis”, Expert Systems with Applications, Vol. 36 (2009), 6697-707.

Griffiths N., K-M. Chao, and M. Younas, “Fuzzy Trust for Peer-to-Peer Systems”, In: Proceedings of the 26th IEEE International Conference Workshops on Distributed Computing Systems, vol. 1, 2006, 73-78.

Golbeck, J. A., “Computing and applying trust in web-based social networks,” Doctor of C.S. dissertation, University of Maryland, 2005.