

## 집계형 구조의 CGK기반 입자모델 설계 및 성능평가

염찬욱\*, 곽근창\*\*

IT-Bio융합시스템농업교육연구단 전남대학교\*, 조선대학교\*\*

### CGK-Based Granular Model Design and Performance Evaluation of Aggregated Structure

Chan-Uk Yeom\*, Keun-Chang Kwak\*\*

Center for IT-Bio convergence system agriculture Chonnam University\*, Chosun University\*\*

**Abstract** - 본 논문은 컨텍스트기반 GK 클러스터링을 이용한 CGK기반 입자모델과 선형회귀, 신경회로망, 방사기저함수신경망을 계층적 구조로 결합한 집계형 구조의 CGK기반 입자 모델을 설계한다. 기존의 입자모델들은 대규모 데이터를 처리할 경우 입력변수에 따라 규칙의 수가 기하학적으로 증가하는 문제점이 발생한다. 이러한 문제를 해결하기 위해 일반적인 예측 모델들과 CGK기반 입자모델을 이용하여 합리적인 정보입자를 생성한다. 제안하는 방법의 타당성을 검증하기 위해 콘크리트 압축 강도 데이터베이스를 이용하여 예측 성능을 분석한다.

그림 1은 출력공간에서 컨텍스트가 3개 생성되고 각각의 컨텍스트에서 클러스터가 3개 생성되는 CGK기반 입자모델의 구조를 나타낸다. 결론부 변수는 출력공간에 생성된 컨텍스트를 나타내며, 전제부 변수는 각각의 컨텍스트에서 생성된 클러스터의 중감 즉 입력공간에 생성된 클러스터를 나타낸다. CGK기반 입자모델의 최종 출력  $Y$ 는 수식 1을 이용해 계산된다.

### 1. 서 론

인공지능 분야에서 추론 엔진은 새로운 정보를 추천하기 위해 지식기반에 논리적 규칙을 적용하는 시스템의 구성요소이다. 예측 관련 인공지능 모델로 퍼지추론시스템, 인공신경망 등 다양한 모델들이 있으면 서로 다른 처리 과정을 가지고 다양한 예측 문제를 해결해왔다. 입자컴퓨팅(GrC, Granular Computing)은 데이터 및 정보에서 지식을 추출하고 데이터를 추상화하는 과정에서 발생하는 정보입자(IG, Information Granule)라고 하는 정보개체의 처리와 관련된 컴퓨팅 이론이다[31].

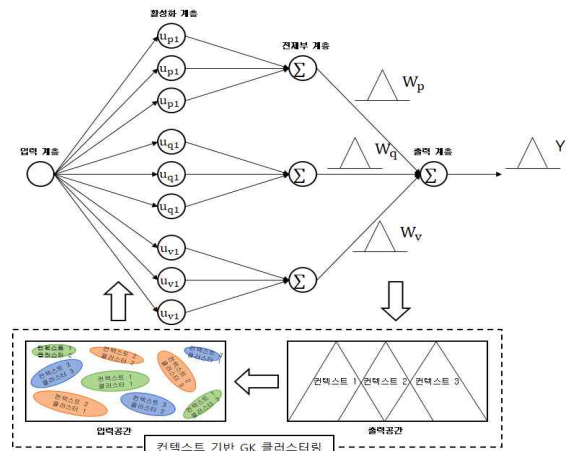
일반적으로 사용되는 퍼지추론시스템, 신경회로망, 딥러닝 모델들의 출력은 크리스프(crisp) 혹은 숫자 형태로 나타난다. 크리스프 형태의 출력은 실제 출력과 수치적인 오차를 계산할 수 있지만, 모델 출력과 실제 출력의 차이를 언어적으로 표현하는데 어려움이 있다. 하지만, 입자컴퓨팅을 사용할 경우 모델의 출력이 소프트(soft) 혹은 퍼지집합 등 형태로 나타나기 때문에 경계가 애매하거나 불확실한 데이터 및 정보를 처리할 때 효과적이다. 실제계에서 인간은 수치적인 표현보다 언어적인 표현을 주로 사용하며 불확실한 환경에서 추론하는 뇌는 수치적인 결과가 아닌 언어적인 결과를 통해 추론하고 결정한다. 이처럼 입자컴퓨팅은 인간이 생각하고 결정하는 과정을 나타낼 수 있다[33]. 입자컴퓨팅에 기초하여 입력공간 및 출력공간에 정보입자를 생성하는 컨텍스트기반 퍼지 클러스터링[47]이 연구되었다.

본 논문에서는 컨텍스트기반 GK 클러스터링을 이용한 CGK기반 입자모델을 설계하고 일반적으로 사용되는 예측 모델인 선형회귀, 인공신경망, 방사기저함수신경망을 결합하여 집계형 구조의 CGK기반 입자모델(AGM, Aggregated Granular Model)을 설계한다. CGK기반 입자모델은 출력공간에 컨텍스트를 생성하고 입력공간에 기하학적 형태의 클러스터를 생성하여 퍼지 규칙을 자동적으로 생성할 수 있는 장점이 있다, 하지만, 입력변수가 많은 대규모 데이터를 처리할 때 규칙의 수가 기하급수적으로 증가하여 의미 있는 규칙을 찾기 어려운 문제가 있다. 이러한 문제를 해결하기 위해 제안된 CGK기반 입자모델과 일반적인 예측 모델들을 계층적 구조로 결합하여 집계형 구조의 CGK기반 입자모델을 설계한다.

### 2. 본 론

#### 2.1 CGK기반 입자모델

CGK기반 입자모델은 출력공간을 고려한 컨텍스트기반 GK 클러스터링을 이용하여 출력공간에 컨텍스트를 생성하고 각각의 컨텍스트에서 기하학적 형태의 클러스터를 생성하여 설계된다.



〈그림 1〉 CGK기반 입자모델의 구조

$$Y = \sum_{\oplus} W_t \otimes z_t \quad (1)$$

여기서, 덧셈 및 곱셈 기호는 각각의 정보입자에 대해 완료된 덧셈 및 곱셈 연산을 나타낸다. CGK기반 입자모델 전제부를 처리하는 과정에서 퍼지집합이 생성된다. 이때, 컨텍스트기반 GK 클러스터링에 의해 생성된 클러스터는 CGK기반 입자모델의 은닉 계층으로 나타낼 수 있다. 은닉 계층과 출력 계층 사이의 언어적으로 설명 가능한 컨텍스트로 표현된다. 전체 컨텍스트를 이용하여 CGK기반 입자모델의 최종출력인 합계 값은 수식 2와 같이 표현할 수 있다.

$$Y = (z_{11} \otimes A_1 \oplus z_{12} \otimes A_1 \oplus \dots \oplus z_{1n1} \otimes A_1) \oplus \dots \oplus (z_{c1} \otimes A_c \oplus z_{c2} \otimes A_c \oplus \dots \oplus z_{cnc} \otimes A_c) \quad (2)$$

#### 2.2 집계형 구조의 CGK기반 입자모델

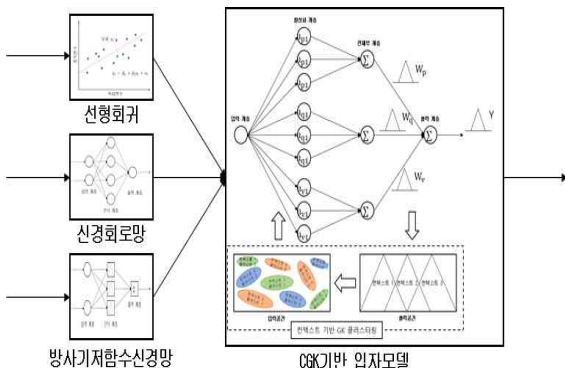
퍼지추론시스템 및 입자모델은 입력변수의 수가 증가함에 따라 규칙의 수가 기하급수적으로 증가하게된다. 대규모 규칙기반은 퍼지추론시스템 및 입자모델의 계산 효율성을 감소시키는 문제가 발생한다. 또한, 입자모델의 작동을 이해하기 어렵게 만들고 규칙과 소속함수를 조정하기 어려워진다. 다양한 예측 관련 응용 분야에는 제한된 양의 데이터를 가지고 있으므로 대규모 규칙기반의 퍼지추론시스템 및 입자모델의 일반화 가능성을 감소시킨다, 이러한 문제를 해결하기 위해 단일 퍼지추론시스템 및 단일 입자모델이 아닌 상호 연결된 구조인 계층적 구조의 입

자모델을 설계할 수 있다. 트리(tree) 구조로 배열되지 때문에 계층적 구조라고도 한다.

본 논문에서는 집계형(aggregated) 구조를 이용한 입자모델을 설계한다. 집계형 구조를 구축할 때 하위모델과 상위모델로 구분할 수 있다, 하위모델에는 선형회귀(LR, Linear regression), 신경회로망(NN, Neural Network), 방사기저함수신경망(RBFN, Radial Basis Function Network)로 구성되며 각각의 예측 모델의 출력을 상위모델인 CGK기반 입자모델의 입력으로 사용하여 최종 출력을 계산한다.

선형회귀는 입력변수와 출력변수 간의 선형 상관 관계를 모델링하는 것으로 단순 선형회귀는 설명변수를 기반으로 하고 다중 선형회귀는 둘 이상의 설명변수를 기반으로 한다. 신경회로망은 인지과학과 기계학습에서 생물학의 신경망에서 영감을받아 만들어진 알고리즘으로 시냅스의 결합으로 신경망을 구축한 노드가 학습을 통해 시냅스의 결합 강도를 변화시켜 문제를 해결하는 능력을 갖춘 모델을 의미한다. 방사기저함수신경망은 일반적인 신경망의 구조에서 활성화함수를 시그모이드함수가 아닌 방사기저함수를 이용한 신경회로망이다. 방사기저함수신경망은 은닉 계층이 하나로써 단순한 구조로 이루어져있으며 출력의 형태가 선형적이기 때문에 가중치 계산에 효과적이다.

위에서 설명한 하위모델에서 나온 출력을 결합하여 상위모델인 입자모델에 입력으로 사용하게 된다. 상위모델인 입자모델에서는 출력공간에 생성되는 컨텍스트의 수와 입력공간에 생성되는 클러스터의 수를 지정하여 정보입자를 생성하고 이를 바탕으로 집계형 구조의 최종 출력을 계산한다. 기존의 입자모델은 대규모 데이터를 처리할 때 입출력공간에서 의미 있는 정보입자를 생성하는 데 어려움이 있으며 계산시간이 오래 걸린다는 문제가 있다. 반면에, 본 논문에서 제안하는 집계형 구조의 입자모델은 하위모델들에서 나온 출력을 결합하여 생성된 데이터를 상위모델 입력으로 사용하기 때문에 의미 있는 정보입자를 생성할 수 있으며 처리 시간을 단축할 수 있는 장점이 있다.



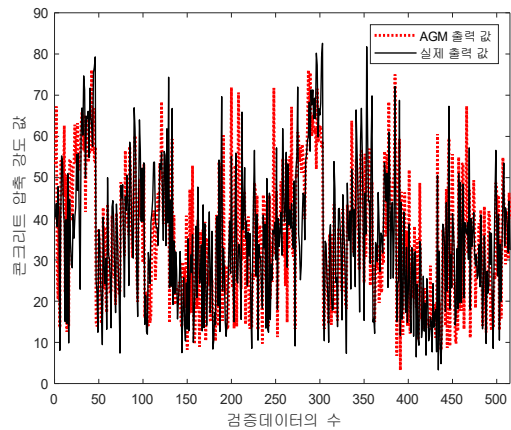
〈그림 2〉 집계형 구조의 CGK기반 입자모델의 구조

### 3. 실험 및 결과분석

본 논문에서 제안하는 방법인 집계형 구조의 CGK기반 입자모델(AGM)의 타당성을 검증하기 위해 예측 분야에서 사용되는 벤치마킹 데이터베이스인 콘크리트압축강도 데이터베이스[73]를 이용하여 실험을 진행했다. 실험에 사용한 데이터베이스는 대한정화대학교에서 수집된 데이터베이스로 인스턴스의 수는 1030, 변수의 수는 9로 구성되어있다. 출력변수는 콘크리트의 압축 강도이다.

본 실험에서는 일반적으로 사용되는 성능평가 방법이 아닌 정보입자 및 입자모델에 적합한 성능평가 방법인 성능지표[55] 방법을 통해 입자모델의 예측 성능을 평가한다. 정보입자의 범위성과 특이성을 이용하여 입자모델의 성능을 평가하는 성능지표 방법은 두 가지 특성이 상충(trade off) 관계이기 때문에 두 가지 특성을 동시에 극대화하는 방법을 모색해야한다. 실험 방법은 다음과 같다. 각각의 데이터베이스를 학습데이터 50%, 검증데이터 50%로 나누어 사용했다. CGK기반 입자모델과 제안하는 방법의 컨텍스트(P)의 수와 클러스터(C)의 수를 2에서 6까지 증가하면서 실험을 진행하였으며, 퍼지화 계수는 2로 고정했다.

또한, 컨텍스트를 균등하게 생성하는 방법과 유연하게 생성하는 방법을 사용하여 실험을 진행했다. 그림 3은 집계형 구조의 CGK기반 입자모델의 출력과 실제 출력의 오차를 보여주고 있으며, 표 1은 기존의 CGK기반 입자모델과 제안하는 방법의 예측 성능을 보여주고 있다,



〈그림 3〉 AGM의 출력과 실제 출력 비교(컨텍스트 균등하게 생성, 컨텍스트의 수 = 6, 클러스터의 수 = 4)

〈표 1〉 콘크리트 압축 강도 예측 실험 결과

입자모델	컨텍스트 생성 방법	성능지표		
		P / C	학습 성능지표	검증 성능지표
CGK-GM	균등하게 생성	6 / 4	0.4743	0.4700
AGM	균등하게 생성	6 / 4	0.5287	0.5208

### 4. 결 론

본 논문은 CGK기반 입자모델과 선형회귀, 신경회로망, 방사기저함수신경망을 계층적 구조로 결합한 집계형 구조의 CGK기반 입자모델을 설계했다. 실험 결과, 기존의 CGK기반 입자모델보다 제안하는 방법이 우수한 예측 성능을 보이는 것을 확인할 수 있었다. 향후 다양한 형태의 정보입자를 생성하는 연구와 입자모델 성능 최적화에 대한 연구를 수행할 계획이다.

### 감사의 글

본 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2018R1D1A1B07044907)

### [참 고 문 헌]

- [1] W. Pedrycz, A. V. Vasilakos, "Linguistic models and linguistic modeling", IEEE Transactions on Systems, Man, and Cybernetics, Part B, Vol. 29, No. 6, pp. 745-757, 1999.
- [2] X. Zhu, W. Pedrycz, Z. Li, "Granular models and granular outliers", IEEE Transactions on Fuzzy Systems, Vol. 26, No. 6, pp. 3835-3846, 2018.
- [3] L. Silva, R. Moura, A. M. P. Canuto, R. H. N. Santiago, B. Bedregal, "An interval-based framework for fuzzy clustering applications", IEEE Transactions on Fuzzy Systems, Vol. 23, No. 6, pp. 2174-2187, 2015.
- [4] UCI Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength>
- [5] T. T. Nguyen, X. C. Pham, A. W. C. Liew, W. Pedrycz, "Aggregation of classifiers a justifiable information granularity approach", IEEE Transactions on Cybernetics, Vol. 49, No. 6, pp. 2168-2177, 2018.