Journal of the Korea Academia-Industrial cooperation Society
Vol. 13, No. 2 pp. 562-574, 2012

http://dx.doi.org/10.5762/KAIS.2012.13.2.562

계층적 X-means와 가중 F-measure를 통한 시뮬레이션 모델 검증 기법

양대길[1], 황보훈[2], 천현재[3], 이홍철[4*]
[1]고려대학교 산업경영공학과, [2]텍사스 A&M 대학교 산업시스템공학과
[3]고려대학교 정보보호연구원, [4]고려대학교 정보경영공학부

Validation Technique of Simulation Model using Weighted F-measure with Hierarchical X-means (WF-HX) Method

Dae-Gil Yang[1], Hun Hwangbo[2], Hyun-Jae Cheon[3] and Hong-Chul Lee[4*]
[1]**Dept. of Industrial and Management Engineering, Korea University**
[2]**Industrial and Systems Engineering Division, Texas A&M University**
[3]**The Information Security Institute, Korea University**
[4]**Dept. of Information and Management Engineering, Korea University**

요 약 기존 대부분의 연구에서 사용하고 있는 시뮬레이션 검증 기법은 통계적 분석기법으로, 총 처리량이나 자원 이용률의 평균 및 분산을 통해 분석하여 왔다. 그러나 이러한 방식은 모델의 개별적인 요소들에 대한 신뢰성을 보장하기 어려웠다. 이를 해결하기 위해 제시된 방법이 가중 F-measure를 사용한 검증이다. 하지만 가중 F-measure는 Tact time 값 하나에 대해 하나의 클래스를 할당하기 때문에 수많은 Tact time 값들을 갖는 복잡한 시스템에 적용하기 어려운 문제를 가지고 있다. 한편, 가중치의 범위가 정해져 있지 않기 때문에 평가기준(Threshold)의 선정에 있어서 어느 정도의 수준이 만족할만한 수준인지 정하기가 어려웠다. 따라서 본 논문에서는 이러한 문제점을 개선하기 위해 군집분석을 적용한 가중 F-measure를 제시한다. 군집의 클래스화를 통해 클래스의 수를 현저히 줄일 수 있고 다양한 시스템으로의 적용 또한 가능해진다. 또한 객관성을 저하시키지 않는 범위 내에서 최소한의 가중치를 부여하는 방식으로 가중치의 범위를 지정하여 검증 방법을 향상시켰다. 이를 입증하기 위해 국내 'L사'의 LCD공정설비를 대상으로 시뮬레이션 모델링 및 환경을 구축하였고, 그 결과를 통해 타당성을 증명하였다.

Abstract Simulation validation techniques which have been employed in most studies are statistical analysis, which validate a model with mean or variance of throughput and resource utilization as an evaluation object. However, these methods have not been able to ensure the reliability of individual elements of the model well. To overcome the problem, the weighted F-measure method was proposed, but this technique also had some limitations. First, it is difficult to apply the technique to complex system environment with numerous values of interarrival time because it assigns a class to an individual value of interarrival time. In addition, due to unbounded weights, the value of weighted F-measure has no lower bound, so it is difficult to determine its threshold. Therefore, this paper propose weighted F-measure technique with cluster analysis to solve these problems. The classes for the technique are defined by each cluster, which reduces considerable number of classes and enables to apply the technique to various systems. Moreover, we improved the validation technique in the way of assigning minimum bounded weights without any lack of objectivity.

Key Words : Discrete Event Simulation, Validation technique, Cluster Analysis, X-means, Weighted F-measure

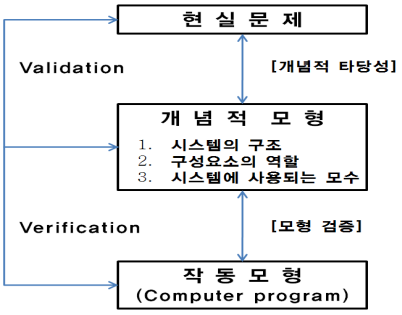
---

본 논문은 2단계 두뇌한국 BK21사업에 의하여 지원되었음.
*교신저자 : 이홍철(hclee@korea.ac.kr)
접수일 12년 01월 03일 수정일 (1차 12년 01월 30일, 2차 12년 02월 06일) 게재확정일 12년 02월 10일

1. 서론

현대 산업의 급속한 발전으로 시스템은 점점 대형화, 복잡화 되고 있다. 이러한 시스템들에 대하여, 기존 시스템의 효율적 개선을 위해 시스템을 평가하거나, 새로운 시스템을 검증할 때, 간단한 분석적인 방법이나 비현실적인 가정들을 수반하는 수리적인 방법은 적합하지 않다. 이러한 경우를 위하여 다양한 시뮬레이션 검증기법이 활용되어 왔다[9].

시뮬레이션은 기존 시스템이 여러 가지 상이한 운용상황에 놓이게 될 때 각각의 성능 및 대안을 쉽게 비교측정할 수 있으며, 분석적(Analytical) 방법을 적용할 수 없는 실세계의 복잡한 시스템에 대해 확률적(Stochastic) 요소를 가지고 모델링 할 수 있다[9]. 다음은 일반적인 시뮬레이션 모델의 벨리데이션(Validation)과 베리피게이션(Verification)이다.



[그림 1] 시뮬레이션 모델의 Validation과 Verification
[Fig. 1] Validation and Verification of simulation model

그림 1[12]과 같이 시뮬레이션 모델이 분석 목표에 맞게 충실히 시스템 동작을 재현했는지, 대상 시스템을 이해할 수 있도록 잘 표현했는지에 대하여 타당성을 검토하는 작업이 벨리데이션(Validation)이다. 시뮬레이션에서 얻은 결과를 적용하기에 앞서, 의사결정자들은 모델이 타당한가에 대한 의문을 제기한다. 따라서 시뮬레이션에 대한 신뢰도를 높이기 위해서는 시뮬레이션을 시작할 때부터 미래의 사용자와 함께 시뮬레이션 모델을 만들고 그러한 모델의 타당성을 검토하는 것이 매우 중요한 작업이라고 할 수 있다[7].

대부분의 연구에서 사용하고 있는 시뮬레이션 검증 기법 중 통계적 기법은 총 처리량이나 자원 이용률의 평균 및 분산을 통해 분석을 하기 때문에 모델의 개별적인 요소들에 대해 신뢰성을 보장하기 어려웠다. 이를 해결하기 위하여 특정 통계량으로 전체에 대해 검증하는 것 대신에 개별적인 시계열 데이터를 검증하는 것이 바람직하다. 따라서 검증 과정에서 비교할 출력 데이터로 총 처리량이나 자원의 이용률 등이 아닌 시계열 데이터를 사용하고, 단순히 평균이나 분산에 대해 비슷한지를 검증하는 것이 아니라 시간의 흐름에 따라 도출된 개별적인 값들을 검증해야 한다[6,12].

따라서 본 연구에서는 이러한 시뮬레이션 모델에 대해 F-measure 개념을 활용하여 보다 구체적으로 타당성을 검토하고자 한다. 이를 위하여 기존에 연구된 가중 F-measure 검증 방법의 문제점들을 보완하여 보다 다양한 시스템에 적용 가능하도록 제시할 것이다. 먼저 군집분석(Cluster Analysis)을 통해 클래스를 재정의 하고, 다음으로 가중치의 재정의를 통해 검증의 정밀성을 확보하고자 한다.

2. 선행연구

2.1 시뮬레이션 모델의 검증 방법

기존에 시뮬레이션 모델 검증 방법의 선행연구들은 Sargent[16]와 Balch[2]의 연구에 잘 정리되어 있다. Sargent는 정성적 또는 정량적인 방법을 통해 모델을 검증하였고, Balch는 테스트 방법에 대해 분류하였다.

시뮬레이션 모델과 대상 시스템을 비교하는 데에는 지금까지 다양한 방법론과 알고리즘들이 연구되어 왔으며, Balci는 시뮬레이션 모델의 검증 및 테스트 방법 중 일부인 통계적 방법들을 표 1과 같이 소개하고 있다[2].

[표 1] 시뮬레이션 모델 검증을 위한 통계적 기법
[Table 1] Statistical techniques for the validation of simulation model

Analysis of variance(ANOVA)
Confidence intervals/regions
Nonparametric goodness-of-fit tests • Kolmogorov-Smirnov Test • Chi-square Test
Nonparametric tests of means • Mann-Whitney-Wilcoxon Test • Analysis of Paired Observations
Regression analysis
Theil's inequality coefficient
t-Test
Time series analysis • Spectral Analysis • Schruben's standardized time-series analysis • Intervention Analysis

분산분석(Analysis of variance, ANOVA)은 분산, 총평균과 각 데이터 평균의 차이에 의해 생긴 분산의 비교를 통해 만들어진 F분포를 이용하여 가설검정을 하는 방법이다[21]. 그러나 분산분석은 평균에 대한 검정을 하며, 이렇게 F분포를 이용하는 과정은 t분포를 이용하는 과정과 한 치의 차이도 없는 동일한 결과를 얻어낸다. 신뢰구간(Confidence interval), 대응 관측분석(Analysis of Paired Observations), t-검정 (*t*-Test) 등의 통계적 기법들은 두 개의 시스템의 평균에 대한 검증이며, 표본평균과 표본표준편차를 사용하여 통계량을 계산하고 그에 따라 모델을 검증한다[20]. 윌콕슨 순위합 검정(Wilcoxon rank-sum test) 또는 맨-위트니 검정(Mann-Whitney test)은 정규분포를 따르지는 않지만 독립이며 연속인 표본들의 두 분포의 평균이 같은가를 검정할 때 사용된다. 그러나 이러한 방법들과 같이 평균에 대해 검정할 경우에는 평균과 전체적인 편차의 정도뿐만 아니라 각각의 시점에서 관측값이 같은가를 판단해야 하는 시계열 데이터의 검증에서 그 타당성을 평가하기가 어렵다.

카이제곱검정(Chi-square test)은 시뮬레이션 모델과 실제 시스템에서 각 값의 빈도 차이를 확인한다[18]. 그러나 빈도를 측정하는 구간의 설정이 어려우며 비정규성을 가진 경우에는 분포가 구간에 더 민감하게 반응한다.

K-S 검정(Kolmogorov-Smirnov test, K-S test)은 분포에 크게 영향을 받지 않으며, 카이제곱 검정과는 다르게 모델의 각 데이터들의 누적 확률 값과 실제 시스템 데이터들의 누적 확률 값을 비교함으로써 구간의 설정에 제약을 받지 않는다[22]. 이러한 방법들은 단순히 각 값들의 빈도에 대해서만 평가를 하기 때문에 데이터들의 순서나 트렌드와 상관없이 값의 분포만 비슷하면 귀무가설을 채택한다. 이 역시 시계열 데이터의 검증에는 적합하지 않다.

Theil의 부등 계수(Theil's inequality coefficient, TIC)는 두 시스템의 값의 차이의 제곱 합과 각각의 시스템의 값의 제곱의 합을 더한 값과의 비율로 얻어 진다[13]. 이 방법 또한 데이터들의 합으로 계수 값을 산출하므로 각각의 데이터들이 의미 있는 값을 가지고 있는가를 판단하기 어렵다.

회귀분석(Regression analysis)은 실제 시스템과 모델에 대해 $y=\beta0+\beta1x$라는 1차 선형회귀방정식을 세우고 상수항 $\beta0= 0$, 기울기 $\beta1= 1$의 만족 여부를 검정한다[17]. 그러나 회귀분석을 통해 나온 값이 유사한 모델을 기각해 버리는 경우가 많아 모델의 검증에 적합하지 않다고 판단되어 현재는 t-검정을 통해 평균과 분산이 같은지를 검정하도록 제안하고 있다[5].

앞에서 살펴본 바와 같이 시계열 분석(Time series analysis)을 제외한 나머지 방법들은 시스템을 하나의 통계량으로 전체에 대해 평가한다는 측면에서 개별적인 데이터의 검증에 적합하지 않다[6]. 시계열 분석에서는 개별적인 시계열 데이터에 대한 검증이 가능하고 이는 스펙트럴 분석(Spectral analysis), Schruben의 표준화 시계열 분석(Schruben's standardized time-series analysis), 간섭분석(Intervention Analysis), 가중 F-measure법 (Weighted F-measure Method) 등으로 나눌 수 있다[12,23].

스펙트럴 분석에서는 정상성(Stationarity)을 갖는 시계열을 사인(sine)함수와 코사인(Cosine)함수로 이루어진 무수히 많은 주기함수(Periodic function)들의 합으로 나타낸다. 이 과정을 푸리에 변환(Fourier transform)이라고 하며, 이를 이용하여 시계열에 대한 각각의 주기 함수들의 영향력을 스펙트럴 밀도 함수로 나타냄으로써 영향력이 큰 주기를 찾는다[23]. 이 방법은 값의 차이가 큰 주기에 대해서는 비교적 우수한 능력을 보이지만 차이가 작은 주기에서는 신뢰도가 낮으며, 고차원적 수학적 기술을 요하는 단점을 가지고 있다[23].

Schruben의 표준화 시계열 분석은 긴 반복수를 필요로 하며, 점 추정치의 분산 추정에 편기되어지는 경향을 가지고 있다. 또한 정상시계열(Stationary time-series)을 모델화하기 위해서 중심극한 정리를 사용하는 한계점이 있다. 다음으로, 간섭분석(Intervention Analysis)은 간섭 사건들에 의해 영향을 받는 시계열의 정도를 고려하여 진행된다. 간섭분석은 하나의 시계열 상에서 간섭사건들의 영향정도를 알아내기 위하여 개발되어진 기법이다[12]. 하지만 이들 분석방법은 초기에 설정된 가정으로 인해 다양한 시스템으로의 적용에 한계가 있다[12].

2.2 가중 F-measure

F-measure는 데이터 마이닝 분야에서 사용되는 분류기의 평가방법으로 이 개념을 이용한 가중 F-measure는 시뮬레이션의 구성요소인 이벤트 기반으로 클래스를 분류하고, 이렇게 분류된 각각의 클래스에 대하여 타당성을 평가하는 방법이다[6]. 다른 시계열 분석 방법들이 시계열의 정상성이나 다양한 가정에 의존하고, 이를 제외한 다양한 통계적인 방법들이 하나의 통계량으로 전체에 대해 검증하는 반면, 가중 F-measure는 시계열 데이터에 대한 특별한 가정 없이 각각의 이벤트 집합을 평가할 수 있다는 의미에서 정밀한 검증을 가능하게 한다. 하지만 이는 도착간격 시간(Interarrival time) 데이터를 검증의 대상으로 삼고, 각각의 개체(Entity)의 서로 다른 도착간격 시간에 대해 동일하게 인식하여 하나의 클래스를 할당함으로써 복잡한 시스템에 적용하기 어렵다는 단점이 있다. 따라서 본 연구에서는 복잡한 시스템에도 적용할 수 있

는 가중 F-measure를 고안하여 보다 다양한 시스템에서 정밀한 검증을 할 수 있는 방법을 제안하고자 한다. 이를 위해 클래스를 정의하고 구분할 수 있는 방법이 필요하며, 군집 분석을 통해 클래스를 구체화하고자 한다.

2.3 군집 분석의 적용

군집 분석은 데이터 분류와 이미지 처리 등의 많은 실용적 분야에 적용할 수 있는 탐색적 데이터 분석 기법 중 하나이다. 군집화(Clustering)는 입력 데이터 집합을 유사한 결과 값들의 군집들로 구분하여 데이터 집합 속에 존재하는 의미 있는 정보를 얻는 과정이다.

군집화를 위한 알고리즘들은 다양하게 제안 되어왔으며, 크게 계층적 기법 (Hierarchical algorithms)과 분할기법 (Partitioning algorithms)으로 나눌 수 있다. 가장 많이 사용되는 분할 기법은 보통 k-means라고 불리는 MacQueen(1967)의 Lloyd's algorithm이다[14].

이 알고리즘은 d-차원 공간상의 점으로 이루어진 입력 집합을 필요로 한다. 목표는 k개의 군집 중심을 찾고, 입력 집합을 나누는 것이다. k-means의 장점은 단순하고, 데이터 분석의 기초가 되며, 일반적으로 효율적으로 적용된다는 데에 있다. 하지만 사용자가 미리 군집의 개수를 설정해야 하고, 예외 값을 다루기 어렵다는 단점을 가지고 있다.

기존 알고리즘의 대부분은 고정된 k값에서 최적의 집단을 찾아내는 k-means나 EM알고리즘의 Wrapper 알고리즘으로, 이들은 특정 조건을 만족할 때까지 집단의 분할 / 병합을 통해 k를 증가 / 감소시켜 나간다. Pelleg와 Moore(2000)가 제시한 X-means 알고리즘은 일정 범위의 k값에 대해 k-means 알고리즘을 수행하고 각 결과를 BIC(Bayesian Information Criterion) 점수(score)를 이용하여 평가한다[4]. 이는 여타의 추가적인 파라미터를 필요로 하지 않으면서 최적의 k값을 찾아주기 때문에 매우 효과적이다. 따라서 본 연구에서는 군집분석 방법으로, k-means의 장점인 시공간 상의 효율성을 가지면서 군집의 개수 선정에 객관성을 확보할 수 있는 X-means를 적용하고자 한다.

3. 계층적 X-means와 가중 F-measure를 통한 모델 검증 방법

기존에 제시된 가중 F-measure는 각각의 이벤트 집합을 평가할 수 있다는 의미에서 정밀한 검증을 가능하게 하지만, Tact time 값 하나에 하나의 클래스를 할당하기 때문에 규모가 큰 시스템에 적용하기 어려웠다. 다시 말해서, 행렬 $(S_n \times R_n)$구성 시 그 크기가 너무 방대해지는 문제가 발생한다. 이는 다음 그림 2에서 나타난다.

		시뮬레이션 데이터			
		S_1	S_2	····	S_n
실제 데이터	R_1	F_{11}	F_{12}	····	F_{1n}
	R_2	F_{21}	⋱		
	⋮	⋮		⋱	
	R_n	F_{n1}			F_{nn}

[그림 2] F-measure의 적용을 위한 혼동행렬
[Fig. 2] Confusion matrix for the application of F-measure

한편, 기존의 연구에서는 F-measure의 이분법적 정확성에 국한된 평가를 보완하기 위하여 가중치를 부여하였다. 가중치는 S_n과 R_n의 각각의 클래스에 대해 평가의 다양성 확보를 위해 부여된 것이다. 하지만 가중치의 범위가 정해져 있지 않기 때문에, 평가기준(Threshold)을 선정하는 데에 있어서 어느 정도의 수준이 만족할만한 수준인지 정하기가 어렵다.

본 연구는 이러한 가중 F-measure의 한계점들을 클래스와 가중치의 재정의를 통해 보완하고, 이를 통해 보다 정밀하고 객관적인 시뮬레이션 모델 검증 방법을 제안하고자 한다.

3.1 계층적 X-means를 통한 클래스의 재정의

3.1.1 Tact time의 정의와 특성

본 연구에서는 이벤트들의 결과물인 Tact time을 기본 데이터로 하여 개별적 시계열 데이터를 검증한다. Tact time이란 요구하는 생산목표를 달성하기 위하여 제품 하나를 생산하는 관리 기준 시간을 말한다. 즉, 식 (1)과 같이 제품 한 단위를 추가적으로 생산하는 데에 필요한 시간을 의미한다.

$$Tact\ Time = \frac{T}{Q} \tag{1}$$

여기서 T는 생산라인의 가용시간이며 Q는 하루 동안의 총생산량이다. 또한 위의 식에서의 Tact time을 식 (2)와 같이 평균의 개념으로 본다면, 최종적으로 작업물의 공정이 끝나는 지점(A)에서의 도착간격 시간(Interarrival time)을 그 제품 한 단위를 추가적으로 생산하는 데에 걸리는 시간이라고 볼 수 있다. 이는 다음의 식 (3)에 나타나 있다.

$$Tact\ Time = Average\left(Tact\ Time_i\right) \tag{2}$$

$$Tact\ Time_i = interarrival\ \ time_i^A \tag{3}$$

제품 각각에 대한 Tact time을 검증의 기본 단위로 삼는다면, 이를 통해 각각의 로드(Entity)들이 어떤 이벤트와 프로세스를 거쳐 최종 목적지까지 도착하는지를 좀 더 쉽게 확인할 수 있다. 결국 Tact time은 특정 로드에 대해 그 로드가 거쳐 온 이벤트 집합들의 시간을 대표하는 대푯값으로 사용될 수 있으며, 이벤트 기반의 시뮬레이션 모델의 검증에 매우 적합하다.

3.1.2 X-means 알고리즘

[표 2] X-means 알고리즘
[Table 2] X-means algorithm

```
1  k = 1; //number of clusters
2  p_score=-∞; //previous score
3  c_score = evaluate_BIC(data,k); //current score
4  p_mean = mean(data); //previous mean
5
6  while c_score>p_score
7       p_score=c_score
8       for cl=1 to k // cluster index
9        pca=do_pca(data(cl)); //local split direction
10
11          for sd = 1 to data_dimension
   //split direction
12              c_mean=remove(p-mean,cl);
    //remove one cluster
13              c_mean=
     add_mean(c_mean,c_mean±pca(sd));
     //add n clusters
14              s_mean(cl,sd)=
     K_means(data(cl),c_mean,2);
     //split
15              s_score(cl,sd)=
     evaluate_BIC(data,k+1,s-mean(cl,sd));
     //evaluate
16       end;
17     end;
18
19     [c_score,c_mean]=max(s_score,s-mean);
20     if c_score>p_score
21     k=k+1; //increse the number of clusters
22     p_mean=K_means(data,c_means,k+1);
23    else
24     break;
25    end;
26    end;
27
28  return p_mean;
```

X-means 알고리즘은 하나의 집단에서 시작하여 BIC(Bayesian Information Criterion) 점수를 최대로 증가시키는 집단을 반복적으로 분할하는 하향식 방법이다[4]. BIC 점수를 최대한 증가시키다가 더 이상의 점수 증가가 없을 때 분할을 멈춘다. 표 2는 X-means의 수행과정을 나타낸 수도코드(Pseudo code)이다.

하나의 집단이 분할된 후에 각 집단에 할당되는 데이터들은 k-means알고리즘을 통해 결정된다. line22에서는 전체 데이터에 대해 k-means알고리즘을 수행하고 있으며 이는 국부적인 집단 분할 중 BIC 점수를 최대로 하는 분할에 대해 전체적으로 집단들을 갱신하는 과정이다. line 22의 세부적 단계는 Step 1 초기 중심점 선택, Step 2, 중심점 할당 및 k개의 군집 형성, Step 3, 중심점 재설정, Step 4 중심점의 변화로 정의된다. Step 1의 초기 중심점 선택[1]은 임의로 선택하는 방법을 많이 사용하지만, 그 결과는 종종 빈약하다. 또한, 임의로 선택된 초기 중심점의 사용에 대한 문제는 여러 번의 반복적인 수행으로도 극복할 수 없다. 따라서 본 연구에서는 유클리드 거리(Euclidean distance, L_2)개념[11]을 사용한 k-means++[3]의 초기 중심점 선택법을 사용한다. 그 방법은 다음의 표 3과 같다.

[표 3] 초기 중심점 선정
[Table 3] Selected as the initial center

1	1번째 center는 랜덤으로 정한다.
2	for I = 2 ~ k
3	모든 인스턴스 x마다 아래의 확률을 부여한다
4	$p(x) = D(x)^2 / \sum x' D(x')^2$ [x'는 인스턴스집합 x의 원소들]
5	$D(x) = x \sim center(\min ; distance)$
6	i th center ; random set

k-means++는 확률을 기반으로 중심점들이 밀집되지 않고 이상치는 되도록 피하며 k개의 초기 중심점을 선정하는 방법이다.

각각의 중심점을 만들어 갈 때, 지금까지 만들어진 중심점 중에 가장 가까운 중심점과의 거리를 $D(x)$라 한다. 위의 표 3의 Line 4에서 분모는 모든 인스턴스 x의 $D(x)^2$의 합을 뜻하므로 모든 x에 대해서 같다. 그러므로 분자인 $D(x)^2$이 큰 값을 가질 때 군집 중심으로 선정될 확률 $p(x)$도 커진다.

군집의 질을 평가하는 목적함수로는 오차 제곱합(SSE : Sum of the squared error)[15]을 사용한다. k-means의 실행으로 두 개의 서로 다른 군집이 주어졌을 때 가장 작은 오차의 제곱을 가진 것을 선호하게 되는데, 그 이유는 해당 군집의 중심점들이 군집내의 점들을 더 잘 나타내기 때문이다. SSE는 다음과 같이 정의된다.

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} dist(c_i, x)^2 \quad (4)$$

여기서 $dist$는 유클리드 공간에서 두 객체 사이의 표준 유클리드 거리(L_2)를 의미한다. 이러한 가정 하에서 군집의 SSE를 최소화하는 중심점은 군집의 평균임을 알 수 있다.

i번째 군집의 중심점(평균)은 다음의 식 (5)와 같다.

$$c_i = \frac{1}{m_i} \sum_{X \in C_i} X \tag{5}$$

line22의 Step 3과 4에서는 SSE를 직접적으로 최소화하는 일을 한다. 먼저 Step 3에서는 주어진 중심점들에 대한 SSE를 최소화하기 위해 점들을 가장 가까운 중심점에 할당하여 군집들을 형성한다. Step 4에서는 SSE를 추가로 최소화하기 위해 중심점들을 다시 계산한다. 수렴할 때까지 그 과정을 계속 반복하는데 그 이유는 데이터의 인접성 측정 기준과 군집화의 목표에 따라서 중심점이 변화할 수 있기 때문이다. 일반적으로 군집화의 목표는 보통 목적함수(Objective function)에 의해 표현된다. 인접성 측정 기준과 목적함수를 지정하면, 중심점의 선택은 수학적으로 결정될 수 있는 것이다.

데이터 D와 서로 다른 값의 k를 가지는 군집화 모델 M_j가 주어질 경우, 모델들을 비교하는 방법에는 여러 가지가 있다. 앞서 언급하였듯이 본 연구에서는 모델의 적합성을 판단하는 기준으로 BIC를 사용하고, 이는 다음의 식 (6)과 같이 나타난다[4].

$$BIC(M_j) = \hat{l}_j(D) - \frac{p_j}{2} \cdot \log R \tag{6}$$

이때 p_j는 파라미터의 개수, R은 데이터의 개수, 그리고 $\hat{l}$은 데이터의 Log-likelihood값을 나타낸다. 식 (6)의 BIC는 Rissanen의 MDL(Minimum description length)과 동일한 것으로 알려져 있다.

여기서 Log-likelihood는 주어진 데이터 D를 가장 잘 표현하는 모델에서 최대값을 가지게 된다. 각 집단의 평균과 분산에 대한 MLE(Maximum likelihood estimation)[10]값은 각각 식 (7),(8)과 같이 구할 수 있다.

$$\hat{\mu_i} = \frac{1}{N_i} \sum_{j \in C_i} x_j \tag{7}$$

$$\widehat{\sum}_i = \frac{1}{N_i - 1} \sum_{j \in C_i} (x_j - \hat{\mu_i})^2 \tag{8}$$

식 (8)에서 N_i는 i번째 집단에 속하는 데이터의 개수를 나타내고, C_i는 i번째 집단에 속하는 데이터의 부분집합을 나타낸다. 평균과 분산에 대한 MLE 값을 이용하여 각 데이터 포인트의 확률은 식 (9)와 같이 구할 수 있다.

$$\hat{P}(x_j) = \frac{R_i}{R} \frac{1}{\sqrt{2\pi}\,\widehat{\sigma^M}} \exp\left(-\frac{1}{2} \| x_i - \mu_{(i)} \|^2\right) \tag{9}$$

여기서 $\mu_{(i)}$와 σ는 각각 j번째 데이터 포인트가 속한 집단의 평균과 분산을 나타낸다. 다음은 식 (9)를 이용하여 i번째 집단에 속하는 데이터의 Log-likelihood를 구한 식이다.

$$l(D) = \log \prod_i P(x_i) = \sum_i \left(\log \frac{1}{\sqrt{2\pi}\,\sigma^M} - \frac{1}{2\sigma^2} \| x_i - \mu_{(i)} \|^2 + \log \frac{R_{(i)}}{R}\right) \tag{10}$$

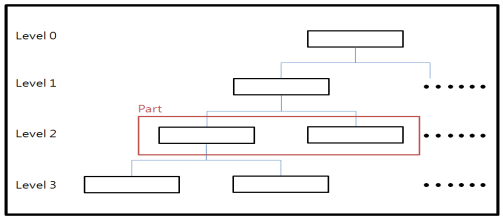
이 모델의 파라미터 개수는 가우시안 컴포넌트 확률, K개의 D차원 평균 벡터, 그리고 K개의 $D \times D$분산 행렬에 의해 식 (11)과 같이 정의된다.

$$p_j = (K-1) + D \cdot K + \frac{D(D+1)}{2} \cdot K \tag{11}$$

3.1.3 계층적 X-means를 통한 클래스 구성

X-means 알고리즘을 통해 Tact time 값들을 군집화하고 얻어진 각 군집들을 하나의 클래스로 선정함으로써 소수의 클래스를 생성할 수 있게 된다. 그러나 하나의 클래스가 반드시 하나의 이벤트 집합을 대표한다고 할 수 없다. 또한, 유사한 Tact time 값들이 하나의 클래스에 할당되었지만 Tact time 값들이 유사하다고 해서 반드시 해당 이벤트 집합들이 유사한 것은 아니다. 따라서 하나의 클래스는 여러 개의 이벤트 집합들을 포함하고 있는 하나의 상위 그룹으로 생각할 수 있다.

검증의 수준을 높이고자 한다면 서로 다른 이벤트 집합들에 대해 다시 개별적인 검증을 해야 할 필요가 있다. 요구되는 검증 수준은 시스템의 특성에 따라 상이하며, 다양한 검증 수준을 처리할 수 있는 일반적인 검증 방법이 필요하다. 따라서 본 연구에서는 X-means 알고리즘을 여러 수준(Level)에 걸쳐 계층적으로 수행하는 방식을 제안하며, 이러한 수준의 최대값은 시스템의 특성을 반영하여 의사결정자에 의해 결정되도록 한다. 아래의 그림은 이러한 계층적 X-means의 구조를 도식화 한 것이다.



[그림 3] 계층적 X-means 구조도
[Fig. 3] hierarchical X-means structure

그림 3의 Level 0은 초기 시뮬레이션 데이터이고 Level 1은 X-means 알고리즘을 통해 한 번의 군집화를 수행한 결과이다. 이렇게 얻어진 군집들을 하나의 파트(part)로 분류한다. 한편, 파트에 속한 하나의 군집은 그 스스로가 다음 수준에서 파트를 가질 수 있다. 즉, Level 1의 각각의 군집은 Level 1에서의 특정 파트에 속할 뿐 아니라 Level 2에서 자신의 파트를 가질 수 있게 된다. Level 1에서 네 개의 군집이 생성되었다면, Level 1에서는 한 개의 파트가 Level 2에서는 네 개의 파트가 존재하게 된다. 여기서 하나의 파트에 대해 하나의 혼동행렬을 구성할 수 있고 이를 통해 가중 F-measure의 적용이 가능하다. 하나의 파트는 부분군집(Subcluster)들의 집합이며, 가중 F-measure에서 사용될 클래스들의 집합이라 할 수 있다.

이러한 계층의 수준은 클래스가 잘못 설계되었다고 판단될 때, 또는 특정 클래스에 대해 보다 구체적인 분석을 요구할 때 등의 필요한 경우에 확장이 가능하다. 앞서 언급하였듯이 군집 계층의 최대 수준$(N>0)$은 의사결정자의 결정사항이다. 이렇게 수준을 끝없이 늘려나가게 되면$(N\to\infty)$, 결국 최종 수준에서는 하나의 데이터 값에 대해 하나의 군집이 생성되며 이는 기존의 가중 F-measure에서의 클래스 정의 방식과 같은 결과를 얻어낸다.

3.2 가중치의 재정의

기존의 가중 F-measure 값은 $(-\infty,1]$ 에서 정의되므로 평가 기준을 어느 정도의 값으로 선정해야 하는지를 판단하기가 어렵다. 따라서 $[0,1]$ 의 범위를 정의하는 것이 보다 유용한 정보를 제공한다.

3.2.1 $[0,1]$ 범위의 가중 F-measure

기존의 가중 재현성(WR)과 가중 정밀도(WP)를 구하는 식은 다음의 식 (12), 식 (13)과 같다. ω_{aj}와 ω_{ia}는 식 (14)와 식 (15)에 나타나 있으며, 이 가중치들의 최대값이 정의되지 않음에 따라 가중 재현성과 가중 정밀도의 최소값이 정의되지 않는다. 이로 인해, 두 값의 조화평균인 가중 F-measure 또한 최소값을 갖지 않게 된다.

$$WR_a=\frac{\sum_{j=1}^{n}f_{aj}-\sum_{j\neq a}(1+\omega_{aj})f_{aj}}{\sum_{j=1}^{n}f_{aj}} \tag{12}$$

$$WP_a=\frac{\sum_{i=1}^{n}f_{ia}-\sum_{i\neq a}(1+\omega_{ia})f_{ia}}{\sum_{i=1}^{n}f_{ia}} \tag{13}$$

$$\omega_{aj}=\frac{|s_j-s_a|}{E(s_n)} \tag{14}$$

$$\omega_{ia}=\frac{|r_i-r_a|}{E(r_n)} \tag{15}$$

$(1+\omega_{aj})$와 $(1+\omega_{ia})$ 는 $[1,\infty)$ 에서 정의되므로, 이를 $[0,1]$ 범위에서 정의되는 ω^*_{aj}와 ω^*_{ia}로 대체함으로써 가중 F-measure를 [0, 1]에서 정의한다. 이를 통해 가중 재현성과 가중 정밀도 및 가중 F-measure는 $(1+\omega^*_{aj})$, $(1+\omega^*_{ia})$ 부분을 각각 ω^*_{aj}와 ω^*_{ia}로 대체하여 다음의 식 (16), (17), (18)과 같이 구할 수 있다.

$$WR_a=\frac{\sum_{j=1}^{n}f_{aj}-\sum_{j\neq a}\omega^*_{aj}f_{aj}}{\sum_{j=1}^{n}f_{aj}} \tag{16}$$

$$WP_a=\frac{\sum_{i=1}^{n}f_{ia}-\sum_{i\neq a}\omega^*_{ia}f_{ia}}{\sum_{i=1}^{n}f_{ia}} \tag{17}$$

$$WF_a=\frac{2\left\{\sum_{j=1}^{n}f_{aj}-\sum_{j\neq a}^{n}\omega^*_{aj}f_{aj}\right\}\left\{\sum_{i=1}^{n}f_{ia}-\sum_{i\neq a}^{n}\omega^*_{ia}f_{ia}\right\}}{2\sum_{i=1}^{n}\sum_{j=1}^{n}f_{aj}f_{ia}-\sum_{i\neq a}^{n}\sum_{j=1}^{n}\omega^*_{ia}f_{aj}f_{ia}-\sum_{i=1}^{n}\sum_{j\neq a}^{n}\omega^*_{aj}f_{aj}f_{ia}} \tag{18}$$

[0, 1]에서 정의되는 새로운 가중치에 대해서는 다음절에서 구체적으로 논의하도록 한다.

3.2.2 가중치 ω

데이터마이닝 분야에서 사용되는 F-measure는 잘못 예측된 모든 클래스에 대해 1의 가중치를 부여하는 것으로 생각할 수 있다. 황보훈 등(2009)은 F-measure를 $n=2$ 이상의 멀티클래스에 대해 평가하면서, 잘못 예측된 $(n-1)$개 클래스에 대해 실제 값과의 차이에 따라 패널티를 부과하기 위해 가중치 개념을 추가하였다.

한편, 앞서 제시한 것과 같이 기존의 $(1+\omega)$ 대신에 ω를 사용하게 될 경우에는 ω값이 0에 가까워질수록 모델의 정확도와 상관없이 가중 F-measure는 상당히 높은 값을 갖게 된다. 이는 모델의 검증 방법으로써 적합하지 않다. 따라서 가중치는 가중 F-measure 값에 크게 영향을 미치지 않으면서(1에 가까우면서), 잘못 예측된 클래스들 사이의 차이를 구분할 수 있도록 설계되어야 한다.

ω^*_{aj}를 구하기 위해 $[0,1]$ 범위의 다음의 비율을 정의한다.

$$\frac{|s_j-s_a|}{s_n-s_i} \tag{19}$$

이는 n개 클래스의 점 추정치(각 클래스의 평균값)의 최대 차이에 대한, 평가하고자 하는 클래스 a와 임의의 잘못 예측된 클래스 j의 점 추정치 차이의 비율로 정의된다. 이러한 비율의 정의는 잘못 예측된 클래스들이 전체 구간에 비해 얼마나 잘못 예측되었는지를 구분할 수 있게 한다. 이 비율을 이용하여 ω^*_{aj}를 다음의 식 (20)과 같이 정의한다.

$$\omega^*_{aj} = (1-\alpha) + \frac{\alpha|s_j - s_a|}{s_n - s_1} \quad (20)$$

식 (20)에 의해 ω^*_{aj}는 $[1-\alpha,\ 1]$에서 정의되며, α값을 통해 가중치의 최소값을 조정할 수 있다. α는 행렬에서 각 클래스별 효과적인 가중치 적용을 위해 부여해준 값이다. 1에 가까운 값을 갖는 가중치를 설계하기 위해 α값은 0.1이하로 설정하는 것이 바람직할 것이다. α를 0.1로 설정할 때의 가중치는 $1-\alpha = 0.9$ 를 최소값으로 갖게 된다. ω_{ia}도 같은 방법으로 식 (21)과 같이 구할 수 있다.

$$\omega^*_{ia} = (1-\alpha) + \frac{\alpha|r_i - r_a|}{r_n - r_1} \quad (21)$$

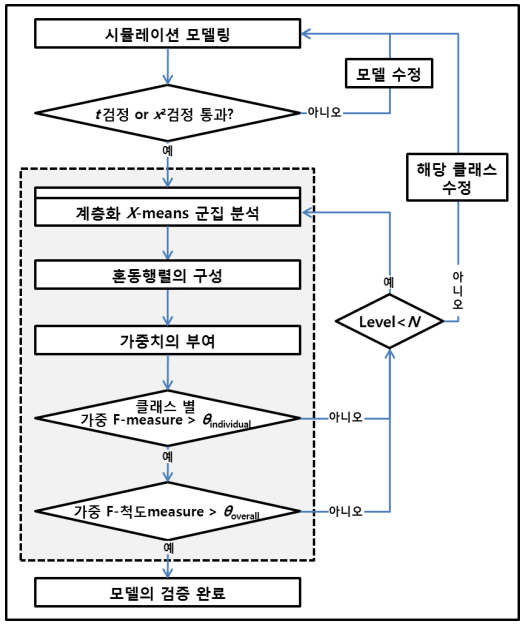
이러한 가중치들을 통해 가중 F-measure에서 평가 기준의 선정이 보다 객관성을 갖게 될 수 있다.

3.3 계층적 X-means를 통한 가중 F-measure (Weighted F-measure with Hierarchical X-means Method, WF-HX Method)

계층적 X-means와 가중 F-measure를 사용한 시뮬레이션 모델의 검증 과정에 대해 그림 4로 나타내었다. 기존 연구에서 한계점으로 지적되었던 보편성과 객관성을 확보하기 위하여 군집분석 과정이 추가되고 가중치가 새롭게 정의 되었다.

전체적인 진행 과정은 다음과 같다. 시뮬레이션 데이터를 t-검정이나 카이제곱 검정 등의 기존 방법들을 통해 1차적 평가를 진행한 후 시뮬레이션 데이터에 대해 X-means 알고리즘을 통해 초기 군집화를 실시한다.

군집화 과정을 거친 군집을 혼동행렬의 각 클래스로 구성한다. 여기에 새롭게 정의된 가중치를 통해 가중 F-measure 값을 구한 후 $\Theta_{individual}$, $\Theta_{overall}$ 과 비교 분석한다. 검증 결과 잘못 설계된 클래스들 또는 검증 수준을 높이고자 하는 클래스들에 대해 군집 계층의 수준이 N이 될 때까지 그림 4에 음영처리 되어 있는 계층적 X-means를 통한 가중 F-measure의 과정을 반복한다. 현 상황에서 수정이 필요한 클래스들을 검토하고 수정한 후 전체 과정을 반복한다. 군집 계층의 수준을 만족하는 모든 클래스들의 가중 F-measure의 값이 $\Theta_{individual}$값 이상이고 모든 파트(part)들의 가중 F-measure 평균값이 $\Theta_{overall}$값 이상이면 검증 과정을 마무리한다.



[그림 4] 계층적 X-means를 통한 가중 F-measure
[Fig. 4] Weighted F-measure with Hierarchical X-means

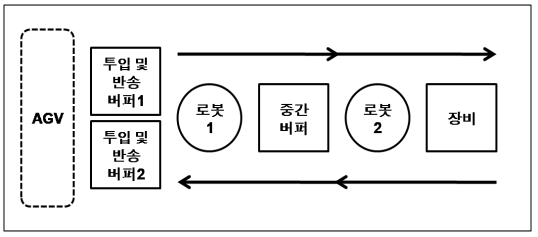
이 과정에서 군집 분석 기법을 사용함으로 인해 합리적인 수의 클래스를 지정할 수 있게 되고, X-means 알고리즘을 계층적으로 구성함으로써 모델의 부분적인 검증이 가능할 뿐만 아니라 검증의 보편성까지 얻을 수 있게 된다. 또한, 군집 계층의 수준을 의사결정자가 직접 결정하고 추가적으로 확장할 수 있기 때문에 각 요소에 대해 이상 판단을 한 경우, 즉시 추가적인 분석 및 수정이 가능하다.

4. 사례 연구(Case Study)

앞서 제시한 검증 방법론의 효과를 검증하기 위해서 이 장에서는 국내 'L사'에서 수행한 시뮬레이션을 통해 클래스와 가중치가 재정의 된 가중 F-measure를 사용하여 기존의 방법들과의 검증결과의 차이에 대해 살펴볼 것이다. 본 사례 연구에서는 기존의 가중 F-measure 연구에서 사용된 데이터와 동일한 LCD공정 시스템 데이터를 사용하였다.

4.1 시뮬레이션 모델 제시

아래에 제시된 모델은 국내 'L'사의 LCD공정 시스템이다. 해당 시스템에서는 TV 및 모니터에 장착될 디스플레이 액정이 생산되며 모델은 로봇 2대와 버퍼 3개, 그리고 장비 1대로 구성되어있다. 로봇은 각 로드(Entity)를 운반하는 역할을 하며, 버퍼에는 해당 시스템으로 로드의 투입을 위한 버퍼 2개와 선·후처리 공정을 하는 버퍼가 있다. 각 로봇과 장비의 용량은 1EA이며, 모델의 배치는 아래의 그림 5와 같이 구성된다.

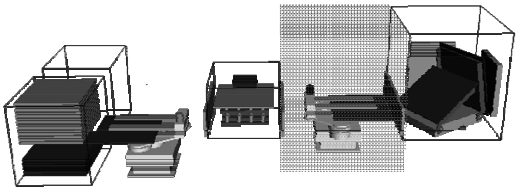


[그림 5] 시스템 모델의 배치도
[Fig. 5] Layout of the system model

총 25단으로 구성된 투입 및 반송버퍼와 2단의 중간버퍼 그리고 장비 사이의 로봇으로 구성된다. 투입 및 반송버퍼에서는 선·후처리 공정이 이루어지며 장비는 총 2가지 공정을 수행하고 모든 로드는 2가지 작업을 차례로 거치게 되며 장비에는 한번에 2개의 로드가 실릴 수 있다. 시스템의 전체적인 공정은 무인운반차(AGV : Automated guided vehicle)를 통해 로드들이 투입 버퍼에 투입된 후 순서대로 쌓이고, 로봇 1을 거쳐 중간버퍼에 투입된다. 선처리 공정을 마친 후 로봇 2를 통해 장비로 이동하여 공정 1을 마친 후 공정 2를 수행한다. 공정을 모두 마친 로드는 로봇 2를 통해 중간 버퍼에 투입되며 후처리 공정이 완료되면 로봇 1을 거쳐 다시 반송(투입) 버퍼에 차례로 쌓이게 된다.

4.2 모델 검증 과정

본 연구에서는 시뮬레이션 모델링 S/W로 AutoModⅡ를 사용하였다. AutoModⅡ는 시스템을 효율적으로 분석하기 위해 만들어진 시뮬레이션 프로그램으로 GUI기반의 데이터 입력과 3D를 지원함으로써 사용자가 보다 쉽게 공정시스템을 구현할 수 있도록 만들어진 소프트웨어이다.



[그림 6] 시뮬레이션 모델 실행 화면
[Fig. 6] Simulation model execution screen

시뮬레이션 모델에서 얻어진 데이터와 실제의 데이터에 대해 기존에 연구되어진 방법들로 모델 검증을 해 본 결과가 표 4와 같이 나타났다. 모델 검증을 위해 SPSS12.0k와 Excel을 이용했다.

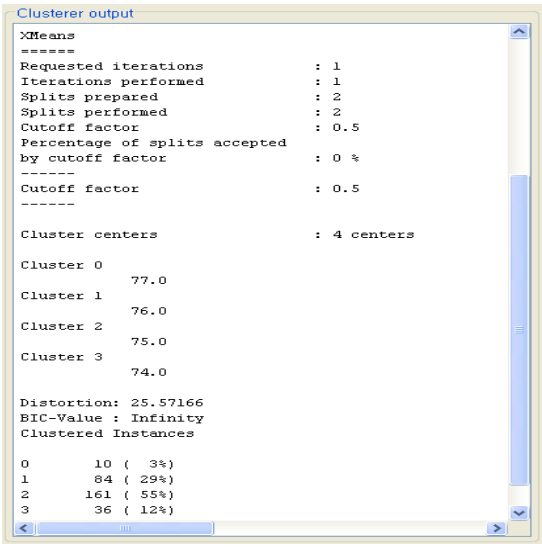
[표 4] 타 방법들에 의한 초기 모델 검증
[Table 4] The initial model validation by other methods

검증 방법	통계량 값	기각역	유의 확률
t-검정	0	-	1
윌콕슨 순위합 검정	-0.028	-	0.978
카이제곱 검정	5.253	-	0.148
K-S 검정	0.010	-	1
TIC 검정	0.0004	0.3	-

대부분의 검정을 통해서 거의 1에 가까운 매우 높은 유의 확률이 얻어졌으며, TIC 검정과 같은 경우 0.3보다 작은 값을 가지면 충분히 잘 설계되었다고 할 수 있는데, 얻어진 TIC 값은 0.3에 비해 매우 작은 값이다. 카이제곱 검정의 경우에는 다른 방법에 비해서 유의확률이 작게 나왔지만 유의수준 0.05보다는 크기 때문에 두 시스템이 유사하다는 결론을 내린다. 따라서 기존의 방법들은 현재의 모델과 실제 시스템은 같다는 가설을 어떠한 근거로도 기각하지 못한다.

다음으로 군집분석을 적용한 가중 F-measure 검증을 시행한다. 검증 과정을 진행하기에 앞서 의사결정자에 의해 평가기준(θ)과 군집 계층의 최대 수준(N)이 결정되어야 한다. 평가 기준은 $\Theta_{individual}$=0.6, $\Theta_{overall}$=0.8로 기존의 연구와 동일하게 한다. 또한, 시스템이 단순하기 때문에 군집 계층의 최대 수준 $N=1$로 한다.

군집분석을 위해서 Waikato Environment Knowledge Analysis(WEKA) 3.6.4를 사용하였으며, 그림 7은 해당 모델의 검증을 위해 X-means 군집 분석을 실행한 결과를 보여준다.



[그림 7] X-means 군집분석 결과
[Fig. 7] X-means clustering analysis

위의 그림 7에서는 군집의 중심과 개수를 보여주며, 군집의 개수는 BIC 점수에 의해 4개로 결정되었다. 그림 8의 그래프는 나란히 정렬된 데이터끼리 하나의 군집을 나타낸다.

[그림 8] 시뮬레이션 데이터의 군집화 그래프
[Fig. 8] Graph Clustering of simulation data

이러한 군집들을 바탕으로 구성한 혼동행렬은 아래의 표 5와 같다. 시뮬레이션 데이터를 바탕으로 군집을 생성하였고 실제 데이터의 클래스들과 시뮬레이션 데이터의 클래스들이 동일해야 하기 때문에, 실제 데이터는 동일한 값을 갖는 시뮬레이션 데이터가 포함된 클래스에 포함시키도록 한다.

[표 5] 혼동행렬 구성
[Table 5] Configuration of confusion matrix

S \ R	군집0	군집1	군집2	군집3	군집4	군집5
군집0	0	0	0	0	0	0
군집1	0	0	4	5	1	0
군집2	0	1	27	40	16	0
군집3	2	7	39	98	13	2
군집4	0	1	15	14	5	1
군집5	0	0	0	0	0	0

이때 실제 데이터 중에 시뮬레이션 데이터에는 없는 값이 존재한다면 이들은 어떤 군집, 즉 어떤 클래스에도 포함되지 못한다. 이러한 데이터들에 대해 시계열상의 해당 위치에 별도의 군집을 추가하여 포함시켜 준다. 표 5에서 군집 0과 군집 5가 이에 해당한다.

혼동행렬을 통해 가중 재현성, 가중 정밀도, 가중 F-measure를 구한 결과는 다음의 표 6과 같다. 군집 0과 군집 5에서 가중 재현성의 값이 0이므로 가중 F-measure의 값은 존재하지 않으며 이는 '-'로 표기하였다. 표 6의 결과에서 가중 F-measure의 값이 군집 3을 제외한 나머지 값들은 현저히 낮게 나타난다. 현재의 군집 계층의 수준이 1이고 군집 0과 군집 5는 실제 데이터에만 존재하는 클래스이므로, 이들과 군집 3 클래스를 제외한 나머지 클래스(군집 1, 군집 2, 군집 4)들에 대해 각각의 이벤트가 제대로 반영되었는지 검토한다. 검토결과 기압 변경 과정의 처리에 문제가 있는 것을 발견하고 모델을 수정하였다.

[표 6] 초기 모델의 척도
[Table 6] Measure of the initial model

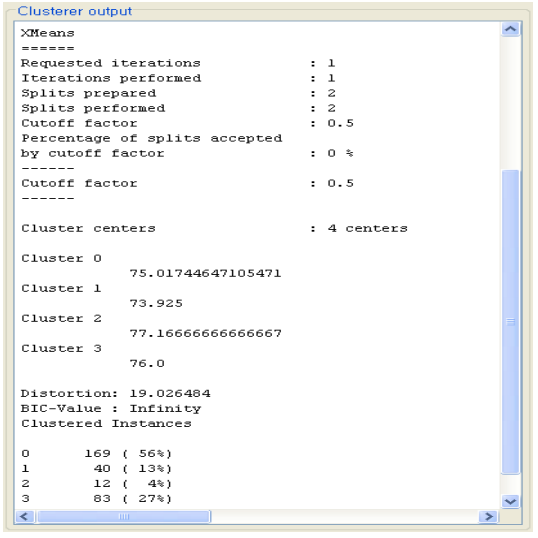
	군집0	군집1	군집2	군집3	군집4	군집5
가중 재현성	0	0.066	0.372	0.638	0.197	0
가중 정밀도	0.04	0.060	0.369	0.654	0.201	0.033
가중 F-measure	–	0.063	0.370	0.646	0.199	–

모델 수정 결과 초기 모델보다 확연히 데이터간의 유사성을 확인할 수 있었으며, 이때 기존의 방법들로 모델의 검증을 한 결과 표 7과 같은 유의 확률을 얻었다.

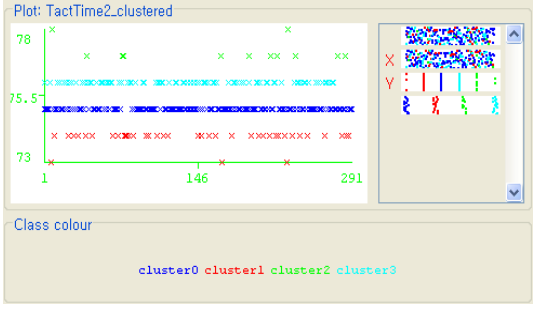
[표 7] 타 방법들에 의한 초기 모델 검증
[Table 7] The initial model validation by other methods

검증 방법	통계량 값	기각역	유의 확률
t-검정	-0.107	-	0.914
윌콕슨 순위합 검정	-0.145	-	0.885
카이제곱 검정	0.279	-	0.960
K-S 검정	0.083	-	1
TIC 검정	0.0002	0.3	-

수정 후에도 대부분의 검정을 통해서 1에 가까운 유의확률을 얻어낼 수 있으며, 카이제곱 검정의 유의 확률도 눈에 띄게 증가한 것을 알 수 있다. 다음으로 수정된 모델을 통해 X-means 군집 분석을 실시한다. 아래의 그림 9는 수정된 모델의 군집 분석결과를 나타낸다. 수정된 모델에 대해서도 4개의 군집으로 군집화한 것을 알 수 있다.



[그림 9] X-means 군집분석 결과
[Fig. 9] X-means clustering analysis



[그림 10] 수정된 모델의 군집화 그래프
[Fig. 10] Clustering a graph of the modified model

그러나 수정된 모델의 군집은 실제 데이터의 값들을 모두 포함하고 있기 때문에 추가적인 군집을 부여하지 않는다. 이는 모델의 수정이 올바로 진행되었음을 보여준다.

분석결과를 통해 표 8과 같이 (4×4)의 혼동행렬이 구성된다. 이때의 가중 F-measure 값은 표 9와 같이 구할 수 있으며, 군집 1과 군집 2에 대해서는 1에 가까운 값이 나오는 것을 확인할 수 있다. 그러나 군집 3의 가중 F-measure가 $\Theta_{individual}$=0.6보다 낮게 나타나므로 해당 클래스에 대해 추가적인 수정 작업을 진행한다.

[표 8] 수정된 모델의 혼동행렬
[Table 8] Confusion matrix of the modified model

S \ R	군집 0	군집 1	군집 2	군집 3
군집 0	110	2	0	44
군집 1	3	36	0	1
군집 2	0	0	11	1
군집 3	44	0	0	39

[표 9] 수정된 모델의 척도
[Table 9] Measure of the revised model

	군집 0	군집 1	군집 2	군집 3
가중 재현성	0.7127	0.9068	0.9768	0.4831
가중 정밀도	0.7091	0.9513	1	0.4732
가중 F-measure	0.7109	0.9285	0.9883	0.4781

다음은 또 다시 모델을 수정한 후의 혼동행렬이며 혼동행렬의 구성을 통해 모델의 수정이 제대로 이루어 졌음을 눈으로도 확인할 수 있다.

[표 10] 최종 모델의 혼동행렬
[Table 10] Confusion matrix of the final model

S \ R	군집 0	군집 1	군집 2	군집 3
군집 0	146	3	0	8
군집 1	2	36	0	0
군집 2	0	0	11	0
군집 3	7	1	1	76

[표 11] 최종 모델의 척도
[Table 11] Measure of the final model

	군집 0	군집 1	군집 2	군집 3
가중 재현성	0.9326	0.9513	1	0.8976
가중 정밀도	0.9440	0.9068	0.9229	0.9071
가중 F-measure	0.9382	0.9285	0.9599	0.9023

표 11에서 가중 F-measure의 값이 모두 $\Theta_{individual}$=0.6 이상이고, 이들의 평균이 $\Theta_{overall}$=0.8 이상이므로 여기서 검증을 완료한다.

5. 결 론

본 연구는 시뮬레이션 분석 절차 중 모델의 타당성과

신뢰성을 입증하는 모델의 검증단계에서 계층적 X-means와 가중 F-measure를 적용하여 모델의 타당성을 검증하는 방법을 제시하였다. X-means를 통해 클래스를 재구성하여 복잡한 시스템에서도 소수의 클래스를 구성할 수 있도록 하였고 이를 계층적으로 구성하여 의사결정자가 원하는 수준까지 모델을 검증하고 수정할 수 있도록 하였다. 또한, 본 연구는 기존에 연구되어진 가중치의 범위를 [0,1]로 한정하여 평가 기준의 선정에 보다 객관성을 갖게 하였다.

이를 검증하기 위해 가장 먼저, 시뮬레이션 모델을 구축하였고 도착간격시간(Interarrival time)을 검증의 대상으로 하였다. 첫 번째 시뮬레이션 모델의 타당성 검증 결과 X-means를 적용한 가중 F-measure값이 군집0, 군집3, 군집5를 제외한 모든 군집에 대해 $\Theta_{individual}$ 값 이하로 낮게 나타났다. 이후 시뮬레이션과 실제 시스템의 상이점을 발견하여 수정 후 다시 모델을 검증한 결과 $\Theta_{individual}$과 $\Theta_{overall}$값 모두를 만족하여 구축한 시뮬레이션 모델이 타당함을 검증하였다. 또한 검증의 신뢰성을 높이기 위해 사례연구에서 기존 방법론과 비교하였다. 비교대상은 t-검정, 윌콕슨 순위합 검정, 카이제곱 검정, K-S검정, TIC검정을 이용하여 동일하게 모델의 검증을 수행한 결과 계층적 X-means와 가중 F-measure를 통한 검증 기법의 우수성을 입증할 수 있었다.

References

[1] A. Kumar, Y. Sabharwal, and S. Sen, “A simple linear time (1+ε)-approximation algorithm for k-means clustering in any dimensions”, Proc. 45th IEEE Symposium on Foundations of Computer Science, pp.454-462, 2004.

[2] Balci, o., “Validation, verification, and testing techniques throughout the life cycle of a simulation study", Annals of Operations Research, Vol.53, pp.121-173, 1994.

[3] David Arthur, Sergei Vassilvitskii, “k-means++: the advantages of careful seeding”, Proceeding SODA '07 Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp.1027-1035, 2007.

[4] Dan Pelleg and Andrew Moore, “X-means : Extending K-means with Efficient Estimation of the number of clusters", proceedings of the 17th International Conference on Machine Learning, pp.727-734, 2000.

[5] Harrison, S. R., “Regression of a model on real-system output: An invalid test of model validity”, Agricultural systems, Vol.34, No.3, pp.183-190, 1990.

[6] Hwangbo hun , Cheon hyeon-jae , Lee hong-cheol, "Using the weighted F-measure of Trace-Driven Simulation model verification methods," Journal of Simulation, Vol.18, No.4, pp.185-195, 2009.

[7] Jack P.C. Kleijnen, “Verification and Validation of Simulation Models", European J. of Operational Research, pp.82, 145-162, 1995.

[8] Jack P.C. Kleijnen, Bert Bettonvil, Willem Van Groenendahl, “Validation of Trace -Driven Simulation Models : Regression Analysis Revised", Proceedings of the WSC, pp.352-359, 1994.

[9] Lee young-hae, Gwak seong-geun, Kim suk-han, "Under a distributed environment, research on Web-based simulation", Journal of Korea Society for Simulation, No. 7 No. 2, pp.79-90, 1998.

[10] L. Excoffier and M. Slatkin, “Maximum - Likelihood Estimation of Molecular Haplotype Frequencies in a Diploid Population", Oxford Journals Life Sciences & Medicine Molecular Biologyand Evolution, Volume12, Issue5, pp. 921-927, 1995.

[11] L. Yen, D. Vanvyve, F. Wouters, F. Fouss, M. Verleysen, M. Saerens, “Clustering using a random walk based distance measure", proceedings - European Symposium on Artificial Neural Networks Bruges April, pp.317-324, 2005.

[12] Lee young-hae, Choe gwan-grim, "Intervention Analysis techniques, utilizing research on Validation of simulation models," Journal of Korea Society for Simulation, pp.175, 1997.

[13] Murray-Smith, D. J., “Methods for the external validation of continuous system simulation models”, Mathematical and computer modelling of dynamic systems, Vol.4, No.1, pp.5-31, 1998.

[14] S. Lloyd, “Least Squares Quantization in PCM", IEEE Transactions on Information Theory, Vol.28, pp.129-137, 1982.

[15] S. N. Crozier, D. D. Falconer, S. A. Mahmoud, “Least sum of squared errors (LSSE) channel estimation”, Radar and Signal Processing, IEEE Proceedings F, Volume 138, Issue 4, pp.371-378, 1991.

[16] Sargent, R. G., “Verification and Validation of Simulation Models", Proceedings of the 2010 Winter Simulation Conference, pp.166-183, 2010.

[17] Smith, E. P. and Rose, K. A., "Model goodness- of-fit analysis using regression and related techniques", Ecological modelling, Vol.77, No.1, pp.49-64, 1995.

[18] Wei, X. C. and Li, E. P., "Reflection of transmitting antenna in reverberation chamber and its effect on chi-square validation", Antennas and propagation society

international symposium 2006, pp.3573-3576, 2006.
[19] Box G. E. P, Jenkins G. M, and Reinsel G. C., “Time Series Analysis ; Forecasting and Control”, pp.463-479, Prentice Hall 3rd, 1994.
[20] Kang seok-bok, “Statistical estimation and hypothesis testing”, pp.156-318, Gyeongmunsa, 2002.
[21] Kim dae-hak, “Utilizing spss, analysis of variance”, pp.39-65, Gyowoosa, 2004.
[22] Ross, S. M., Simulation 4th Ed., pp.294-306, Elsevier Academic Press, 2006.
[23] Law, A. M. and Kelton, W. D., “Simulation modeling and analysis 3rd Ed.”, pp.283-290, McGraw-Hill, 2000.

양 대 길(Dae-Gil Yang) **[정회원]**

- 2008년 2월 : 단국대학교 산업공학과 (산업공학 전문공학사)
- 2010년 9월 ~ 현재 : 고려대학교 일반대학원 산업경영공학과 석사과정

<관심분야>
SCM, Discrete Event Simulation, Data Mining,

황보 훈(Hun Hwangbo) **[정회원]**

- 2008년 2월 : 고려대학교 산업공학과 학사 (산업공학 공학사)
- 2010년 2월 : 고려대학교 일반대학원 산업경영공학과 석사 (공학석사)
- 2011년 3월 ~ 현재 : Texas A&M Univ. Industrial and Systems Engineering 석사과정

<관심분야>
Discrete Event Simulation, SCM, Integer Programming

천 현 재(Hyeon-Jae Cheon) **[정회원]**

- 1997년 2월 : 인천대학교 산업공학과 학사 (산업공학 공학사)
- 1999년 2월 :고려대학교 산업공학과 석사 (공학석사)
- 2006년 2월 : 고려대학교 산업공학과 박사 (공학박사)
- 2009년 2월 ~ 현재 : 고려대학교 정보보호연구원 연구교수

<관심분야>
SCM, 데이터 마이닝, Discrete Event Simulation

이 홍 철(Hong-Chul Lee) **[정회원]**

- 1983년 2월 : 고려대학교 산업경영공학과 (산업공학 공학사)
- 1988년 2월 : Univ. of Texas 산업공학과 석사 (공학석사)
- 1993년 2월 : Texas A&M Univ. Industrial and Systems Engineering 박사 (공학박사)
- 1996년 3월 ~ 현재 : 고려대학교 정보경영공학부 교수

<관심분야>
모델링&시뮬레이션, SCM, 생산 및 물류 정보시스템, PLM