

# Design of CBFS Analyzer for the Digital Dashboard of Nuclear Power Plant

Sung Wook Jeong<sup>1</sup>, Chang Jin Kang<sup>1</sup>, Sangmun Shin<sup>1</sup>, Yonghee Lee<sup>2</sup>, Jun Woo Kim<sup>1\*</sup>

<sup>1</sup>Department of Industrial & Management Systems Engineering, Dong-A University, Busan 604-714, Republic of Korea

<sup>2</sup>Co-won Nuclear Engineering Center, Co-won Co. Ltd., Busan 604-714, Republic of Korea

## ABSTRACT

원자력 발전소에서 발생하는 문제점들의 상당수가 기계적 원인이 아닌 인적 오류에서 기인함이 알려지면서 원자력 발전소의 안전 문화 확립이 중요한 과제로 부각되고 있다. 안전 문화 확립을 위해서는 관련 분야 종사자 개개인의 의식 수준 향상과 함께, 정량적으로 안전 문화 수준을 모니터링하고 평가할 수 있는 체계가 필요하다. 아울러, 안전 문화 수준과 관련된 다양한 변수들을 지속적으로 수집하고, 기존에 기계적 측면의 모니터링 위주이던 원자력 발전소 제어실 디지털 대시보드에서 이러한 변수들에 대한 모니터링과 분석을 통합적으로 제공하여야 한다. 그러나 이러한 과정에서 관리 대상 변수들의 수가 급격히 증가하게 되고, 이는 사용자의 모니터링을 어렵게 하거나 여러 가지 분석 알고리즘들의 수행시간과 정확성을 악화시킬 수 있다. 이에 본 논문에서는 원자력 발전소 디지털 대시보드에 적절한 특징 선택(feature selection) 분석기를 탑재하는 것을 제안하며, 구체적으로는 CBFS(correlation based feature selection) 방법에 기반한 분석기의 기본적인 구조를 설계하고자 한다. 특히, 제안하는 분석기는 기존의 원자력 발전소 관리 시스템과의 유기적으로 통합되어 디지털 대시보드에서 편리하게 사용할 수 있으며, 특징 선택 과정에서 발생하는 정보들을 효과적으로 제공하여 대화형(interactive) 분석이 가능하다. 이러한 특징 선택 분석기는 향후 다양한 데이터 분석 알고리즘과 결합되어 원자력 발전소의 안전 문화 수준 제고에 기여할 수 있을 것으로 기대된다.

Keywords: Feature selection, Correlation based feature selection, Safety culture, Digital dashboard, Nuclear power plant main control room

## 1. Introduction

다양한 산업 시설 중에서도 원자력 발전소는 사고 시 광범위한 위험을 발생시킨다는 특수성을 가지며, 이로 인해 원자력 발전소에서의 안전 관리의 중요성은 지극히 높다(Carvalho et al., 2006). 아울러, 최근 원자력 발전소에서 발생하는 문제점들의 원인은 기술적, 기계적인 측면보다는 주로 인간적인 요소(human factor)에 있음이 지적되고 있고, 이에 따라 원자력 발전소 관련 분야 종사자들에 대한 안전 문화의 중요성이 부각되고 있다(Lee and Harrison, 2000). 이에 따라 종사자 개개인을 위한 상세한 업무 지침 등의 개발과 함께, 안전 문화를 모니터링하고 평가하기 위해 다양한 데이터의 수집과 분석이 요구된다.

한편, 기존에도 원자력 발전소에서는 발전기의 작동 상태 및 운영 관리와 관련된 다양한 정보들을 데이터베이스로 구축하고 이를 제어실에서 모니터링 및 분석해왔으며, 최근에는 정보통신기술의 발달과 함께 다양한 기능을 가진 디지털 대시보드가 이러한 목적을 위해 활용되고 있다. 아울러, 이러한 디지털 대시보드는 데이터들의 단순 조회 및 모니터링 기능 외에도 다양한 위험 요인 평가 방법을 제공하는 것이 바람직하다(Stephane, 2012). 나아가, 향후에는 기존의 기술적 지표들 외에도 안전 문화와 관련된 다양한 변수들을 함께 측정 및 관리하고, 이들에 대한 분석 방법도 함께 제공하는 것이 필요할 것이다.

그러나, 이렇게 안전 문화와 관련된 데이터들이 추가되는 경우에는 관리 대상 데이터들의 차원

(dimensionality)이 높아지면서 다양한 문제점들을 야기할 수 있다. 이에 본 논문에서는 원자력 발전소의 디지털 대시보드에 탑재될 수 있는 특징 선택(feature selection) 분석기를 설계하고자 한다.

## 2. Correlation based Feature Selection

고차원의 데이터는 다변량 분석이나 데이터마이닝 등의 분석을 실행하는데 많은 시간이 소요되고, 과잉적합(overfitting) 등으로 인해 분석 결과의 품질도 저하될 수 있다는 등의 문제점들을 가지고 있다. 이러한 현상은 흔히 차원의 저주(curse of dimensionality)라고 일컬어지며, 이를 피하기 위해서는 데이터 분석을 하기 전, 사전처리(pre-processing) 단계에서 데이터를 적절히 가공하는 것이 바람직하다. 특징 선택은 차원 감소(dimensionality reduction)를 위해 행하는 대표적인 사전처리 방법으로, 데이터가 가진 필드 중 중요한 것만을 남기고 나머지는 삭제하는 작업을 의미한다(Tan et al., 2005).

특징 선택 방법은 여러 가지가 개발되어 왔다. CBFS(correlation based feature selection)는 이러한 방법들 중 하나로, 지정된 반응변수에 높은 영향력을 갖는 독립변수들의 조합을 효과적으로 탐색할 수 있다(Hall and Smith, 1997; Hall, 1999). 아울러, CBFS는 여타의 교사 학습(supervised learning) 알고리즘과는 독립적으로 작동하는 필터 방식의 특징 선택 기법으로, 특히 범주형 변수들로 구성된 데이터에 적합하다. CBFS의 가장 큰 특징은 이러한 범주형 독립변수들의 집합과 반응변수 간의 상관관계를 측정하는 방법에 있으며, 전통적인 데이터 집합 불순도(impurity) 지표 중의 하나인 엔트로피(entropy)를 이용하여 두 개의 범주형 변수 간 상관관계를 측정한다. 이에 기반하여, CBFS는 반응변수에 대해 높은 상관관계를 갖는 독립변수들로 구성되면서 이러한 독립변수들 간의 상관관계는 높지 않은 경우, 바람직한 독립변수 집합으로 평가한다. 이러한 CBFS는 다른 데이터 분석 기법과 결합되었을 때, 좋은 결과를 보이는 것으로 알려져 있다(Shin et al., 2007).

그러나 원자력 발전소 제어실과 같은 중앙집중형 대시보드에서 이러한 특징 선택 기법을 활용하기 위해서는 다음과 같은 사항들이 고려되어야 한다. (1) 기존에 구축되어 있는 legacy 시스템 및 데이터베이스와 유기적으로 통합되어야 하고, 이를 위해 적절한 분석 모듈 및 사용자 인터페이스가 필요하

다. (2) 전통적으로 특징 선택 기법 및 데이터 분석 알고리즘들이 자동화된 방법으로 분석을 실시하고 최종적인 결과만을 사용자에게 제공해왔던 것과는 달리, 빅데이터(big data)로 일컬어지는 방대한 분량의 데이터의 등장과 함께 데이터의 분석 과정에서 인간의 감각적인 인지 능력과 추론 능력을 적극적으로 활용할 필요성이 제기되고 있다(Gharehchopogh, 2010). 특히, 향후 안전 문화와 관련된 다양한 변수들에 대한 수집이 이루어질 경우, 특징 선택 과정에서도 현장 전문가에 의한 인간의 개입(human intervention)이 필요할 가능성이 있다. (3) 앞의 사항들을 고려하여 분석 과정이 유연하고 분석 과정에 대한 정보들이 효과적으로 사용자에게 전달되어야 한다.

이에 따라 본 논문에서는 디지털 대시보드에 탑재되기 적합한 형태의 CBFS 기반 특징 선택 분석기를 설계하여 향후 원자력 발전소에서 안전 문화와 관련된 다양한 데이터들을 효과적으로 관리하고자 한다.

## 3. CBFS Analyzer

Figure 1은 본 논문에서 제안하는 CBFS 분석기의 전체적인 분석 절차를 보여준다. 본 논문에서는 기존 시스템에 범주형 변수들로 이루어진 데이터가 존재한다는 것을 가정하였고, 이 데이터는 원자력 발전소 운영 전반에 걸친 다양한 변수들을 지속적으로 기록한 시계열로 볼 수 있다. 아울러, Figure 1의 절차를 통해 이러한 데이터의 특정 변수에 영향력을 미치는 다른 변수들의 집합을 선별할 수 있다.

최초에 사용자는 시스템에서 관리하는 범주형 변수들의 목록에서 반응변수로 사용할 것을 1개 지정한 다음, 이에 영향을 미치는지 관찰하고 싶은 독립변수들을 여러 개 지정한다. 이후에는 분석기간과 분석방법을 설정하여 1회의 특징 선택을 의미하는 분석 프로젝트를 시작할 수 있다. 분석 프로젝트는 사용자가 설정한 분석기간 동안의 데이터를 분석하여 지정된 반응변수에 영향을 미치는 독립변수들을 선별하는 것을 목적으로 하며, 분석방법은 독립변수 집합을 탐색 전략을 의미하며, depth-first, width-first, best-first 등 다양한 전략을 선택할 수 있다.

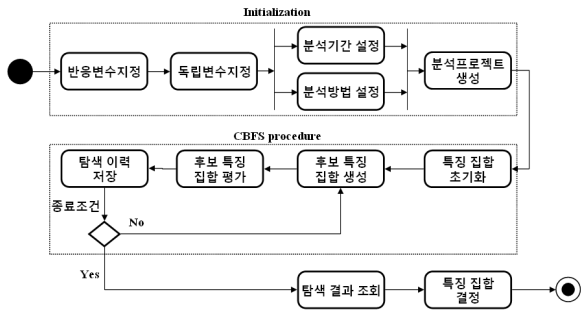


Figure 1. The overall procedure of CBFS analyzer

분석 프로젝트가 개시된 후에는 CBFS 분석 알고리즘이 작동한다. 처음에는 아무 독립변수도 포함하지 않는 초기 특징 집합에서 독립변수들을 추가해가며 후보 집합을 생성하여 평가한 후, 더 좋은 특징 집합을 찾아나가는 작업을 반복한다. 또, 이러한 탐색 과정 도중의 각 반복(iteration)에서 얻어진 독립변수 집합들에 대한 이력을 지속적으로 저장하여, 향후 사용자가 관찰할 수 있도록 한다. 알고리즘이 가장 좋은 독립변수 집합을 찾은 이후에는 사용자가 분석 결과 및 이력을 조회할 수 있다.

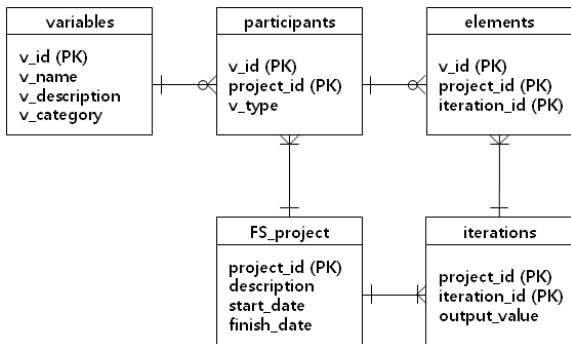


Figure 2. The data structure of CBFS analyzer

Figure 2는 앞에서 설명한 CBFS 분석기가 사용하는 데이터들의 구조를 나타낸다. Variables 테이블은 기존 시스템에서 관리되고 있는 변수들의 목록을 저장하며, 사용자가 생성한 분석 프로젝트는 FS\_project 테이블에 저장된다. 프로젝트 생성 시에는 사용자가 지정한 종속변수 및 독립변수들이 participants 테이블에 기록되고, 탐색 과정에서 1회 반복을 할 때마다 iterations 및 elements 테이블에 각 반복에서 얻은 독립 변수 집합의 성능 및 이 집합에 소속된 독립 변수들의 목록을 기록하여 분석이 완료된 이후, 사용자가 해당 내용을 참조할 수 있도록 한다.

## 4. Conclusion

실제 원자력 발전소 관련 분야에 종사하는 관리자들의 경우, 특정 변수가 악화되었을 때 그 원인을 파악하고자 하는 경우가 많을 것이다. 특히, 이러한 특정 변수가 기술적인 항목이 아니라 안전 문화와 관련된 항목인 경우에는 그 원인을 쉽게 알아내기 어려울 수 있다. 또한, 과거에 특정 변수에 영향을 주는 다른 변수들을 파악했다더라도 이러한 관계는 시간이 지나면서 변할 수 있으며, 과거의 상관관계가 현재도 유효한지를 확인해야 하는 경우도 있을 수 있다. 본 논문에서 제안하는 CBFS 분석기는 이러한 경우에 해당 변수에 영향을 미치는 다른 변수들을 편리하게 선별하고, 이를 바탕으로 기타 다양한 교사 학습을 실시하여 해당 변수를 효과적으로 관리할 수 있도록 해 준다.

뿐만 아니라, 분석 도중의 탐색 이력을 사용자에게 제공함으로써, 사용자는 독립변수 집합의 변화 과정을 쉽게 관찰할 수 있다. 따라서 CBFS 분석기가 생성한 독립변수 집합이 실제로 활용하기 어려운 경우에는 과거 이력을 보면서 다른 대안을 사용자가 직접 찾아 사용하는 것도 가능하다.

본 논문의 저자들은 향후 실제 제안하는 분석기의 프로토타입을 구현하고, 실제 원자력 발전소 운영과 관련된 데이터들에 적용함으로써 보다 실증적으로 이러한 유용함을 입증해나갈 계획이다.

## Acknowledgements

This work was funded by grants from Korean Ministry of Trade, Industry & Energy (Development of a public monitoring system and competence enhancement technologies for strengthen Korean safety culture in nuclear organizations, Grant # 2013T100200058).

## References

- Carvalho, P.V.R., dos Santos, I.L. and Vidal, M.C.R., "Safety Implications of Cultural and Cognitive Issues in Nuclear Power Plant Operation", *Applied Ergonomics*, 37(2), 211-223, 2006.
- Lee, T. and Harrison, K., "Assessing Safety Culture in Nuclear Power

- Stations”, *Safety Science*, 34(1), 61-97, 2000.
- Stephane, L., “Advanced Interaction Media in Nuclear Power Plant Control Rooms”, *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41, 4537-4544, 2012.
- Tan, P.N., Steinbach, M. and Kumar, V., *Introduction to Data Mining*, Addison-Wesley, 2005.
- Hall, M.A. and Smith, L.A., “Feature Subset Selection: A Correlation based Filter Approach”, *Proceedings of the International Conference on Neural Information Processing and Intelligent Information Systems* (pp.855-858), Berlin, 1997.
- Hall, M.A., *Correlation-based Feature Selection for Machine Learning*, Doctoral Dissertation, The University of Waikato, 1999.
- Shin, S., Guo, Y., Choi, Y. and Kim, C., “Development of a Robust Data Mining Method using CBFS and RSM”, *Lecture Notes in Computer Science*, 4378, 377-388, 2007.
- Gharehchopogh, F.S., “Approach and Review of User Oriented Interactive Data Mining”, *Proceedings of the 4th International Conference on Application of Information and Communication Technologies* (pp.1-4), 2010.