

소프트웨어 품질 진단을 위한 고장예측모델

정혜정 (평택대학교)¹⁾

국문 요약

최근 AI 기능을 가진 소프트웨어가 많이 개발되어지면서 다양한 예측 기능을 가진 소프트웨어 제품이 늘어나고 있으며, 그런 영향으로 소프트웨어 품질의 중요성이 높아지게 되었다. 특히 AI 기능을 가진 제품의 기능 안전에 대한 고려가 높아지면서 국가적 차원에서 소프트웨어 품질에 대한 관리를 진행하고 있다. 특히 국가적 차원에서 진행하고 있는 소프트웨어 제품의 품질 인증제도로 GS 품질인증제도를 들 수 있는데, GS 인증제도에서도 AI 제품에 대한 품질 평가 방안을 연구하고 있다. 본 연구에서는 소프트웨어 신뢰성 검증을 위해서 제시된 다양한 품질 평가 모델에서 국제표준에 근거해서 소프트웨어 품질의 기본 조건이 충족되는 평가모델을 제시하려 한다. 인공지능 부분의 소프트웨어 품질 특성을 고려하여 품질 평가 모델에 대한 연구를 하고, 품질을 진단하고 고장을 예측하려 한다. 본 연구에서는 소프트웨어 신뢰성 성장모델을 기반으로 하여 인공지능 국제 표준 모델을 제시하여 평가 모델을 제시하고 모델을 통해 품질 진단을 할 수 있는 방안을 제시하려 한다. 이러한 측면에서 본 연구는 앞으로 소프트웨어에서 발생하게 되어지는 고장시간에 대한 예측을 통해서 고장을 미리 예측하고 사전에 고장을 찾아서 위험을 예방할 수 있는 차원에서 중요한 의미가 있다고 판단된다. 특히 안전과 관련된 다양한 소프트웨어에서는 고장을 예측한다는 것은 중요한 의미가 있을 것으로 판단된다.

■ 중심어: 인공지능, 품질평가 메트릭, 신뢰성, 평가모델, 진단 및 예측

I. 서론

최근 소프트웨어는 대부분 많은 기능이 인공지능(Artificial Intelligence:AI) 부분을 탑재한 형태로 운영되어지고 있어, 일반적인 소프트웨어 품질 평가 모델로 제품의 품질을 평가하게 되어지면 정확한 인공지능 기능을 평가하기 어려운 부분이 있어 이 부분의 문제점 해소를 위한 연구가 다양하게 진행되어지고 있다. 국내에서는 소프트웨어 제품의 품질 향상을 위해서 ISO/IEC 25023이라는 국제 표준 모델에 준해서 품질 평가를 실시하고 있으나, 현재 사용되는 ISO/IEC 25023으로 소프트웨어 제품을 평가하게 되어지면 인공지능 부분의 기능에 대한 평가에 한계가 있다(ISO/IEC 25000, 2005), (ISO/IEC 25010, 2011), (ISO/IEC 25023, 2015))

이러한 문제점을 해결하기 위해서는 정확한 소프트웨어 제품에 대한 신뢰도 평가도 진행되어야 하며 소프트웨어 고장 관련 데이터를 수집하고, 수집된 고장 관련 시간 데이터를 소프트웨어 신뢰성 성장 모델에 적용하여 신뢰도를 평가하여야 한다. 소프트웨어 고장 관련 데이터를 소프트웨어 신뢰성 성장 모델에 적용하면 소프

** 본 연구는 평택대학교 연구 지원비에 의해 이루어졌음

1)제1저자: 평택대학교 데이터정보학과 교수, jhjung@ptu.ac.kr

· 투고일: 2024-05-22 · 수정일: 2024-06-19 · 게재확정일: 2024-06-20

트웨어 신뢰성을 예측하게 되어진다. 이러한 소프트웨어 신뢰성에 대한 연구 결과를 활용해서 앞으로 발생하게 되어지는 고장을 예측할 수 있게 된다(정혜정, 2019). 본 연구는 소프트웨어 신뢰성 성장 모델을 적용하여 소프트웨어의 품질 평가를 진행하고 인공지능 기능을 평가하기 위해서 추가적으로 평가해야 하는 평가 방안을 제시한다.(정혜정, 2020).

고장 관련 데이터는 현재 국제표준을 기반으로 소프트웨어 시험을 하고 있는 소프트웨어 시험 기관에 의뢰해서 소프트웨어 시험 과정에서 발생한 고장 관련 고장시간에 대한 자료를 수집하고 수집된 자료를 분석하려 한다(Goel,A.L. & Okumoto,K, 1979).

분석된 소프트웨어 고장 관련 고장 시간 자료는 소프트웨어 신뢰성 성장 모델에 적용하여 신뢰성을 평가할 수 있는 모수를 측정하고 활용할 수 있는 방안에 대한 연구를 진행하려 한다. 본 연구는 고장 관련 자료 수집에서 부터 수집된 자료에 대한 분석과 분석된 자료를 소프트웨어 신뢰성 성장 모델에 적용해서 신뢰도에 대한 베이지추정치를 구하고, 구한 결과를 국제표준에서 제시하고 있는 기능정확성을 평가하기 위한 메트릭에 적용하는 방안에 대해서 제시한다(정혜정, 2014).

정확한 소프트웨어의 신뢰성 평가를 위해서는 고장 관련 자료수집과 수집된 자료를 다양한 경우의 소프트웨어 신뢰성 성장 모델에 적용하고 적용한 모델의 모수 추정 방법을 통해서 모수를 추정하고 추정된 모수를 이용해서 추국제표준에서 제시하는 기능정확성 등을 평가할 수 있어야 한다. 본 연구에서는 소프트웨어 신뢰성 성장 모형으로 베타사전분포를 이용해서 베이지추정치를 구하고 추정치를 이용해서 품질을 평가하는 방법을 제시한다(정혜정, 2003). 본 연구는 ISO/IEC 25023에서 제시하고 있는 국제 표준을 기반으로 하여 소프트웨어 품질을 평가할 수 있도록 8가지 품질특성을 정리하고, 제시된 품질평가 메트릭을 통해서 품질을 평가하게 되어진다. 8가지 품질 특성으로는 소프트웨어 제품의 기능을 중심으로 평가하는 기능적합성(Functional suitability), 제품의 시간과 자원의 효율을 측정하는 효율성(Performance efficiency), 제품을 사용하는 사용자가 얼마나 쉽게 제품을 사용할 수 있는지를 평가하는 사용성(Usability), 제품이 고장 없이 가동할 수 있는지를 평가하는 신뢰성(Reliability), 제품의 설치와 제거를 평가하는 이식성(Portability), 소프트웨어 제품이 보안 관련 내용이 있을 때 평가하는 보안성(Security), 데이터 등의 호환을 확인하기 위한 호환성(Compatibility), 제품의 수정을 통해서 제품이 유지될 수 있는지를 평가하는 유지보수성(Maintainability)이 있으며, 이러한 8가지 품질 특성 중 신뢰성에 대한 평가가 가장 어려운 실정으므로 본 연구에서 신뢰성에 대한 모델을 제시하고 데이터를 적용하는 방안과 메트릭에 적용하는 방안을 소개하려 한다. 2장에서는 현재 AI 소프트웨어 부분에 소프트웨어 품질 평가에 대한 연구 내용을 소개하고, 3장에서는 이 부분을 이용해서 소프트웨어 신뢰성 성장 모델과 메트릭을 적용하는 방안을 제시하고, 4장에서는 적용된 결과에 대한 분석을 통해서 품질을 평가하는 방안과 향후 연구과제에 대해서 제시하려 한다(정혜정, 2019b).

II. AI 소프트웨어 품질 평가 관련 연구

2.1 AI 소프트웨어 품질 평가 모델

현재 국내에서는 소프트웨어 관련하여 GS(Good Software) 소프트웨어 품질인증제도를 운영하여 제품에 대한 평가를 실시하고 있다. 국내에서 적용하고 있는 소프트웨어 품질 평가는 국제표준문서 ISO/IEC 25023을 이용하고 있으며, ISO/IEC 25023에서는 기능적합성(Functional Suitability), 성능효율성(Performance Efficiency), 호환성(Compatibility), 사용성(Usability), 신뢰성(Reliability), 보안성(Security), 유지보수성(Maintainability), 이식성(Portability)과 같은 8가지 품질특성을 기반으로 소프트웨어 품질 평가를 제시하고 있으며, 각각의 품질 특성은 품질 부특성과 품질 평가 메트릭을 구성하고 있다(정혜정, 2018)

그러나 최근 들어 많은 소프트웨어에 AI 기능이 탑재되면서 소프트웨어 제품 평가의 기준으로 삼고 있던 국

제표준 ISO/IEC 25023으로는 인공지능 기능에 대해서 소프트웨어가 정확히 평가 되지 않는 부분이 있어, 이 부분에 대한 평가를 위한 연구가 진행되고 있다(KS X ISO/IEC TR 24028, 2020). 현재 국제표준에서 제시하고 있는 인공지능시스템에 대한 소프트웨어 품질 평가에서 연구되어지고 있는 내용으로는, 첫째 AI 기능이 탑재된 소프트웨어는 기능적합성을 평가할 때 기능적응성에 대한 내용을 추가하여 평가해야 한다는 주장을 하고 있다. 둘째, 사용성에 대한 평가는 사용자 제어성과 투명성에 대한 평가를 추가해야 한다는 결과가 발표 되었다. 셋째, 신뢰성에 대한 평가에서는 기존의 평가 결과에 강건성에 대한 평가를 추가해야 한다고 발표하였으며, 넷째, 보안성에 대한 평가는 개입성에 대한 평가를 추가해서 평가해야 한다고 제시하고 있다. AI 기능이 있는 소프트웨어에서 추가적으로 평가해야 한다고 제시된 내용에서 기능적합성의 기능적응성은, 인공지능 시스템이 스스로 적응하는 능력을 의미하는 것으로 학습데이터에 대해서 시스템이 이전에 취한 행동의 결과를 적응하는 것이라 해석 할 수 있다. 그리고 사용성의 사용자 제어성이라 함은 인공지능 시스템의 경우 인간이 직접 인공지능 시스템의 기능 작동에 개입할 수 있는 능력을 의미하는 것으로 인공지능 시스템이 부정적인 결과를 내거나 예기치 못한 결과가 나올 경우에 사용자 제어성의 기능은 필요하다고 제시되어 있어 AI 제품의 경우 사용자 제어성의 평가가 필요하다고 제시하고 있다. 다음은 사용성의 투명성의 경우는 사용성에 대한 품질 평가 모델에서 인공지능 시스템에 대한 정보가 이해관계자에게 전달되는 정도를 의미하는 것으로 사용자 자신의 요구에 맞는 시스템을 선택하는데 도움을 줄 수 있는 능력을 가지고 있는지를 평가하게 되어진다(양해술, 2013)

국제표준에서 제시하고 있는 신뢰성의 강건성은 데이터 편향에 의한 문제로 발생하는 것에 대한 문제점과 외부 간섭에 대한 문제점 등에 대한 것을 평가하는 것으로, 개발자에 의해서 의도된 올바른 기능 작동과 관련되는 속성들을 평가하도록 제시되어 있다. 여기에서 강건성의 유형에는 데이터로 인해서 발생하는 다양한 문제를 고려할 수 있다. 보안성의 개입성이라 함은 인공지능 시스템이 기능 작동 상태를 관찰하고 안전하지 않은 상태를 안전한 상태로 전환하는 능력을 가지고 있는지를 의미하는 것으로 인공지능 시스템에 있어서 시스템에 문제가 발생하거나 안전상에 큰 리스크를 초래하는 것을 방지하는 목적으로 작동하고 있는지를 평가해야 한다고 제시하고 있다.

인공지능 시스템에 대한 평가는 국제표준 ISO/IEC 25059에서 간접적으로 제시하고 있다. 또한 국제표준 ISO/IEC 25023에서는 일반 소프트웨어를 평가하도록 제시되어 있으나, 현재 인공지능 시스템의 기능에 대한 평가를 소프트웨어적인 측면에서 평가할 수 있는 방안을 검토하고 있어 추가 모델에 대한 제안을 하고 있다.

<표1> 인공지능 품질 평가 국제 표준 모델 적용 가이드

번호	표준명	내용
1	KS X ISO/IEC 25023	소프트웨어 품질평가 모델
2	ISO/IEC 25024	System and software Quality Requirement and Evaluation(SQuaRE)-measurement of data quality
3	ISO/IEC 25059	System and software Quality Requirement and Evaluation(SQuaRE)-Quality model for AI system
4	ISO/IEC 24028	Information technology-Artificial intelligence-overview of trustworthiness in artificial intelligence

<표1>에 제시된 것과 같이 소프트웨어 품질 평가로 제시된 ISO/IEC 25023에서는 일반 소프트웨어 품질 평가에 대한 내용을 수록하고 있으며, ISO/IEC 25024에서는 데이터에 대한 품질 평가를 제시하고 있다. 위의 두 개 표준 모델을 기반으로 ISO/IEC 25059에서는 AI 시스템에서 강조하고 있는 새로운 평가 모델을 추가하여 제시하고 있으며, <표1>에서 제시하고 있는 국제표준 모델을 기반으로 국내 GS 인증을 관리하고 있다고 할 수 있으며 본 연구에서는 GS 인증을 위해서 AI 제품 기능을 가지고 있는 제품에 대한 평가를 위해서 ISO/IEC

25059 적용 방안에 대해서 제시한것이다.

<표1>의 ISO/IEC 24028에서 제시하고 있는 인공지능 신뢰성 개요를 참조하면 인공지능 시스템은 신뢰성과 관련하여 투명성, 설명가능성, 제어성을 통해 인공지능 시스템을 구축해야 함을 강조하고 있다. 즉, 인공지능 시스템에 대한 기술적 함정 및 일반적인 위협과 가용성, 회복력, 정확도 등에 대해서 정의하고 있으며 적용방안에 대해서 제시하고 있다(권원일, 2010).

소프트웨어에 인공지능 기능이 탑재되면 일반적으로 신뢰성에 영향을 미치게 되며, 데이터의 영향에 따라서 인공지능 시스템은 도덕적이고 윤리적인 원칙을 준수해야 하며, 인공지능 시스템에 대한 정보가 악의적으로 조작되지 않도록 보호되어야 한다. 앞에서 설명했던 것처럼 이런 의미에서 ISO/IEC 25010 소프트웨어 품질 평가 모델에서 제시하고 있는 8개 품질 특성의 부특성에 인공지능 시스템의 특성을 고려하여 소프트웨어 품질을 평가할 수 있는 모델을 개발해야 한다. 또한 AI 시스템의 경우 데이터에 의존적이므로 데이터 품질을 평가할 수 있도록 제시된 국제표준 ISO/IEC 25012를 통해서 통합적인 평가 방안을 제시해야 한다.

ISO/IEC 25059에서는 인공지능에 투명성에 대한 접근방식, 새로운 보안에 대한 관리, 개인정보보호에 대한 내용, 강건성 및 회복 탄력성과 관련된 내용, 기계학습 알고리즘의 선택과 설정에 대한 내용, 데이터와 관련된 고려 사항 등에 대해서 제시하고 있으므로 AI 관련 소프트웨어 품질 평가를 위해서 적용방안에 대한 검토가 필요하다.

국내에서도 2022년 한국정보통신기술협회에서는 신뢰할 수 있는 인공지능개발안내서를 발간하고 인공지능 평가에 대한 가이드를 제시하였다.

한국정보통신기술협회에서 제시한 신뢰할 수 있는 인공지능개발안내서에서는 인공지능 신뢰성에 대한 문제를 데이터의 편향성 문제로 소개하고 있으며 인공지능 시스템에 내재되어 있는 위협이나 문제점을 해결하고 인공지능을 활용하고 확산하는 과정에서 부작용을 준수해야 하는 기준을 제시하고 있다. 본 안내서에서는 인공지능 신뢰성을 확보하기 위해서는 안전성(safety), 설명가능성(explainability), 투명성(transparency), 견고성(robustness), 공평성(fairness)이 확보되어야 한다고 제시하고 있다(ISO/IEC 25059, 2023).

특히 신뢰할 수 있는 인공지능 개발 가이드에서는 데이터 수집 및 처리에 있어 데이터 정보에 대한 내용과 데이터 수집에 대한 내용과 데이터 라벨링에 대한 내용과 데이터 편향에 대한 내용으로 나누어 품질 평가의 기준을 제시하고 있으며, 인공지능 시스템에 영향을 미치고 있는 데이터에 대한 내용을 품질 평가에서 평가하도록 제시하고 있다(한국정보통신기술협회, 2022).

본 연구에서는 최근에 연구 실적으로 제시된 내용들을 중심으로 국제표준에서 제시하고 있는 평가 메트릭을 기반으로 하여 인공지능 소프트웨어 품질 평가에 대한 평가를 위해서 평가 방안을 제시한다.

2.2 추가된 인공지능 소프트웨어 평가모델

현재 연구된 내용에서 데이터에 대한 내용을 살펴보면 인공지능 시스템에 있어 인공지능시스템의 정확성을 평가하기 위해서는 데이터에 대한 이상값을 정확하게 식별할 수 있어야 하며, 인공지능 시스템의 예측에 대한 정상과 결함을 점검할 수 있어야 한다. 평가 항목과 데이터 다양성 확보와 데이터 편향에 대해서 점검하는가에 대한 평가관리가 진행되어야 하며, 데이터 편향에 대해서 특성을 배재했는지에 대해서 평가가 되어야 한다. 인공지능시스템의 경우 데이터 처리에 대한 평가와 인공지능 모델에 대한 것을 동시에 고려하여 평가할 수 있도록 해야 한다.

인공지능시스템의 경우 제품에 대한 평가를 위해서는 데이터 확보 방안과 데이터 관리에 대한 방안이 중요하며 제대로 되고 있는가에 대한 정확한 평가가 진행되어야 한다. 특히 데이터 편향으로 인해서 인공지능시스템 적용 시 발생할 수 있는 리스크가 클 수 있으므로 정확한 평가 방안이 제시되어야 한다.

국제표준 ISO/IEC 25023에서 제시되고 있는 소프트웨어 품질평가 방안은 품질특성 8개, 품질 부특성은 31

개, 평가메트릭은 86개로 구성되어 있다. 본 연구도 이러한 국제표준에서 제시하는 내용에서 AI 소프트웨어 평가를 위한 방안 제시라고 할 수 있다.

2.3 추가된 인공지능 소프트웨어 평가모델

인공지능시스템 평가를 위한 국제표준 ISO/IEC 25023에서 제시하고 있는 소프트웨어 품질 평가를 위한 품질 특성과 부특성, 부특성에 따른 평가메트릭의 내용 중 신뢰성 관련 평가 항목은 아래 <표2>와 같다.

<표2> 신뢰성 부특성

신뢰성	성숙성	결함수정, 평균고장사이시간 등
	가용성	시스템가용성, 평균다운시간 등
	결함허용성	고장회피율, 평균결함통보 등
	복구성	평균복구시간, 백업데이터완벽성

<표2>에서 제시한 소프트웨어 신뢰성 평가는 4개의 부특성으로 구성되어 있으며 성숙성(Maturity)은 소프트웨어에서 발생하는 고장이나 결함 관련 평가에 대한 내용을 제시하고 있으며, 가용성(Availability)은 소프트웨어 사용이 요청될 때 제품이 끊기지 않고 계속적으로 동작하고 있는 상태를 평가할 수 있도록 제시하고 있다. 결함허용성(Fault Tolerance)에 대한 평가는 제품이 하드웨어나 소프트웨어 결함에도 불구하고 의도한대로 작동하고 있는지를 평가하고 있으며, 복구성은 소프트웨어가 작동 중에 중단이나 고장이 발생했을 때, 소프트웨어를 복구하고 재설정 할 수 있는 정도를 파악할 수 있는지를 평가한다. 신뢰성 평가에 대해서 제시하고 있는 시스템 가용성이나 고장률에 대한 정확한 평가항목은 <표3>과 같다.

<표3> 신뢰성 품질 메트릭

메트릭이름	측정내용	측정함수
시스템가용성	시스템이 정해진 운영시간 동안에 중단 없이 실제적으로 운영 가능한 시간의 비율을 측정함	$X = \frac{A}{B}$ A: 실제적으로 운영된 시스템 운영시간 B: 정해진 운영시간으로 명시된 전체 운영시간
고장률	제품이 운영되고 있는 시간 동안에 평균적으로 고장이 몇 번 일어났는가를 측정함	$X = \frac{A}{B}$ A: 제품이 작동되는 동안 측정된 전체 고장의 수 B: 제품의 전체 작동 시간

인공지능 시스템의 품질 평가를 위해서 테스트케이스를 작성하고 작성된 테스트케이스를 중심으로 품질 평가에 대한 특성별 평가점수를 정량적으로 측정해야 한다.

III. 베타분포를 이용한 신뢰성 성장모델

인공지능시스템에 대한 신뢰성을 측정하기 위해서 본 연구에서는 베타사전 분포를 제시하고 베이지 추정치를

구하여 신뢰성을 평가하는 모델을 제시한다. 먼저 본 모델을 이용하여 신뢰성을 구하기 위한 기본 가정은 다음과 같다.

1. 각 테스트 단계는 통계적으로 독립이어야 하며, 각 테스트 단계에서 소프트웨어가 가동할 확률은 R_i 라고 가정한다.
2. 각 테스트 단계에서 테스트 실행을 n_i 번 하고 s_i 번의 소프트웨어 가동을 하였다고 가정한다.
3. 각 단계에서 소프트웨어의 결함을 발견하고 수정하게 되면 신뢰도가 측정된다. 이때 신뢰도는 아래와 같다 $R_1 \leq R_2 \leq \dots \leq R_m$

다음은 신뢰도 함수에 대한 우도함수 $L(R_1 \dots R_m)$ 를 구하면 다음과 같다

$$L(R_1, \dots, R_m) = \prod_{i=1}^m R_i^{s_i} (1 - R_i)^{n_i - s_i}$$

본 연구에서 사용된 i 번째 테스트 단계에 대한 사전분포는 앞에서 제시한 베타분포를 사용하였으며 베타분포를 이용한 결과는 아래와 같다.

$$p(R_i | \alpha_i, \beta_i) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} R_i^{\alpha_i - 1} (1 - R_i)^{\beta_i - 1}$$

베타사전 분포를 이용해서 계산된 우도함수와 사전분포를 이용한 사후분포 $p(R_1, \dots, R_m)$ 는 아래의 식과 같다.

$$p(R_1, \dots, R_m) = R_1^{S_1} (1 - R_1)^{n_1 - s_1} \dots R_m (1 - R_m)^{n_m - s_m} \cdot \frac{\Gamma(\alpha_m + \beta_m)}{\Gamma(\alpha_m)\Gamma(\beta_m)} R_m^{\alpha_m - 1} (1 - R_m)^{\beta_m - 1}$$

단 $R_1 \leq R_2 \leq \dots \leq R_m$

사후분포를 이용하여 베이즈추정치를 구하기 위하여 계산된 비례상수를 구하면 아래의 식과 같다

$$\frac{1}{K_m} = \sum_{h_1=S_1}^{g_1} \dots \sum_{h_{m-1}=S_{m-1}+h_{m-2}}^{g_{m-1}} \prod_{j=1}^{m-1} B(S_j + h_{j-1}, S_{j-1} + f_j - h_{j-1}) \binom{g_j}{h_j} B(S_m + h_{m-1}, S_{m-1} + f_m - h_{m-1})$$

여기에서 B 는 베타분포를 의미하며 베타분포의 모수를 간단히 하기 위해서 $S_{(i)}$, $f_{(i)}$ 로 정리하였으며 h_i 는 단계별 테스트 횟수를 의미한다. 이때 $S_{(i)}$, $f_{(i)}$ 는 아래의 수식과 같다.

$$S_i = s_i + \alpha_i, f_i = n_i + \beta_i - s_i - 1$$

$$S_{(i)} = S_1 + \dots + S_i = \sum_{j=1}^i s_j + \sum_{j=1}^i \alpha_j$$

$$f_{(i)} = f_1 + \dots + f_i = \sum_{j=1}^i n_j + \sum_{j=1}^i \beta_j - \sum_{j=1}^i s_j - i$$

$$g_i = f_{(i)} + S_{(i)} + 1 = \sum_{j=1}^i n_j + \sum_{j=1}^i \alpha_j + \sum_{j=1}^i \beta_j - i + 1$$

위의 값을 이용하여 일양분포를 사전분포로 이용했을 경우와 베타분포를 사전분포로 이용했을 경우 사후확률의 정확도를 확인하기 위하여 모의실험을 하여 본 결과는 아래 <표4>와 같다.

<표4> 사전분포를 이용한 베이즈추정치

사전분포확률	2 단계	3 단계	4 단계
실제값	0.7	0.75	0.8
일양분포	0.69	0.78	0.821
베타분포	0.701	0.76	0.805

<표4>에서 살펴보면 베이지추정치는 사전분포로 일양분포를 사용하는 경우보다 베타분포를 사용하여 추정하는 경우가 더 정확하다는 것을 알 수 있다. 추정치의 정확성이 높아야 신뢰도를 정확하게 추정할 수 있으므로 이 경우는 사전분포로 일양분포 보다 베타분포가 더 정확하다는 것을 알 수 있다.

이 경우 신뢰도 함수를 구하면 다음과 같다.

$$R(n, n+h) = \exp(-(D(n+h) - D(n))), \quad (n, h = 0, 1, \dots)$$

여기에서 n 은 단위시간 당 고장의 평균값이며 h 는 단위시간을 의미한다.

위의 식에서 $D(n)$ 은 오류 검정과정에서 소프트웨어 신뢰도성장모형의 비동질적포아송과정에서의 평균값함수이다. 이때 평균값함수 $D(n)$ 는 포아송 프로세스의 기본적 성질을 만족한다. 즉 베이지추정치를 통해서 측정된 신뢰도는 <표3>에서 제시된 시스템 가용성 평가에 시스템이 계속적으로 가동될 신뢰도 측정치로 활용되어질 수 있다.

베이지추정치를 이용하여 신뢰도를 측정하고 측정된 신뢰도 측정치를 품질 특성별 메트릭에 적용한다.

다음은 AI 소프트웨어의 경우에 기능적합성의 부특성인 기능정확성에 인공지능 모델의 정확성을 평가하여 평가 항목으로 추가해서 관리하게 되면 인공지능 소프트웨어의 정확성에 대한 품질을 측정하게 될 것이다.

인공지능 모델에서 기능의 정확성은 다음과 같이 측정한다

$$\text{정확성} = \frac{\text{올바르게 예측한 샘플수}}{\text{전체 샘플수}}$$

즉, 정확성에 대한 평가는 일반적으로 AI 소프트웨어의 분류 기능에 있어서 정확하게 분류한 수를 측정하여 평가한다. 즉 아래와 같이 제시할 수 있다.

$$\text{정확성} = \frac{TP + TN}{TP + TN + FP + FN}$$

여기에서

TP: True Positive TN: True Negative

FP: False Positive FN: False Negative

본 연구에서는 현재 AI 소프트웨어 품질 측정을 위해서 기능정확성의 추가항목에 대한 제시와 신뢰성의 가용성에 신뢰도를 추가하여 측정하는 방안을 제시하였다. 신뢰도성장모델을 이용하여 국제표준에 적용해서 신뢰도를 측정할 수 있는 방안을 제시했으며 사전분포의 유형에 따라서 사후분포를 구해서 사후분포의 확률을 신뢰도로 측정하는 방안을 제시했다. 또한 일반 소프트웨어의 기능정확성 평가 항목에다가 인공지능 소프트웨어에서 정확성을 측정하기 위한 측정함수를 적용하여 평가하는 방안에 대해서 제시하였다. 위와 같이 정확성을 추가 측정하여 제품에 대한 품질평가를 진행하면 정확한 품질평가가 가능해 질것으로 판단된다.

IV. 결론 및 인공지능제품 추가 평가메트릭

기존의 소프트웨어 품질 평가 모델에서 평가할 수 없는 AI 제품에 대해서 품질을 평가하기 위한 방안을 제시하기 위하여 본 연구에서 2장에서 현재 국내 인공지능 제품의 품질평가 메트릭에 대한 연구를 제시하였으며, 3장에서는 사전분포를 주고 베이지 추정치를 추정하여 신뢰도를 평가하는 방안에 대해서 제시하였으며 기능적합성에 기능정확성을 추가 측정하여 품질을 평가 할 수 있는 방안을 제시하였다. 본 연구에서 인공지능제품에 대한 품질평가를 위해서는 무엇보다도 데이터에 대한 관리가 정확히 이루어져야 하므로 ISO/IEC 25012에서 제시하는 데이터 품질 관리 방안도 추가적으로 적용하여야 할 것으로 판단된다. 한 예로 데이터 관리에 대한 내용에서 예시를 보이면 아래의 <표5>와 같다.

<표5> 데이터 품질 매트릭

번호	테스팅이름	적용항목	방법
1	이상치	이상치를 제거하는 방안을 검토하였는가?	z-score
2	데이터 에러	발생가능 한 에러에 대해서 시각화하고 체크 할 수 있는가?	히스토그램 커널분포
3	편향	발생할 수 있는 데이터 편향에 대해서 관리할 수 있는가?	필터 래퍼 임베디드

인공지능 시스템의 경우 데이터에 의존도가 높기 때문에 데이터에 대한 관리도 동시에 이루어져야 할 것으로 보인다. 현재 인공지능시스템에 대한 평가 방안을 계속적으로 연구하고 있으므로 본 연구에서 제시한 신뢰도 측정방안과 정확도와 데이터 평가 방안을 고려하여 인공지능 시스템 평가가 진행되어야 할 것으로 판단된다. 앞으로 본 연구를 바탕으로 AI 소프트웨어에 대한 평가 항목을 좀 더 세분화해서 평가 할 수 있는 방안 연구를 진행할 것이다. 또한 현재 GS 인증기관에서 사용하고 있는 ISO/IEC 25023에 인공지능 부분이 추가될 수 있도록 테스트 데이터를 수집하여 다양한 사전분포를 적용하고 사후분포를 예측할 수 있도록 연구할 계획이다.

REFERENCE

- 권원일(2010). *개발자를 위한 소프트웨어 테스트*. STA.
- 양해술 (2013). 오픈소스 소프트웨어의 기준 구축의 품질 평가. *The Journal of Digital Policy & Management*, 11(2), 45-62.
- 정혜정(2003). 오류 데이터의 그래프를 이용한 소프트웨어 신뢰성 성장 모델의 효과성 측정, *Korea Information Processing Society*, 3(1), 41-48.
- 정혜정(2014). 소프트웨어 효과 분석, *The Journal of Digital Policy & Management*, 12(1), 56-62.
- 정혜정(2018). 소프트웨어 품질 평가 매트릭을 적용한 신뢰성 측정, *The Journal of Multimedia*, 15(1),55-61.
- 정혜정(2019a). 소프트웨어 오류 데이터를 기반으로 한 소프트웨어 신뢰성 성장 모형. *Journal of the Korea Convergence Society*, 10(3), 59-66.
- 정혜정(2019b). 소프트웨어 신뢰성 성장 모델을 기반으로 한 소프트웨어 품질 측정. *Journal of the Korea Convergence Society*, 10(4), 45-50.
- 정혜정(2020). 소프트웨어 시험보고서의 텍스트 분석, *Journal of the Korea Convergence Society*, 11(3), 59-66.
- 한국정보통신기술협회(2022), *신뢰할 수 있는 인공지능 개발안내서*
KS X ISO/IEC TR 24028(2020), *정보기술-인공지능-인공지능의 신뢰성 개요*
- Goel, A. L. & Okumoto, K. (1979). Time dependent error-detection rate model for software reliability and other performance measures, *IEEE Trans. Reliability*, R-28, 206-211.
- ISO/IEC 25000.(2005), *System and software engineering: System and software Quality Requirements and Evaluation(SQuaRE) -Guide to SQuaRE*.
- ISO/IEC 25010. (2011), *System and software engineering-System and software Quality Requirements and Evaluation(SQuaRE) -System and software quality model*.
- ISO/IEC 25023. (2015), *System and software engineering-System and software Quality Requirements and Evaluation(SQuaRE) - Measurement of system and software product quality*.
- ISO/IEC 25059 (2023), *Software engineering-System and software engineering: System and software Quality Requirements and Evaluation(SQuaRE) -Quality model for AI system*

Failure Prediction Model for Software Quality Diagnosis

Jung Hye-jung¹⁾

Abstract

Recently, as a lot of software with AI functions has been developed, the number of software products with various prediction functions is increasing, and as a result, the importance of software quality has increased. In particular, as consideration for functional safety of products with AI functions increases, software quality management is being conducted at a national level. In particular, the GS Quality Certification System is a quality certification system for software products that is being implemented at the national level, and the GS Certification System is also researching quality evaluation methods for AI products. In this study, we attempt to present an evaluation model that satisfies the basic conditions of software quality based on international standards among the various quality evaluation models presented to verify software reliability. Considering the software quality characteristics of the artificial intelligence sector, we study quality evaluation models, diagnose quality, and predict failures. In this study, we propose an international standard model for artificial intelligence based on the software reliability growth model, present an evaluation model, and present a method for quality diagnosis through the model. In this respect, this study is considered to be important in that it can predict failures in advance and find failures in advance to prevent risks by predicting the failure time that will occur in software in the future. In particular, it is believed that predicting failures will be important in various safety-related software.

Keyword: Artificial intelligence, quality evaluation metrics, reliability, evaluation model, diagnosis and prediction

1)Author, Full Professor of PyeongTaek University, jhjung@ptu.ac.kr

제1 저자 소개

■ 정혜정(Jung, Hye-jung)

■ 평택대학교 데이터정보학과 정교수, 통계학박사

ISO/IEC JTC1/SC7 위원, ISO/IEC JTC1/SC34 위원, 한국정보통신기술협회 소프트웨어 인증위원, 한국산업기술시험원 소프트웨어인증위원, 한국기계전자시험연구원 소프트웨어 인증위원장, 한국융합시험연구원 소프트웨어인증위원, 동의대학교산학협력단 소프트웨어인증위원, 한국표준협회 AI+ 소프트웨어 인증위원장, 한국정보통신기술협회 클라우드 상호운영성 인증위원, 과학기술정보통신부 데이터 품질자문위원장, 한국정보통신기술협회 데이터인증위원장 등

<관심분야>: 빅데이터분석, 소프트웨어 품질평가, 인공지능 소프트웨어 품질평가, 소프트웨어 품질 인증, 데이터 품질평가, 데이터 품질인증 등