Journal of the Korea Industrial Information Systems Research Vol.25 No.6, Dec. 2020 :15-23
http://dx.doi.org/10.9723/jksiis.2020.25.6.015 ISSN:1229-3741

농작물 생육환경정보와 생체정보 분석을 위한 빅데이터 모델+

(Big Data Model for Analyzing Plant Growth Environment Informations and Biometric Informations)

이 종 열[1], 문 창 배[2], 김 병 만[3]*
(JongYeol Lee, ChangBae Moon, and ByeongMan Kim)

요 약 기후 변화에 대응하기 위한 농업분야의 연구활동이 활발하게 이루어지고 있는 가운데 4차 산업혁명에 맞춰 정보통신기술을 활용한 스마트농업이 새로운 트랜드가 되었다. 이에 따라 다양한 노지 환경과 토양 조건에서 농작물의 스트레스를 모니터링하여 생육 이상 징후를 미리 식별하고 대응하려는 연구가 진행되고 있다. 다양한 센서를 거쳐 실시간으로 수집되는 데이터들을 인공지능 기법이나 빅데이터 기술을 활용하여 분석하려는 시도도 있다. 본 논문은 빅데이터 분석을 위해 기존 관계형 데이터베이스를 이용하여 농작물의 생육환경정보와 생체정보 분석에 효과적인 빅데이터 모델을 제안한다. 모델의 성능은 데이터 양에 따른 쿼리에 대한 응답 시간으로 측정하였다. 그 결과 최대 23.8%의 시간 단축 효과가 있음을 확인할 수 있었다.

핵심주제어: 빅데이터모델, 생육환경정보, 생체정보, 스마트농업, 생육지표

Abstract While research activities in the agricultural field for climate change are being actively carried out, smart agriculture using information and communication technology has become a new trend in line with the Fourth Industrial Revolution. Accordingly, research is being conducted to identify and respond to signs of abnormal growth in advance by monitoring the stress of crops in various outdoor environments and soil conditions. There are also attempts to analyze data collected in real time through various sensors using artificial intelligence techniques or big data technologies. In this paper, we propose a big data model that is effective in analyzing the growth environment informations and biometric information of crops by using the existing relational database for big data analysis. The performance of the model was measured by the response time to a query according to the amount of data. As a result, it was confirmed that there is a maximum time reduction effect of 23.8%.

Keywords: Big data model, Growth environment informations, Biometric informations, Smart-agriculture, Plants growth index

* Corresponding Author: bmkim@kumoh.ac.kr
+ 본 논문은 농촌진흥청 연구사업(세부과제번호: PJ015050042020)의 지원에 의해 이루어진 것임. 이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2018R1C1B6001042)

Manuscript received September 14, 2020 / revised October 14, 2020 / accepted November 08, 2020
1) 금오공과대학교 소프트웨어공학과, 제1저자
2) 금오공과대학교 ICT융합특성화연구센터, 제2저자
3) 금오공과대학교 소프트웨어공학과, 교신저자

1. 서 론

최근 농업 분야에서는 농작물의 생산부터 소비까지 자동화하는 스마트농업 실현을 위한 다양한 연구가 진행되고 있다. 스마트농업 (Smart Agriculture)은 기존 농업에 정보통신기술 (Information and Communications Technologies; ICT)이 융합된 최첨단 농업 생산방식을 통칭한다. 농업 생산 시스템이 과거의 '경험 기반'에서 점차 '빅데이터 기반 예측 관리'로 전환되어 농부의 직간접적인 경험과 노하우에서 벗어나 데이터의 정확한 분석과 예측을 통한 맞춤 처방을 목적으로 하고 있다. 스마트농업은 정밀농업과 스마트팜 (Smart Farm)으로 대표되는 실내농업, 스마트축사, 스마트양식 등의 분야에서 폭넓게 적용되고 있다 (Lee and Seol, 2019; Choi and Jang, 2019; Korea Agro-Fisheries and Food Trade Corporation. 2019; Ministry of Science and ICT, 2019).

정보시스템을 활용한 스마트농업의 핵심 기술은 데이터, 시설, 장비로 요약된다. 데이터의 경우, 시설이나 노지에서 농작물의 생육환경을 감시하거나 제어하기 위한 시설로부터 생성된다. 주로 센서에서 측정한 아날로그 신호가 디지털 신호로 변환되어 데이터화 되는데, 측정 데이터는 측정 주기와 형식에 따라 생성되는 양에서 큰 차이가 있지만 오랜 시간 축적하면 데이터를 정밀하게 선별하여 활용 가능한 빅데이터 형태를 띠게 된다. 빅데이터는 구조가 복잡하고 데이터양이 많아서 기존의 데이터베이스 관리 도구로 처리하기 어려운 데이터 집합이라고 말해 왔다. ICT 기술의 발전으로 많은 양의 데이터를 처리하는데 시간과 비용이 현저하게 줄어들었지만 새롭게 생성되는 데이터의 증가 속도는 더욱 빨라졌다. 따라서 빅데이터 기술은 데이터의 양이 증가한 것에 대응하는 기술만을 의미하는 것이 아니라 데이터의 형식이나 저장 방식 그리고 입출력 성능 등이 함께 고려된 기술로 이해해야 한다.

빅데이터에 대한 정의는 사람, 관점, 분야에 따라 다양하게 내려지고 있고 빅데이터의 속성도 계속 발견되고 있다. 가장 일반적인 속성으로는 데이터의 크기를 나타내는 규모 (Volume), 다양한 종류의 데이터를 수용할 수 있는 다양성 (Variety), 데이터를 빠르게 처리하고 분석할 수 있는 속도 (Velocity) 등이 있다. 추가적인 속성으로는 소수의 데이터에 문제가 있어도 결과에 영향을 미치지 못한다는 정확성 (Veracity), 빅데이터로 분석 한 결과를 통해 새로운 가치가 도출되는지 여부를 나타내는 가치 (Value) 등이 있다 (Park et al. 2019). 최근에는 어떤 결정을 내리는데 타당한 데이터인지를 판단하는 정확성 (Validity)과 데이터가 얼마나 오래 저장되고 언제까지 의미 있는 데이터로 사용되는지 여부를 나타내는 휘발성 (Volatility) 등의 속성들이 추가로 고려되고 있다(Tomcy and Pankai, 2017).

기존의 빅데이터 관련 연구는 데이터 수집방법과 분석방법 그리고 활용방법에 대해서 이루어지고 있다. 데이터 수집방법은 텍스트 기반 콘텐츠를 입수하여 변환하거나 분석 대상이 있는 웹과 모바일 환경에 게시된 콘텐츠를 자동으로 수집하는 크롤러를 제작하여 활용한다 (Lee et al., 2018). 데이터 분석방법은 통계처리용 프로그램을 이용하여 수치화된 결과를 모델로 생성하는 통계적 방법과 머신러닝이나 딥러닝과 같은 인공지능 학습기법들을 사용하여 모델로 생성하는 지능형 방법이 사용된다. 데이터 활용 방법으로는 정보시스템에 탑재하여 분석 결과를 시각화 하거나 독립 보고서 형태로 분석 결과를 정리하는 방식으로 활용한다. 이를 종합하면 빅데이터 관련 연구는 빅데이터 수집, 구조화, 최적화, 서비스화 등의 관점에서 다루어지고 있다고 볼 수 있다 (Chen et al. 2012; Moon and Do, 2012; Song and Kim, 2014; Song, 2017; Yu et al. 2020).

농업분야에서는 4차 산업혁명에 맞춰 빅데이터에 대한 관심은 많았으나 기존 스마트팜 환경에서 생산되는 데이터들을 빅데이터 관점으로 다루려는 시도는 최근에 시작되어 관련된 문헌이 부족하다. 하지만 이미 구축되어 있는 환경에서 많은 데이터들이 생산되었고 이를 활용하여 의미 있는 분석을 진행하는데 많은 시간이 소요되지 않을 것으로 예상된다. 기존 스마트팜 시스템에서 생산되는 데이터들은 대부분 비용문제와 유지문제로 인해 오픈 소스 기반의 관계형 데이터베이스 시스템 (Relational database

management system; RDBMS)을 활용하고 있어서 빅데이터를 저장하거나 분석하기에는 어려움이 있다. 데이터들 간의 릴레이션쉽을 중요시하는 RDBMS 특성상 자료 조회하기 위한 질의문에 JOIN 연산을 필연적으로 사용해야 하는데 여러 테이블에 대한 JOIN 연산은 레코드가 많을 경우 성능이 급격히 떨어지는 문제가 있다. 또한 한 번에 반환되는 레코드의 양이 많으면 주메모리 공간의 부족으로 가상 메모리 공간에 저장하게 되는데, 가상 메모리의 접근 속도 문제로 전체 시스템의 처리 속도가 떨어지는 문제가 있다.

따라서 본 논문에서는 기존 RDBMS를 활용하여 농작물의 생육환경정보와 생체정보 분석에 효과적인 빅데이터 모델을 제안한다. 테이블 사이의 릴레이션쉽으로 JOIN 연산에 많은 시간을 소비하는 RDBMS의 문제점을 극복하기 위해 릴레이션쉽을 최소화하면서 의미 있는 분석이 가능하도록 빅데이터 모델을 만드는 것이 목표이다. 제안된 모델의 성능 평가를 위해 실험용으로 측정된 생육환경정보와 생체정보를 일반적으로 사용되는 두 개의 RDBMS에 제안된 두 가지 모델로 저장한 후 동일한 분석용 오퍼레이션을 실행하여 반환되는 레코드의 개수에 따른 응답 시간을 측정하고 비교하는 방식으로 진행한다.

2. 빅데이터 모델 설계

2.1 목표 시스템 구성

본 논문에서는 Fig. 1과 같이 빅데이터 모델을 기반으로 하는 생체정보 모니터링 시스템을 구축하기 위한 첫 단계로 빅데이터 모델을 설계하는 것이 목표이다. 생체정보 모니터링 시스템은 네 부분으로 구성된다. 데이터 생성부 (Data generation)에서는 노지와 시설에 설치된 각종 센서로 부터 생육환경데이터와 농작물의 생체데이터를 생성한다. 데이터 전송부 (Data transmission)에서는 인터넷을 통해 센서 측정 데이터를 송수신한다. 빅데이터 모델 (Big data model)에서는 수신된 데이터들을 빅데이터 분석에 최적화된 형태로 저장하고, 빅데이터 분석용 데이터를 추출하여 실시간으로 생육지표와 스트레스지표를 생성하는 역할을 한다. 빅데이터 시각화 (Big data visualization)에서는 실시간으로 수신된 데이터들과 빅데이터 분석된 생육지표와 스트레스지표 값을 동시에 표출하고, 해당 농작물의 생육에 문제가 발생하면 알람을 통해 농가에 통지하는 기능을 수행한다. 빅데이터 모델을 구축하기 위해 관련 데이터들의 속성을 식별하고 빅데이터 분석 요소를 정의하여 RDBMS를 기반으로 하는 빅데이터 모델을 설계한다.

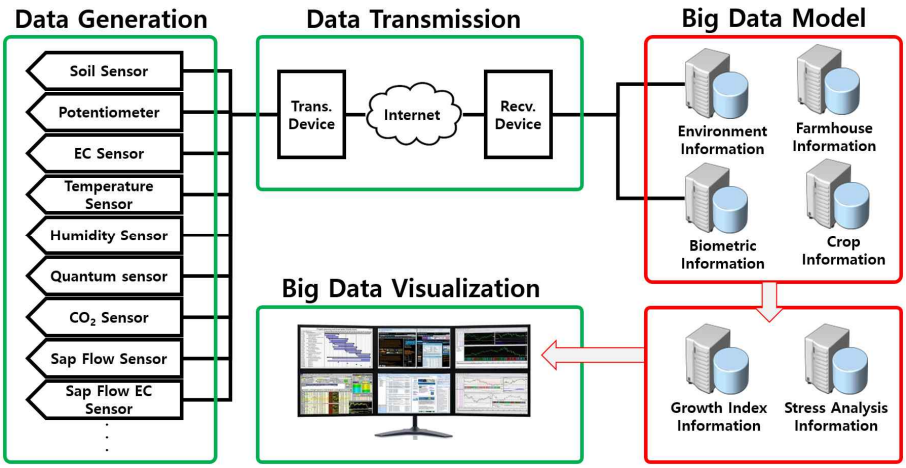

Fig. 1 Conceptual Diagram of Biometric Information Monitoring System based on Big Data Model

2.2 농작물 생육환경정보 속성 식별

농작물의 생육은 대기에서 흡수되는 이산화탄소와 뿌리에서 흡수되는 수분과 양분을 이용하여 태양 빛을 받아 광합성작용을 하고 이로 인해 생성되는 탄수화물을 주원료로 하여 이루어진다. 광합성은 온도, 광량, 수분과 양분, 이산화탄소 등에 영향을 받는다. 특히 농작물의 생리 반응들은 복잡한 물리적, 화학적 반응으로써 온도의 영향을 가장 많이 받는다 (Lee and Seol, 2019).

농작물의 종류에 따라 요구하는 환경요인에도 차이가 있기 때문에 센서를 이용해서 측정 가능한 공통 환경요인들을 식별하고 이를 모델링하는 것이 요구된다. 생육에 관여하는 공통 환경속성은 온도 (℃), 습도 (%), 강우량 (㎜), 이산화탄소농도 (PPM), 광량 (Wm-2), 광합성 광량자속밀도 (㎷) 등의 대기환경정보와 토양수분함량 (vol), 토양수분장력 (㎪), 토양pH (pH), 토양 EC (dS) 등의 토양환경정보가 모델링 대상으로 식별할 수 있다.

각 환경요인들은 고유의 측정 단위가 존재하고 센서의 종류에 따라 측정값의 범위가 다르다. 습도, 이산화탄소 농도, 토양환경 정보들은 정수값으로 측정되지만 나머지는 실수값으로 측정된다. 측정 단위는 시각화 과정에서 고려하면 되기 때문에 모델링 대상에서는 제외된다. 또한 농작물을 노지에서 재배하는지 시설에서 재배하는지에 따라 환경정보에 차이가 발생하지만 이는 측정값의 차이로 나타나기 때문에 재배지에 대한 정보도 모델링 대상에서 제외된다.

2.3 농작물 생체정보 속성 식별

농작물의 생육 상태는 생체정보 모니터링으로 확인할 수 있다. 생체정보 모니터링은 농작물의 생리요소를 자동으로 측정하거나 농작물의 외관을 수동으로 측정하여 정보로 생산한다. 농작물의 종류에 따라 확인해야 할 생체정보에 차이가 있지만, 센서를 이용해서 자동 측정 가능한 생리요소에는 광합성량 (㎛㏖), 엽온 (℃), 수액 EC (dS), 수액흐름양 (㎜), 엽록소양 (SPAD) 등이 있고, 외관 수동 측정으로 획득할 수 있는 생체정보에는 화방 (개), 초장 (㎝), 경경 (cm), 마디수 (개), 엽수 (개), 지상생체중 (g), 지하생체중 (g), 엽면적 (㎠) 등이 모델링 대상 정보로 식별된다. 여기서도 측정 단위와 재배지에 대한 정보는 모델링 대상에서 제외된다.

수동 측정의 경우, 측정할 때마다 측정값이 달라지는 측정오류가 발생할 수 있다. 이는 데이터의 정확성 문제를 야기하지만 데이터 모델과는 무관하며 향후 빅데이터 측면에서는 무시해도 될 만한 수준으로 이해된다.

2.4 빅데이터 분석 요소 정의

농작물의 생육환경정보와 생체정보를 기반으로 농작물의 생육이 어떻게 진행되고 있으며 향후 어떻게 진행될 것인지에 대한 분석 기준으로 삼을 수 있는 것이 농작물 생육지표와 스트레스지표이다. 생육지표는 농작물별로 설정된 생육단계별 목표값에 어느 정도 도달하고 있는지를 확인할 수 있는 척도로 인식되고, 스트레스지표는 농작물의 생육에 문제가 있다면 그 원인을 식별할 수 있는 척도가 된다.

생육지표는 생체정보에서 수동으로 측정되는 화방 (개), 초장 (㎝), 경경 (cm), 마디수 (개), 엽수 (개), 지상생체중 (g), 지하생체중 (g), 엽면적 (㎠) 등을 대상으로 한다. 스트레스지표는 망간 (Mn), 3가 크롬 (Cr3+), 6가 크롬 (Cr6+), 양이온 염 (Na/K/NH4/Mg/Ba/Ca/Pb/Ag), 음이온 염 (NO3/Cl/S/SO4/CO3), 고온, 저온, 다습 등을 대상으로 한다.

2.5 빅데이터 모델

본 논문에서는 기존 RDBMS를 활용하는 두 가지 빅데이터 모델을 제안한다. 첫 번째 모델 (Model-1)은 JOIN 연산이 많은 일반적인 정보처리 시스템에서 다루는 방식이고, 두 번째 모델 (Model-2)은 JOIN 연산을 줄인 새롭게 제안하는 방식이다.

첫 번째 모델 (Model-1)은 Fig. 2와 같이 전형적인 RDBMS로 구성된 생육지표정보 (Growth index information)와 스트레스지표정보 (Stress analysis information)를 JOIN 문장을 사용해서 생성하는 모델

이다. 이 모델에서는 생육환경정보 (Environment information)와 생체정보 (Biometric information)의 속성들이 각 특성에 따라 분리되어 저장된다. 생육환경정보는 대기환경정보 (Air environment information)와 토양환경정보 (Soil environment information)로 분리되어 저장되고, 생체정보는 자동측정생체정보 (Auto-measuring biometric information)와 수동측정생체정보 (Manual- measuring biometric information)로 분리되어 저장된다. 각 정보를 분리시킨 이유는 데이터 측정 위치, 측정 시점, 측정 방법이 다르기 때문이다.

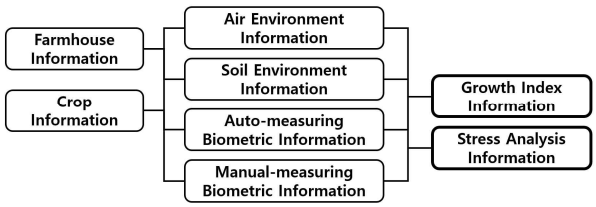

Fig. 2 The Proposed Big Data Model-1

농작물 재배환경은 노지나 시설에서 공통적으로 일정한 재식거리를 유지하고 있다. 토양환경정보 측정용 센서들은 측정 대상 농작물과 가까운 곳에 설치가 가능하지만 대기환경정보 측정용 센서들은 측정 대상 농작물과 대체로 떨어져 설치된다. 또한 센서들마다 고유 측정주기를 가지고 있어서 동일한 시점에 측정이 되지 않는다. 실험용 농작물의 경우 측정주기가 짧게는 10분, 길게는 하루에 두 번 측정하는 경우가 많다. 실제 농가에서는 하루에 한 번에서 두 번 정도 측정하면 되기 때문에 정확하게 측정 시점을 맞출 필요는 없지만 정확한 데이터 확보를 위해 이 부분은 향후 동일 시점에 측정될 수 있도록 맞추는 것이 필요하다. 생체정보는 센서를 이용해서 자동으로 측정되는 데이터와 수동으로 직접 측정할 수밖에 없는 데이터들이 있다. 이 모델에서 생육지표와 스트레스지표를 생성하기 위해서는 총 4번의 JOIN을 거쳐야 한다.

두 번째 모델 (Model-2)은 Fig. 3과 같이 생육지표정보 (Growth index information)와 스트레스지표정보 (Stress analysis information)가 생육환경정보 (Environment information)와 생체정보 (Biometric information)에서 분석에 필요한 것들만 필터링 (Index filtering)해서 생성하도록 구성되었다. 이 모델은 생육환경정보와 생체정보의 속성을 분리하지 않고 각각 통합하여 JOIN 문장의 사용을 줄였다. 생육환경정보를 통합한 이유는 측정 시점을 일치시키기 위한 것이다. 대기환경정보와 토양환경정보가 다른 시점의 자료가 된다면 시간을 기준으로 분석할 때 잘못된 결과가 나올 수 있기 때문이다.

생체정보를 자동측정 데이터와 수동측정 데이터를 통합한 것은 한 시점에 하나의 의미있는 데이터를 생성하기 위해서이다. 이 모델에서 생육지표와 스트레스지표를 생성하기 위해서는 한번만 JOIN을 수행하면 된다. 이 모델은 향후 빅데이터 분석용 인덱스를 생성하는 엔진 (IndexeEngine)을 거쳐 빅데이터에 빠르게 접근할 수 있도록 별도 인덱스 저장소 (Index storage)를 구축하고, 시각화 모듈 (Visualization module)에서 인덱스 저장소를 통해 표출할 데이터에 불러올 수 있도록 할 계획이다.

3. 성능 평가

본 논문에서는 제안된 두 가지 모델의 성능 분석을 위해 실험용으로 측정된 생육환경데이터와 생체데이터를 각각 1만 건 확보하였다. 실험에 필요한 데이터는 1억 건 (10^8)이기 때문에 원본 데이터를 자가복제하여 시험용 데이터로 만들었다. 특히 생체데이터 중에서 수동으로 측정해야 하는 데이터는 자료확보가 어려워 확보된 자료를 별도로 자가 복제하여 1만 건을 생성한 후 다시 1억 건을 생성하였다.

성능 평가는 데이터의 품질이나 정확성과는 무관

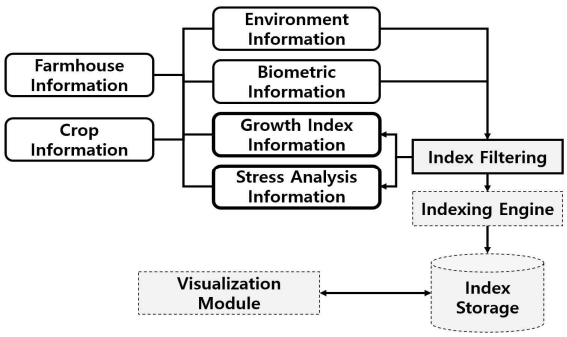

Fig. 3 The Proposed Big Data Model-2

하며 데이터의 양이 중요하기 때문에 자가 복제를 통한 자료 확보에 문제가 없는 것으로 판단된다. 데이터를 1억 건으로 한정한 것은 RDBMS 성능 평가를 자료의 양으로 측정하는데 기준이 되기도 하고, 빅데이터라고 부를 수 있는 시작점이기 때문이다. 자가 복제는 1만 개의 레코드에 대해 "SELECT INTO" 문장을 1만 번 반복 수행하는 저장 프로시저(Stored procedure)를 별도로 작성하여 두 RDBMS에 대해 1회 수행하는 방식으로 진행되었다.

성능 평가에 사용된 RDBMS는 농업관련 시스템에서 제일 많이 사용하고 있는 Oracle 11g와 MySQL 8.0이다. 최근에는 MariaDB도 많이 사용하고 있지만 MariaDB 엔진이 MySQL과 동일하기 때문에 같은 성능으로 생각해도 무방하다. Oracle과 MySQL 각각에 제안된 모델 두 가지를 구현하였다.

첫 번째 모델 (Model-1)에는 7개의 테이블이 구현되었고, 두 번째 모델 (Model-2)에는 6개의 테이블이 구현되었다. 실제 데이터가 들어있는 테이블은 생육환경정보와 생체정보 테이블뿐이다. 1억 건의 자료를 저장시킨 후 Oracle 11g의 데이터베이스 크기는 27.64GB 이였고, MySQL 8.0은 27.91GB 이였다.

각 테이블에는 레코드를 식별하기 위한 고유키(Primary key; PK)와 레코드 생성일자를 저장할 수 있는 필드를 추가하였고, 생육환경정보와 생체정보를 연결하기 위해 농가정보의 고유키를 외래키 (Foreign Key; FK)로 설정하였다. 생육환경정보에서 대기환경정보의 레코드 하나는 48바이트, 토양환경정보의 레코드 하나는 40바이트이고, 이를 통합한 생육환경정보의 레코드 하나는 76바이트이다. 생체정보에서 자동측정 생체정보와 수동측정 생체정보의 레코드 하나는 각각 44바이트이고, 이를 통합한 생체정보의 레코드 하나는 64바이트 이다.

성능 평가 방식은 RDBMS에 저장되어 있는 1억 건의 레코드를 10의 승수씩 증가시켜가며 총 8회 반복하여 조회 질의문을 실행하고 해당 레코드가 반환되는 응답시간을 측정하였다. 반환된 레코드를 화면에 출력하는데 소요되는 시간을 무시하기 위해서 화면에는 10건만 표시하도록 하였다. 모든 반복에 대해 매회 반복하기 전에 프로그램 캐쉬에 저장되어 있는 레코드를 삭제하여 다음 반복에 영향을 주지 않도록 하였다.

첫 번째 모델 (Model-1)의 성능 평가 결과는 Fig. 4와 같았다. Oracle과 MySQL 모두 반환 레코드가 10건부터 10만 건까지는 응답시간이 거의 차이가 없었고, 백만 건부터 비교 가능한 응답시간이 소요되었다. Oracle은 백만 건부터 1억 건까지 3 구간의 평균 응답시간이 20,055ms 이고, MySQL은 28,853ms 였다. 최대 응답시간은 Oracle이 32,081ms이고, MySQL은 43,227ms 였다.

두 번째 모델 (Model-2)의 성능 평가 결과는 Fig. 5와 같았다. 의미 있는 응답시간 구간에서 Oracle의 평균 응답시간은 15,817ms 이고, MySQL은 21,992ms 였다. 최대 응답시간은 Oracle이 25,301ms 이고, MySQL은 32,948ms 소요 되었다.

제안된 두 모델의 전체구간 평균 응답시간 차이는 Oracle이 모델-2에서 2,328ms 빨랐고, MySQL은 모델-2에서 3,629ms 빨랐다. 의미 있는 3 구간의 평균 응답시간 차이는 Oracle이 모델-2에서 4,238ms 빨랐고, MySQL은 모델-2에서 6,861ms 빨랐다.

제안된 두 모델 사이의 응답시간 개선 효과를 비교하기 위해 RDBMS별로 나타내면 Oracle은 Fig. 6과 같고, MySQL은 Fig. 7과 같다. 두 개의 RDBMS 모두 모델-2가 모델-1보다 응답시간이 빠른 것으로 측정되었다.

제안된 모델-2에서 레코드 반환 응답시간이 단축된 것은 생육환경정보의 대기환경정보와 토양환경정보를 통합하고 생체정보의 자동측정 생체정보와 수동측정 생체정보를 통합하여 생육지표와 스트레스지표 생성을 위한 JOIN 문장을 최소화하였기 때문인 것으로 판단된다. 두 개의 RDBMS에서 측정된 전체구간의 응답시간과 의미있는 구간의 응답시간에 대한 시간 단축 비율은 Table 1과 같이 Oracle에서는 21.1% (4.2초), MySQL에서는 23.8% (6.8초) 단축되는 효과가 있는 것으로 확인되었다. 데이터가 1억 건을 초과한다면 제안된 모델-2와 같은 방식이 응답시간 단축에 더 큰 효과가 있을 것으로 기대된다.

두 모델에 대한 성능평가에 사용된 하드웨어 환경은 Intel Core i7-3770 (3.4Ghz) 프로세서와 16GB 메모리를 장착한 HP Compaq 8300 Elite CMT PC를 사용하였다. 운영체제는 Windows 10 Pro (빌드 14393) 64비트를 사용하였으며, Oracle 11g와 MySQL 8.0 이외에는 데이터베이스 관련 소프트웨어는 설치하지 않았다.

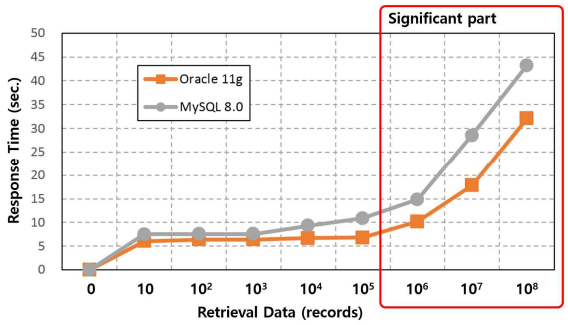

Fig. 4 Response Time of the Proposed Big Data Model-1

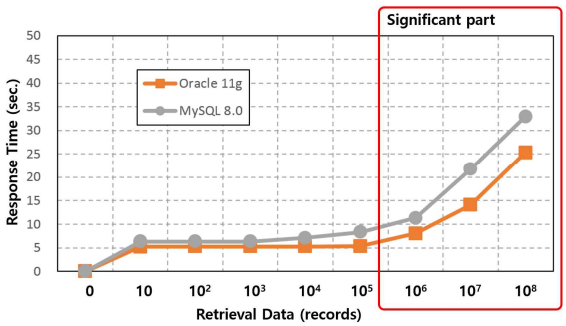

Fig. 5 Response Time of the Proposed Big Data Model-2

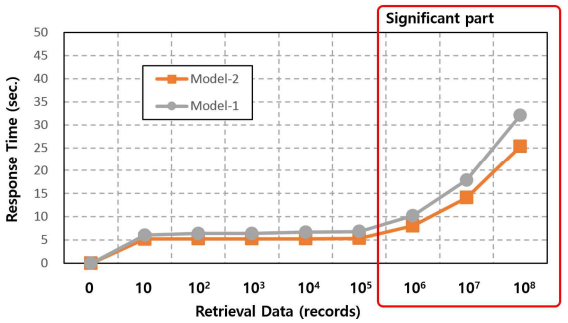

Fig. 6 Comparison of Response Time between the Two Models on Oracle 11g

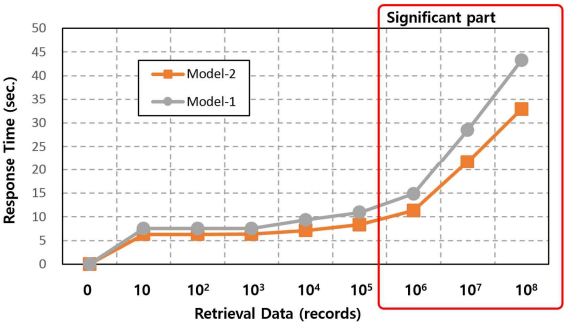

Fig. 7 Comparison of Response Time between the Two Models on MySQL 8.0

Table 1 Response Time of the Two Models (unit: millisecond)

DBMS / Records	Oracle 11g			MySQL 8.0		
	Model-1	Model-2	Difference	Model-1	Model-2	Difference
10	6,041	5,235	-806	7,529	6,317	-1,212
10^2	6,390	5,264	-1,126	7,535	6,322	-1,213
10^3	6,392	5,266	-1,126	7,542	6,328	-1,214
10^4	6,689	5,276	-1,413	9,330	7,112	-2,218
10^5	6,806	5,368	-1,438	10,918	8,322	-2,596
10^6	**10,221**	**8,061**	**-2,160**	**14,830**	**11,304**	**-3,526**
10^7	**17,864**	**14,089**	**-3,775**	**28,501**	**21,724**	**-6,777**
10^8	**32,081**	**25,301**	**-6,780**	**43,227**	**32,948**	**-10,279**
whole section avg.		-2,328			-3,629	
10^6~10^8 section avg.		**-4,238**			**-6,861**	
maximum difference		-6,780			-10,279	
minium difference		-806			-1,212	
10^6~10^8 section reduced time ratio		**21.1%**			**23.8%**	

4. 결 론

본 논문은 농작물의 생육환경정보와 생체정보 분석에 효과적인 저비용 RDBMS 활용 빅데이터 모델을 제안하였다. RDBMS의 단점을 극복하기 위해 JOIN 문장의 사용을 줄이고 분석에 유의미한 자료를 선별하는 방식을 제시하여 성능 평가를 한 결과 빅데이터 분석 속도를 최대 23.8% 개선하였고, 데이터의 양이 많아지면 더욱 효과가 클 것으로 예상된다. 제안된 모델은 기존의 저가형 RDBMS를 그대로 활용할 수 있는 장점이 있어서 스마트농업 활성화와 안정화에 기여할 것으로 기대된다. 이를 위해 향후 생육환경정보와 생체정보에서 지표 필터링 (Index filtering)을 거쳐 빅데이터에 대한 인덱스를 생성하는 인덱싱 엔진 (Indexing engine)을 완성하게 되면 지금보다 응답시간을 더 단축하여 저가형 빅데이터 시스템 구축에 큰 도움이 될 것으로 기대한다.

References

Chen, H., Chiang, Roger, H. L., and Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly: Management Information Systems*, 36(4), 1165-1188. https://doi.org/10.2307/41703503

Choi, Y. C., and Jang, I. H. (2019). Smart Farm in the Era of the 4th Industrial Revolution, *The Journal of the Korean Institute of Communication Sciences*, 36(3), 9-16.

Korea Agro-Fisheries and Food Trade Corporation, (2019). *Cultivation Status and Management of Agricultural Crops using US Smart Agriculture and IT System*, 17-24.

Lee, J. K., and Seol, B. M. (2019). A Study on Intelligent Smart Farm Utilization and Productivity: Focusing on the Case of Tomato Farmers, *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, 14(3), 185-199.

Lee, J. Y., Moon, C. B., and Kim, B. M (2018). Music Crawler for Mood-based Music Classification and Retrieval systems, *Journal of the Korean Information Science Society 2018 Korea Computer Science Conference*, June, 699-701, Jeju.

Ministry of Science and ICT (2019). Science and Technology Ministers' Meeting "*Science and Technology Strategy for Future Agriculture: Focusing on Smart Agriculture and Agricultural Biotechnology,*" Science and Technology Ministers' Meeting, 2, 12-13.

Moon, D. H., and Do, N. C. (2012). Big Data, *Journal of the Korean Computational Design and Engineering Association*, 18(2), 26-29.

Park, K. B., Kim, J. Y., and Kim, H. M. (2019). BigData Research in Information Systems: Focusing on Journal Articles about Information Systems, *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 9(6), 681-689.

Song, M. G. (2017). Development of a Big Data System to accept Energy Big Data, *Journal of the Korean Information Science Society*, 415-417.

Song, J. H., and Kim, J. S. (2014). Analysis of the Best Practices Big Data Services, *Journal of the Korea Contents Association Review*, 12(1), 32-37.

Tomcy, J., and Pankaj, M. (2017). *Data Lake for Enterprises - Leveraging Lambda Architecture for Building Enterprise Data Lake*, Packt Publishing.

Yu, S. S., Choi, K. P., Myung, H., and Oh, H. J. (2020). Prediction Model of Pest According to Individual Farms Based on

Heterogeneous Public Big data, *Journal of Korea Information Technology Society*, 18(6), 1-9.

이 종 열 (JongYeol Lee)

- 정회원
- 금오공과대학교 컴퓨터공학과 공학사
- 금오공과대학교 컴퓨터공학과 공학석사
- 금오공과대학교 소프트웨어공학과 박사수료
- (현재) 금오공과대학교 컴퓨터소프트웨어공학과 강사
- 관심분야: 인공지능, 정보검색, 농업ICT

문 창 배 (ChangBae Moon)

- 정회원
- 금오공과대학교 소프트웨어공학과 공학사
- 금오공과대학교 소프트웨어공학과 공학석사
- 금오공과대학교 소프트웨어공학과 공학박사
- (현재) 금오공과대학교 ICT융합특성화센서 연구교수
- 관심분야: 인공지능, 지식검색, 감성공학

김 병 만 (ByeongMan Kim)

- 비회원
- 서울대학교 컴퓨터공학과 학사
- 한국과학기술원 전산학과 석사
- 한국과학기술원 전산학과 박사
- (현재) 금오공과대학교 컴퓨터소프트웨어공학과 교수
- 관심분야: 인공지능, 정보검색, 정보보안