

# 오픈소스 기반 빅데이터 플랫폼의 에너지 하베스터 최적설계 적용 연구

유은섭\*, 김석찬\*\*, 이한민\*\*, 문두환\*<sup>#</sup>

\*경북대학교 정밀기계공학과, \*\*한국기계연구원 기계시스템안전연구본부

## Application of Open Source, Big Data Platform to Optimal Energy Harvester Design

Eun-seop Yu\*, Seok-Chan Kim\*\*, Hanmin Lee\*\*, and Duhwan Mun\*<sup>#</sup>

\*Department of Precision Mechanical Engineering, Kyungpook National University

\*\*Mechanical Systems Safety Research Division, Korea Institute of Machinery & Materials

(Received 19 February 2018; received in revised form 2 March 2018; accepted 4 March 2018)

### ABSTRACT

Recently, as interest in the internet of things has increased, a vibration energy harvester has attracted attention as a power supply method for a wireless sensor. The vibration energy harvester can be divided into piezoelectric types, electromagnetic type and electrostatic type, according to the energy conversion type. The electromagnetic vibration energy harvester has advantages, in terms of output density and design flexibility, compared to other methods. The efficiency of an electromagnetic vibration energy harvester is determined by the shape, size, and spacing of coils and magnets. Generating all the experimental cases is expensive, in terms of time and money. This study proposes a method to perform design optimization of an electromagnetic vibration energy harvester using an open source, big data platform.

**Key Words** : Electromagnetic Vibration Energy Harvester(전자기형 진동 에너지 하베스터), Hadoop(하둡), Design Optimization(설계 최적화), Machine Learning(기계 학습), R(알)

### 1. 서 론

최근 정보 통신 기술의 발전으로 인해 처리해야 할 데이터의 양과 그 복잡성이 증가해 빅데이터(big data)에 대한 관심이 증가하고 있다. 빅데이터는 단순히 큰 용량의 데이터를 의미하는 것이 아니라 빅데이터의 수집부터 처리까지의 전 과정을 포괄하는

개념이다. 미국의 정보 기술 연구 및 자문 회사인 Gartner의 연구자 laney는 2001년 데이터의 급성장에 따른 이슈와 기회를 데이터의 양(volume), 입출력 속도(velocity), 종류의 다양성(variety)이라는 세가지 차원으로 정의하는 3V 모델을 제시했다<sup>[1]</sup>. 그리고 Gartner는 2012년 빅데이터는 큰 용량, 빠른 속도, 높은 다양성을 갖는 정보 자산으로 이를 통해 의사 결정, 새로운 가치 발견, 프로세스 최적화를 위해서는 새로운 형태의 처리 방식이 필요하다고 정의했다<sup>[2]</sup>.

# Corresponding Author : dhmun@knu.ac.kr

Tel: +82-54-530-1271, Fax: +82-54-530-1278

빅데이터는 주로 서비스 분야에서 활발히 이용되고 있지만 최근 센서의 발달로 인해 제조업에서도 충분한 양의 데이터 얻는 것이 가능해져 제조업에 빅데이터 기술을 적용하려는 연구가 활발히 수행되고 있다. 특히, 제조업에서 생산 공정 최적화 및 제품의 결함 진단이나 플랜트 산업에서 설비의 고장 예지보전에 대한 연구가 많이 이루어지고 있다.

제조업에서의 빅데이터 적용 사례를 살펴보면, 자동차 제조업체인 Volvo는 자동차 주행 중에 발생하는 데이터를 수집하여 종래에 50만대의 차가 팔린 후 발견하였을 결함을 1,000대가 판매된 시점에서 조기 파악할 수 있었다<sup>[3]</sup>. 풍력 발전 터빈 및 풍력단지 설비 제조업체인 Vestas는 풍력 발전소 부지 선정에 위해 날씨, 지리, 산림지도, 해상지도 등 페타바이트 규모의 빅데이터를 수집하고 분석함으로써 종래에 27km<sup>2</sup>로 분석되던 후보지의 넓이를 10m<sup>2</sup>로 대폭 감소시켰다<sup>[4]</sup>. 데이터 저장장치 제조업체인 Western Digital은 제품이 생산라인에 있을 때부터 제품의 상태를 기록 및 추적함으로써 결함 제품을 조기에 발견하여 제품의 결함율을 1.9%로 대폭 감소시켰다<sup>[5]</sup>.

제조업에서 최적설계 등의 목적으로 빅데이터를 활용하기 위해서는 빅데이터를 관리하기 위한 저장소 및 이 저장소와 연동된 데이터 분석 환경을 제공하는 빅데이터 플랫폼의 구축이 필요하다. 그리고 최근에는 비용 절감, 빠르고 유연한 개발, 호환성 및 유연성 확보를 위해서 이와 같은 플랫폼을 오픈소스(open source) 기반 도구들을 활용하여 구축하는 추세이다.

국내 기계가공 및 설계 분야에도 기계학습을 적용하기 위한 연구들이 늘어나고 있지만 기계학습을 위한 데이터를 빅데이터로서 어떻게 관리할지에 대한 연구는 많지 않다<sup>[6,7,8]</sup>.

이 연구에서는 오픈소스 기반 빅데이터 저장소인 하둡(Hadoop), 통계처리 프로그램인 R, 하둡과 R을 연계하는 Rhadoop을 이용하여 빅데이터 플랫폼을 구축하고 전자기형 진동 에너지 하베스터의 최적설계에 적용하는 방법을 제안한다.

2절에서는 오픈소스 기반 빅데이터 플랫폼을 구축하고 전자기형 진동 에너지 하베스터의 기계학습용 데이터를 저장하는 방법을 제시한다. 3절에서는

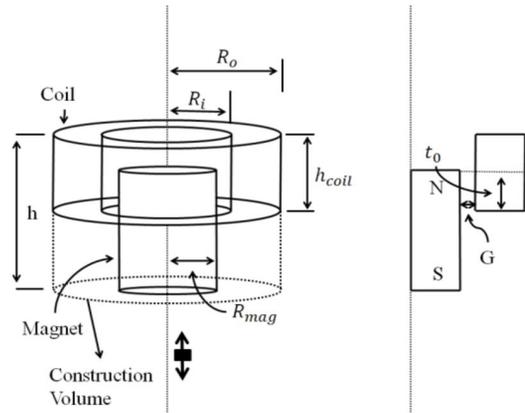


Fig. 1 Design parameters of the harvester

R을 이용하여 인공지능망을 이용한 회귀분석을 통해 전자기형 진동 에너지 하베스터의 최적설계를 수행한 결과를 설명한다. 마지막으로 4절에서 결론을 맺는다.

## 2. 오픈소스 기반 빅데이터 플랫폼을 이용한 기계학습용 데이터 구축

이 연구에서 최적설계 대상으로 하고 있는 하베스터의 기계학습용 데이터로 Kim et al.<sup>[9]</sup>이 수행한 전자기형 진동 에너지 하베스터의 최적설계 연구의 데이터를 사용했다.

### 2.1 하베스터의 구조

하베스터의 구조는 Fig. 1과 같이 중앙에 구멍이 있는 원통형 코일이 고정되어 있고 원통형 자석이 코일의 구멍에 침투하여 상하 진동하는 형태이다.

하베스터의 유동 설계변수는 하베스터 전체 체적의 종횡비( $AR = h/2R_0$ ), 하베스터 전체 체적의 반경 대비 자석의 반경 비( $R_{mag}ratio = R_{mag}/R_0$ ), 하베스터 전체 체적의 높이 대비 코일의 높이 비( $h_{coil}ratio = h_{coil}/h$ ), 코일의 높이 대비 코일에 침투한 자석의 높이 비( $t_0ratio = t_0/h_{coil}$ )가 있다.

하베스터의 고정 설계변수는 하베스터의 전체 체

적( $V_{const}$ )  $8\text{cm}^3$ , 자석의 최대 변위( $Z_{lim}$ )  $0.5\text{mm}$ , 자석과 코일의 간격( $G$ )  $0.5\text{mm}$ , 자석의 밀도( $\rho_{mag}$ )  $7400\text{kg/m}^3$ , 코일 충전률( $k_{co}$ )  $0.6$ , 코일 권선의 직경( $d_{co}$ )  $0.1\text{mm}$ , 코일 권선의 저항( $R_p$ )  $2.2/\text{m}$ 가 있다.

## 2.2 학습데이터의 구조

하베스터 최적설계 데이터의 변수는 입력변수인 하베스터의 유동 설계변수와 출력변수인 하베스터의 설계변수에 따른 출력 전압( $V_{rms}$ )과 전력( $P_{avg}$ )으로 구성되어 있다.

하베스터의 설계변수에 따른 출력 전압 및 전력의 비선형성을 잘 표현하기 위해서는 인자의 수준이 높아야 한다. 하지만 수준이 높아질수록 실험 횟수가 기하급수적으로 늘어나기 때문에 적절한 수준을 결정해야 한다.

이 연구에서 사용한 데이터는 별로 9수준 전산 직교배열  $L_{81}9^3$ 을 이용하여 하베스터의 설계변수 값을 조합했다.  $n$ 개의 설계변수에 대해 9수준의 완전 요인 배치법을 적용하면 9<sup>n</sup>번의 실험이 필요하지만 직교 배열법을 이용하면 81번의 실험을 통해 데이터 세트를 만들 수 있다.

하베스터 최적설계 데이터의 입력변수의 값은 0.1, 0.2, 0.4, 0.8, 1.6, 3.2의 값을 갖는  $AR$ 과 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9의 값을 갖는  $R_{magratio}$ ,  $h_{coilratio}$ ,  $t_{0ratio}$ 를 조합하여  $AR$  별로 81개 총 486개 데이터로 구성되어 있다. 출력변수의 값은 설계변수 값의 조합에 따라서 유한요소 해석을 이용하여 계산한 기계-전기 변환계수와 하베스터의 코일과 자석의 운동을 근사한 미분방정식을 수치적분함으로써 계산한 결과이다.

## 2.3 오픈소스 기반 빅데이터 플랫폼

하둡은 대표적인 오픈소스 기반 빅데이터 저장소로서 간단한 프로그래밍을 통해 다중 클러스터 서버에 저장된 데이터를 효율적으로 처리할 수 있다. 하둡은 단일 서버에서부터 수천개의 다중 서버까지 확장할 수 있으며 각 서버의 고장을 감지하고 처리할 수 있도록 설계되어 안정적인 빅데이터 운용 환경을 구축할 수 있다.

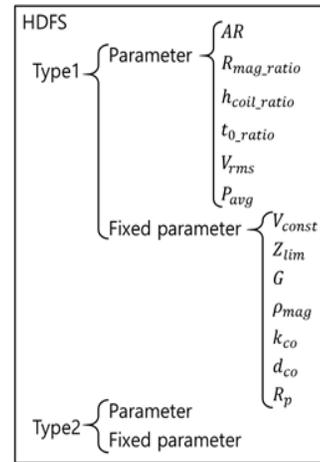


Fig. 2 Structure of training data stored in HDFS

하둡의 주요 구성요소로는 HDFS (Hadoop Distributed File System)과 맵리듀스(MapReduce)가 있다. HDFS는 하나의 데이터 파일을 여러 개의 서버에 중복하여 분산 저장함으로써 하나의 서버가 고장 나더라도 다른 서버를 통해 복구함으로써 데이터 보관의 안정성을 확보하고 데이터 처리시에 각 서버의 자원을 활용함으로써 데이터가 커질수록 처리속도향상의 이점이 있다. 맵리듀스는 여러 개의 서버에 분산되어 있는 데이터를 쉽게 처리할 수 있도록 하는 기능을 제공한다.

하둡의 운용을 위해서는 다소 복잡한 설정을 요구하기 때문에 비숙련자가 직접 운용하기는 어렵다. 이러한 이유로 일부 업체에서는 하둡의 대표적인 기능을 구현하여 제공하고 있다. 대표적인 예로 IBM사의 InfoSphere, Cloudera사의 CDH, Microsoft사의 HDInsight, Hortonworks사의 HDP 등이 있다.

하둡은 HDFS, 맵리듀스 등을 통해 빅데이터의 저장 및 관리를 지원할 수 있지만 통계 처리 등의 해석 기능을 제공하지 않는다. 이를 해결하기 위한 방법으로 오픈소스 통계 처리 프로그램인 R과 하둡을 연동하는 방법이 있다. R은 데이터 처리를 위한 프로그래밍 언어를 지원하며 고급 통계분석 및 가시화를 위한 많은 패키지가 존재한다. 그리고 하둡과의 연동을 위한 Rhadoop 패키지가 존재한다. Rhadoop의 세부 패키지는 R에서 하둡의 HDFS를 제어하는 rhdfs 패키지와 R에서 하둡의 맵리듀스를

제어하는 mmr2 패키지 등이 존재한다.

이 연구에서는 R에서 rhdfs 패키지를 이용하여 학습 데이터를 R 객체 파일 형식으로 하둠의 HDFS에 저장했다. 그리고 저장된 데이터를 R에서 불러와 기계학습을 이용한 최적설계를 수행했다. 데이터의 확장성을 고려하여 HDFS에 저장한 학습 데이터의 구조를 Fig. 2와 같이 정의했다.

### 3. 오픈소스 기반 빅데이터

#### 플랫폼을 이용한 하베스터 최적설계

##### 3.1 하베스터 최적설계 위한 R 패키지 및 방법 선정

이 연구에서는 하베스터 최적설계를 위한 기계학습 알고리즘으로 여러 개의 독립변수에 대한 비선형 회귀분석 기능을 제공하는 인공신경망(ANN, artificial neural network)을 선택했다. 그리고 오픈소스 통계처리 프로그램인 R에서 인공신경망을 사용하기 위한 패키지로 nnet<sup>[10]</sup>를 선택했다.

기계학습 알고리즘으로 생성한 회귀모델이 실제 데이터를 얼마나 잘 표현하는지 확인하기 위한 척도로 RMSE(root mean square error)를 이용했다. RMSE는 입력변수의 값에 따른 실제 결과값( $y_1$ )과 기계학습 알고리즘으로 예측한 값( $y_2$ )의 차이를 나타내는 함수로 계산식은 식(1)과 같다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{1,i} - y_{2,i})^2}{n}} \quad (1)$$

nnet패키지의 경우 학습모델을 튜닝하기 매개변수를 제공하는데 튜닝 매개변수가 높을수록 학습데이터에 대한 예측 정확도는 올라가지만 학습데이터에 포함되지 않은 설계변수 값의 조합에 대해선 예측 정확도가 떨어지는 과적합 문제가 발생할 가능성이 있다. 과적합 문제를 피하고 학습모델을 최적화하기 위해서는 학습데이터를 학습용과 검증용으로 나누고 학습용 및 검증용 데이터에 대한 회귀모델의 RMSE가 모두 최소가 되는 튜닝 매개변수 값을 찾아야한다. 이때 데이터를 가능한 한 고르게

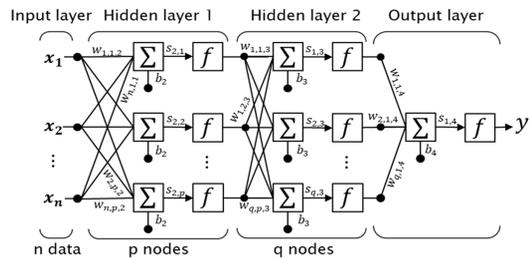


Fig. 3 Structure of artificial neural network

나누기 위하여 AR별로 50개 총 300개 데이터를 무작위로 추출하여 학습용 데이터로, 나머지 31개 총 186개 데이터를 검증용 데이터로 나누었다.

하베스터 최적설계 데이터를 회귀분석한 결과를 토대로 최적설계 치수를 확인하기 위해선 학습이 완료된 회귀모델을 가시화하여 분석해야 한다. 따라서 위 알고리즘을 이용한 회귀분석 결과를 xyz 축 3차원 그래프 및 xy축 등고선 그래프로 가시화 기능을 제공하는 visreg<sup>[11-12]</sup> 패키지를 설치했다.

##### 3.2 인공신경망을 이용한 회귀분석

인공신경망은 생체 신경망의 신호전달 체계인 뉴런을 모방한 수학적 모델이다. 인공신경망의 구조는 데이터를 받아들이는 입력 계층(input layer)과 예측결과를 출력하는 출력 계층(output layer)가 있고 입력 계층과 출력 계층 사이에 보이지 않는 은닉 계층(hidden layer)으로 구성되어 있다. 은닉 계층은 은닉 노드(hidden node)들로 구성되어 있다. 은닉 노드는 신경망에서의 뉴런의 역할을 하며 가중치( $w$ ), 편향( $b$ ), 활성화함수( $f$ )로 구성되어 있다. 이 중에서 실제로 학습이 이루어지는 부분은 가중치와 편향이다. 인공신경망의 구조를 도식화한 그림은 Fig. 3와 같다.

R의 nnet 패키지는 은닉 계층(hidden layer)이 1개인 ANN 알고리즘을 제공한다. 학습모델을 튜닝하기 위한 매개변수는 은닉 노드(hidden node)의 수와 최대 반복 학습 횟수가 있다. 은닉 노드는 은닉 계층에 포함된 노드를 말한다. nnet 패키지의 반복 학습은 매 학습에서 손실함수(loss function) 결과값의 변화량이 일정 수준 이하로 떨어지면 중단되는데 반복 학습 횟수가 최대 반복 학습 횟수에 도달하면

**Table 1 Abnormal RMSEs from training of ANN**

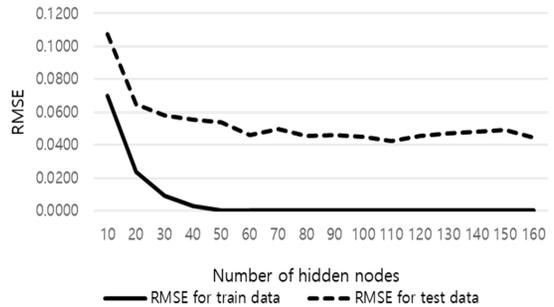
times	number of hidden nodes			
	10		20	
	train	test	train	test
1	0.069	0.086	0.022	0.075
4	0.062	0.11	0.03	0.08
5	0.066	0.087	0.023	0.062
19	0.0741	0.1367	0.0235	0.0730
20	0.0890	0.1129	0.0218	0.0541
21	0.0786	0.1501	0.0254	0.0540
22	0.0727	0.1049	0.0227	0.0759
23	0.0657	0.1017	0.0223	0.0623
24	0.0700	0.1052	0.0243	0.0527
34	0.0660	0.1095	0.0203	0.0629
35	0.0693	0.1057	0.4915	0.0520
36	0.0646	0.0991	0.0246	0.0532
40	0.0632	0.1276	0.0249	0.0649
mean	0.0698	0.1051	0.0356	0.0793
truncated	0.0698	0.1037	0.0234	0.0636

손실함수의 결과값과는 상관없이 학습을 중단한다. 은닉 노드 수와 반복 학습 횟수가 많을수록 학습모델이 복잡하고 학습한 데이터를 잘 표현하지만 과적합 문제가 발생할 가능성 또한 커진다.

최적 은닉 노드 수를 찾기 위하여 최대 반복 학습 횟수를 2000으로 제한했다. 그리고 은닉 노드의 수는 10에서 160까지 10단위로 증가시켰다. 각 은닉 노드 수에 대하여 40개의 회귀모델을 생성했으며 각 회귀모델은 모두 무작위로 추출된 학습용 데이터를 이용하여 학습시켰다. 은닉 노드 수별 회귀모델의 RMSE를 계산한 결과, Table 1과 같이, 주변 값과는 달리 RMSE가 유난히 큰 값이 관찰되었다.

RMSE가 유난히 큰 이유는 학습용 데이터를 무작위로 추출하는데 있어 데이터 조합이 고르게 추출되지 않았거나 회귀모델을 학습하는데 있어 초기 값이 잘못 설정되어 학습이 제대로 이루어지지 않았기 때문이라고 판단했다. 따라서 RMSE가 유난히 큰 값을 제거하기 위하여 은닉 노드 수별로 RMSE 결과의 상하위 20%를 절사평균한 값을 해당 은닉 노드 수에 대한 최종 RMSE로 결정했다. 은닉 노드 수별 학습용 및 검증용 데이터에 대한 RMSE 그래프는 Fig. 4와 같다.

은닉 노드 수별 학습용 및 검증용 데이터에 대한



**Fig. 4 RMSEs according by number of hidden nodes**

RMSE 그래프를 보면 학습용 데이터에 대해선 은닉 노드 수가 50일 때 RMSE가 0에 수렴하고 검증용 데이터에 대해선 은닉 노드 수가 늘어남에 따라 RMSE가 다시 커지는 과적합의 양상이 보이는 않지만 은닉 노드 수가 100개 이후론 RMSE의 변화가 거의 없어 최적 은닉 노드 수를 100개로 결정했다.

하베스터의 최적설계변수를 찾기 위해서 데이터에 최적화된 튜닝 매개변수인 은닉 노드 수 100, 최대 반복 학습 횟수 2000인 ANN 알고리즘에 전체 데이터를 학습시켜 회귀모델을 생성했다. visreg 패키지를 이용하여 생성된 회귀모델을 출력변수인 하베스터의 출력 전압에 대한 설계변수 그래프로 가시화한 결과는 Fig. 5와 같다. 회귀모델을 가시화하여 분석한 결과 최적설계치수를  $AR \approx 0.7$ ,  $R_{mag}ratio \approx 0.6$ ,  $h_{coil}ratio \approx 0.65$ ,  $t_0ratio \approx 0.5$ 로 결정했다.

이 연구의 유효성을 검증하기 위하여 Kim et al.<sup>[9]</sup>이 수행한 하베스터의 최적설계 결과와 비교했다. 비교 결과  $R_{mag}ratio$ ,  $h_{coil}ratio$ ,  $t_0ratio$ 는 0.6, 0.65, 0.5로 이 연구에서 결정한 최적설계 치수와 동일했다. 은 0.8로 이 연구에서 결정된 치수는 0.7가 조금 차이가 있었다. 실제 최적 을 확인하기 위하여 설계변수가  $R_{mag}ratio = 0.6$ ,  $h_{coil}ratio = 0.65$ ,  $t_0ratio = 0.5$ 일 때 AR별 하베스터의 출력전압을 유한요소해석과 미분방정식의 수치적분을 통해 계산했다. 계산 결과 AR이 약 0.7일 때 출력전압이 최대가 되는 것을 확인했다. 두

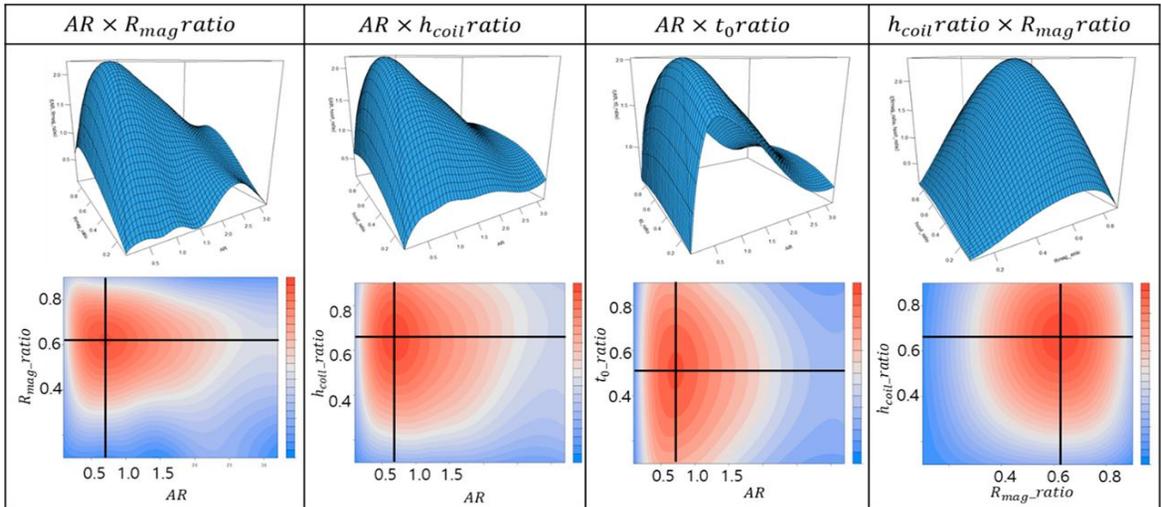


Fig. 5 Regression models estimated using ANN for AR,  $R_{magratio}$ ,  $h_{coilratio}$  and  $t_0ratio$

연구 결과가  $AR$ 에서 차이를 보인 이유는 Kim et

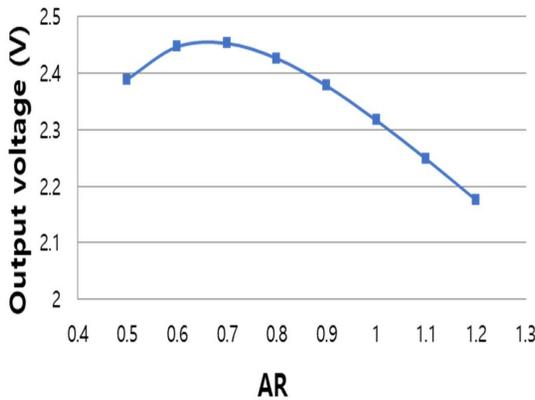


Fig. 6 Output voltage of harvester according by AR

al.이 미리 설정된  $AR$  값을 기준으로 분석한 것과는 달리 이 연구에서는  $AR$ 을 인공지능망이 학습해야 할 변수로 처리했기 때문인 것으로 판단된다.  $AR$ 에 따른 하베스터의 출력 전압 그래프는 Fig. 6과 같다.

#### 4. 결론

제조업에서 최적설계 등의 목적으로 빅데이터를

활용하기 위해서는 빅데이터를 관리하기 위한 저장소 및 이 저장소와 연동된 데이터 분석 환경을 제공하는 빅데이터 플랫폼의 구축이 필요하다. 이 연구에서는 오픈소스 기반 빅데이터 저장소인 하둡, 통계처리 프로그램인 R, 하둡과 R을 연계하는 Rhadoop을 이용하여 빅데이터 플랫폼을 구축하고 전자기형 진동 에너지 하베스터의 최적설계에 적용하였다.

최적설계 결과 하베스터의 최적 치수는  $AR \approx 0.7$ ,  $R_{magratio} \approx 0.6$ ,  $h_{coilratio} \approx 0.65$ ,  $t_0ratio \approx 0.5$ 로 결정되었다. 이 연구에서 도출한 하베스터의 최적 설계 결과는 Kim et al.<sup>[9]</sup>의 선행 연구와 동일하였다. 두 연구에서  $AR$  값은 차이를 보였다. 이와 같은 차이는 Kim et al.은  $AR$ 을 학습대상 변수로 취급하지 않고 고정된 상수로 처리하였지만 이 연구에서  $AR$ 을 학습대상 변수로 취급하여 최적화를 하여 발생된 것으로 판단된다.

#### 후 기

이 논문은 산업통상자원부 산업기술혁신사업 (과제번호: 10082580)의 지원으로 수행된 연구 결과임을 밝힙니다.

## REFERENCES

1. Laney, D., “3D Data Management: Controlling Data Volume, Velocity and Variety”, Gartner, 2001.
2. Beyer, M., “Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data”, Gartner, 2012.
3. Volvo and Teradata, “A Car Company Powered by Data”, Teradata, 2012.
4. Kerber, “Demystifying Big Data”, Tech America Foundation, 2012.
5. Hessman, “Putting Big Data to Work”, Industryweek, pp. 14-18, 2013.
6. Oh, “Prediction of Machining Performance using ANN and Training using ACO”, Journal of the Korean Society of Manufacturing Process Engineers, Vol. 16, No. 6, pp. 125-132, 2017.
7. Punuhsingon, C. S., & Oh, S. C., “Prediction of Surface Roughness and Electric Current Consumption in Turning Operation using Neural Network with Back Propagation and Particle Swarm Optimization”, Journal of the Korean Society of Manufacturing Process Engineers Vol. 14, No. 3, pp. 65-73, 2015.
8. Jeong, Y. H., “Tool Breakage Detection Using Feed Motor Current” ,Journal of the Korean Society of Manufacturing Process Engineers Vol. 14, No. 6, pp. 1-6, 2015.
9. Kim S. C., Kim Y. C, Seo J. H and Lee H. M, “Design Optimization of Electromagnetic Vibration Energy Harvesters Considering Aspect Ratio” Transactions of the Korean Society for Noise and Vibration Engineering, Vol. 27, No. 3, pp. 360~371, 2017.
10. nnet, <https://cran.r-project.org/web/packages/nnet/index.html>, 2018.
11. visreg, <https://cran.r-project.org/web/packages/visreg/index.html>, 2018.
12. Breheny P. and Burchett W., “Visualization of regression models using visreg”, R package, 1-15. 2013.