

IoT 기반의 특징 선택 기법 연구 동향

임환희⁰, 이태호^{*}, 이병준^{*}, 김경태^{*}, 윤희용^{**}

⁰성균관대학교 전자전기컴퓨터공학과

^{**}성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {lhh423, leetacho, byungjun}@skku.edu⁰, kyungtaekim76@gmail.com^{*}, youn7147@skku.edu^{**}

IoT-based Feature Selection Technique Research Trend

Hwan-Hee Lim⁰, Tae-Ho Lee^{*}, Byung-Jun Lee^{*}, Kyung-Tae Kim^{*}, Hee-Yong Youn^{**}

⁰Dept. of Electrical and Computer Engineering, Sungkyunkwan University

^{**}Dept. of Software, Sungkyunkwan University

● 요약 ●

특징 선택이란, 기계학습에서 분류 정확도를 향상시키기 위해서 많은 특징들을 분석해 가장 좋은 성능을 나타낼 수 있게끔 특징의 부분집합을 찾아내는 방법이다. 특징 선택 연구는 수십만개의 변수가 있는 데이터 세트를 이용하는 응용분야에서 주로 연구된다. 이러한 응용 분야는 주로 텍스트 처리, 유전자 배열 분석과 같은 고차원 데이터를 분석하는 분야이다. 또한, IoT 환경은 많은 데이터를 처리하기 때문에, 데이터 분류나 데이터의 가공을 위해서는 특징 선택 기법이 필수적이다. 본 논문에서는 특징 선택 기법에 대해 설명하고, IoT 환경에서 특징 선택 기법을 제안한다.

키워드: 사물인터넷(IoT), 특징 선택(feature selection), 분류(Classification)

I. Introduction

오늘날, 세계는 Big Data 시대에 접어들었습니다. IoT 기기는 이용 가능한 데이터 정보를 엄청나게 증가시키고 있으며 훨씬 저렴한 비용으로 지속적으로 생산되고 저장되고 있다[1]. 이러한 엄청난 크기의 데이터의 계산과 분류를 하려면 특징 선택 기법이 필수적이다. 특징 선택 기법은 기계학습에서 분류 정확도를 향상시키기 위해 많은 특징들을 분석해 가장 좋은 성능을 나타낼 수 있게끔 특징의 부분집합을 찾아내는 방법이다[2]. 본 논문에서는 IoT 환경에서의 적절한 특징 선택 기법을 제안한다. 본 논문 2장에서는 여러 종류의 특징 선택 기법을 설명하고, 3장에서는 IoT 환경에 적절한 특징 선택 기법을 제안한다. 마지막으로, 4장에서는 결론을 서술한다.

1.1.1 Filter methods

Filter Methods는 일반적으로 전처리 단계로 사용한다. 또한 기계 학습 알고리즘과 독립적이며, 결과 변수와의 상관관계를 분석하고 통계적인 방법으로 스코어를 매겨 높은 순서대로 특징이 선택된다.

1.1.2 Wrapper Methods

Wrapper Methods는 특징의 서브셋을 사용하며 학습을 한다. 가장 분류 정확도가 높은 특징 조합을 찾는 것이며, 특징들의 조합을 다르게 하면서 학습을 진행한다. 학습이 완료된 모델 중 가장 분류 정확도가 높은 조합을 선택한다.

1.1.3 Embedded Methods

Embedded Methods는 보통 Filter Method와 Wrapper Method의 혼합으로 구성되어 있다. 특징의 서브셋을 학습 및 생성을 같이 진행하며 최적의 특징을 선택한다. Wrapper는 모든 특징 조합이 학습을 마친 뒤 특징들을 선택하지만 Embedded는 학습과정에서 최적화된 특징을 선택한다는 점이 틀리다.

II. Preliminaries

1. Related works

1.1 Feature Selection

이번 세션에서는 특징 선택의 종류에 대해 서술하며 각 특징 선택 종류의 동작에 대해서 자세하게 서술한다. 특징 선택의 종류에는 크게 3가지로 분류하며 Filter Method, Wrapper Method, Embedded Method로 구분한다.

III. The Proposed Scheme

이번 세션에서는 Feature selection에 관해 자세히 서술하고 IoT 환경에 맞는 기법을 제안한다.

아래의 그림 1은 Filter Method의 전체적인 흐름도이다.

특징 집합 → 최적의 Subset 선택 → 분류 알고리즘 → 성능 측정

Fig. 1. Filter Method 흐름도

Filter Methods 알고리즘은 FCBF, mRMR 등이 있다. 또한, Filter Methods는 계산량이 적어 속도가 빠르지만 분류와 관련된 특징이 제거될 수 있는 단점이 있다.

아래의 그림2는 Wrapper Method의 전체적인 흐름도이다.

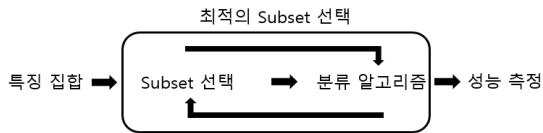


Fig. 2. Wrapper Method 흐름도

Wrapper Methods 알고리즘은 Support Vector Machines (SVM), K-nearest Neighbors (KNN), Naive Bayes (NB) 등이 있다. Wrapper Methods는 계산량이 많아 느리다는 단점이 있지만, 좀 더 정확한 특징 선택이 가능하다.

아래의 그림 3은 Embedded Methods의 전체적인 흐름도이다.

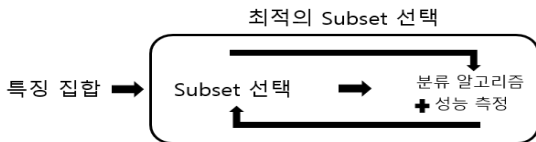


Fig. 3. Embedded Methods

Embedded Methods의 알고리즘에는 Regularized trees, Memetic algorithm, Random multinomial logit 등이 있다.

IoT 환경은 자원 Resource가 부족할 뿐만 아니라, Device의 성능도 일반 컴퓨터보다 떨어진다. 따라서 IoT 환경에서는 Wrapper Methods같이 계산량이 많은 특징 선택 기법보다는 Filter Methods와 같이 계산량이 적은 특징 선택 기법이 더 적합하다.

IV. Conclusions

본 논문에서는 특징 선택 기법들을 서술하고 IoT 환경에 적합한 특징 선택 기법을 제안하였다. IoT는 기하급수적으로 증가하고 있지만 아직까진 성능이나 자원 Resource가 많이 부족한 실정이다. 따라서 향후 연구에서는 Wrapper 기법이나 Embedded 기법을 이용해 Filter 기법의 단점을 보완하고 IoT 환경에도 적합한 알고리즘을 연구할 예정이다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신·방송연구 개발 사업(No. 2016-0-00133, 초연결 IoT 노드의 군집 지능화를 통한 Edge Computing 핵심 기술 연구), SW중심대학지원사업(2015-0-00914), 한국연구재단 기초연구사업(No.2016R1A6A3A11931385, 실시간 공공안전 서비스를 위한 소프트웨어 정의 무선 센서 네트워크 핵심기술 연구, 2017R1A2B2009095, 실시간 스트림 데이터 처리 및 Multi-connectivity를 지원하는 SDN 기반 WSN 핵심 기술 연구), 삼성전자, BK21PLUS 사업의 일환으로 수행되었음.

REFERENCES

- [1] J. Fan, F. Han, H. Liu, "Challenges of Big Data analysis", National Science Review, Vol. 1, No. 2, pp. 293-314, June 2014
- [2] Jain, A. and Zongker, D., "Feature Selection : Evaluation, Application, and Small Sample Performance", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 2, pp. 153-158, 1997