한국컴퓨터정보학회 동계학술대회 논문집 제27권 제1호 (2019. 1)

최대 엔트로피 이론 기반 네트워크 흐름 분류

김민우[O], 이태호[*], 이병준[*], 김경태[**], 윤희용[*]
[O*]성균관대학교 정보통신대학 전자전기컴퓨터공학과
[**]성균관대학교 소프트웨어대학 소프트웨어학과
e-mail: {kimmw95, leetaeho, byungjun}@skku.edu[O*], kyungtaekim76@gmail.com[**], youn7147@skku.edu[*]

Network Flow Classification Based on Maximum Entropy Theory

Min-Woo Kim[O], Tae-Ho Lee[*], Byung-Jun Lee[*], Kyung-Tae Kim[**], Hee-Yong Youn[*]
[O*]Dept. of Electrical and Computer Engineering, Sungkyunkwan University
[**]Dept. of Software, Sungkyunkwan University

● 요 약 ●

최대 엔트로피(Maximum Entropy)는 실증적 데이터에서 관찰된 잠재적인 여러 유용한 특징들을 기반으로 최대 엔트로피를 갖는 추정된 분포를 구축하기 위한 접근법이다. 본 논문에서는 네트워크상의 데이터 전송 시 혼잡한 흐름을 효율적으로 분류하기 위해 최대 엔트로피 알고리즘을 기반으로 한 새로운 네트워크 흐름 분류 모델을 제안한다. 제안한 알고리즘이 기존의 방법들 보다 높은 분류 정확도를 나타내는 것을 목표로 네트워크 서비스 시 효율성을 높이고자 한다.

키워드: 최대 엔트로피(maximum entropy), 흐름 분류(flow classification)

I. Introduction

최근 빠르고 복잡하게 발달하는 IoT 기술과 전통적인 인터넷 응용 프로그램들 뿐 아니라 P2P(Peer-to-Peer) 파일 공유, 인터넷 개인방송, 다시보기 등과 같은 다양한 플랫폼들이 생겨나며 대용량 스트리밍을 기반으로 하는 인터넷의 사용량이 급격히 증가하고 있다. 또한 이러한 추세는 앞으로 더욱더 심화 될 전망이다. 본 연구는 네트워크상의 데이터 전송 시 복잡한 전송 흐름을 효율적으로 분류하여 전송 오류를 낮추는 최대 엔트로피 모델[1]을 제안한다. 정확한 흐름 분류 모델은 실제 네트워크 흐름의 특성을 통한 분류를 통해 네트워크 모니터링에 대해 보다 높은 성능의 네트워크 관리가 가능하다. 이때 네트워크 흐름의 정확한 분류는 네트워크 성능 분석, 혼잡 관리 및 트래픽 평준화에 대해 매우 중요한 가치를 가지며 서비스의 품질을 향상시킨다. 본 논문의 나머지 부분은 다음과 같이 구성된다. 2절에서는 최대 엔트로피 이론이 적용된 관련 연구에 대해 서술하며 3절에서는 네트워크 흐름 분류 원리와 최대 엔트로피 이론을 포함하는 이 논문에서 제안된 방법을 소개한다. 마지막으로 4절에서는 현재 연구 중인 작업들을 통해 향후 연구 방향을 제시한다.

II. Preliminaries

1. Related works

1.1 최대 엔트로피 이론

최대 엔트로피 모델은 분류 문제를 해결하기 위해 기계학습에서 사용되는 방법 중 하나이다. 현재까지 자연 언어 처리에 대한 다양한 분야에서 많은 연구가 진행되어 왔으며 품사 태깅을 이용한 작업에 어휘 문맥 정보를 사전화하며 확률 분포 추정과 같은 확률 계산 등의 연구가 대표적이다[2]. 또한 최대 엔트로피 응용프로그램을 통해 다양한 분야에서 꾸준한 연구가 이루어지고 있다[3].

III. The Proposed Scheme

본 연구에서는 네트워크 흐름 분류의 포괄적인 모델을 구축하기 위해 최대 엔트로피 원리를 적용하였다. 최대 엔트로피는 사전 지식이 요구되는 기계 학습 방법이다. 일반적으로 더 많은 데이터의 학습과 학습 결과는 입력과 출력간의 관계를 정확하게 반영할 수 있어 분류 정확도가 높다. 무작위로 생성된 데이터 흐름 세트를 통해 측정된 흐름에 대해 특징들이 확률 분포로 표현되고 이에 대한 평균값을 얻는 접근법을 통해 모델을 구현하였다. 하지만 실시간적인 실제 상황에서는 대표적인 샘플을 선택해야하며 이는 시스템 성능 점수가 좋은 샘플이여야 한다. 실험은 Naive Bayes를 기반으로 비교하여

더 높은 분류 정확도를 얻을 수 있었다. 또한 분류에 대해 포괄적인 분류 기준보다 가장 많이 사용되는 특징을 기반으로 분류를 진행하였다. 무작위로 생성된 데이터 세트를 사용하여 안정적인 결과를 얻을 수 있었지만 실제 네트워크 흐름의 실험을 추가로 연구해야한다. 네트워크 흐름은 여러 요인들로 인해 변경되며 네트워크 흐름 분류의 정확성을 향상시키기 위해 불확실한 정보들을 사전에 처리하는 작업이 필요하다. Fig. 1은 네트워크 흐름의 분류 과정을 간략히 나타낸 것이다.

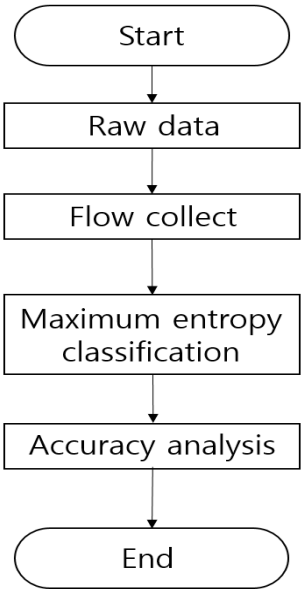

Fig. 1. 네트워크 흐름 분류 원리의 흐름도

IV. Conclusions

제안된 최대 엔트로피 모델의 효율성을 높이기 위해서는 사전에 네트워크 흐름의 특징을 추출하는 기법을 적용하여 선택된 특징들을 기준으로 흐름을 분류하는 접근법을 통하여 효율을 높이는 연구가 필요하다. 이를 위해서는 특징을 선택할 때 효율적인 특징 선택을 위한 기법이 선 연구되어야 한다. 이러한 특징 선택 기법은 대표적으로 Naive Bayes와 SVM(Support Vector Machines) 등이 있으며 향후 모델의 성능 향상을 위해 다양한 특징 선택기법을 비교할 것이다. 또한 비교를 통해 가장 좋은 성능의 접근 방법을 제안된 최대 엔트로피 모델에 적용하는 연구가 필요하다.

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), BK21PLUS 사업의 일환으로 수행되었음.

REFERENCES

[1] Hieu X. Phan, Minh L. Nguyen, S. Horiguchi, Bao T. Ho, Y. Inoguchi, “Classification with Maximum Entropy Modeling of Predictive Association Rules”, European Conference on Machine Learning(ECML), Vol. 3720, pp. 682-689, Oct. 2005

[2] Gideon Mann, Ryan McDonald, Mehryar Mohri, Nathan Silberman, Daniel D. Walker, “Efficient Large-Scale Distributed Training of Conditional Maximum Entropy Models”, Advances in Neural Information Processing Systems 22 (NIPS), 2009

[3] Yi Yin, Dan Feng, Yue Lid, Shuifang Yine and Zhan Shi, “Text prediction method based on multi-label attributes and improved maximum entropy model”, Journal of Intelligent & Fuzzy Systems, pp. 1097-1109, 2018