

빅 데이터에서 기계학습을 통한 온톨로지 생성에 관한 연구

황치곤¹ · 윤창표²

¹경민대학교 · ²경기과학기술대학교

A Study on Ontology Generation by Machine Learning in Big Data

Chi-Gon Hwang¹ · Chang-Pyo Yoon²

¹Kyungmin University · ²GyeongGi College of Science and Technology

E-mail : duck1052g@gmail.com

요 약

최근 데이터 처리를 통한 의사결정 수단으로 기계학습 기법을 도입한 개념이 많이 등장하고 있다. 기계학습은 기존의 데이터를 기반으로 학습한 결과를 이용하여 의사결정의 수단이 된다. 기술의 발전으로 생성되는 데이터는 방대하다. 이러한 데이터를 빅 데이터라 한다. 이러한 데이터에서 필요한 데이터를 추출하는 것은 중요하다. 이를 위해 본 논문에서는 온톨로지를 구축하기 위한 연관데이터를 기계학습을 통해서 추출하는 방법을 제시한다. 기계학습의 결과는 의미론적 관점에서 관계성을 부여할 수 있으며, 이것은 온톨로지에 추가됨으로써 어플리케이션의 요구에 따라 관계성을 지원할 수 있다.

ABSTRACT

Recently, the concept of machine learning has been introduced as a decision making method through data processing. Machine learning uses the results of running based on existing data as a means of decision making.

The data generated by the development of technology is vast. This data is called big data. It is important to extract the necessary data from these data. In this paper, we propose a method for extracting related data for constructing an ontology through machine learning. The results of machine learning can be given a relationship from a semantic perspective. it can be added to the ontology to support relationships depending on the needs of the application.

키워드

Ontology, Machine Learning, Semantic

1. 서 론

최근 의사결정을 위한 방법으로 기계학습을 위한 기법들이 제안되고 있다. 그리고 온톨로지는 의사결정을 위한 지식 저장소로서의 역할을 해오고 있었다.

온톨로지는 정보의 의미적 연관성을 밝힘으로써 모든 자원들 간의 유사성을 제공하고, 이질적인 문제를 해결하여 사람과 컴퓨터 간의 정보 교환을 도와주는 지식과 같다[1].

온톨로지에 저장된 지식을 공유하고 재사용할 수 있도록 하며, 메타데이터의 스키마와 같은 기능을 수행함으로써 구문론과 의미론적 모두에서 사람과 컴퓨터 간의 정보 교환을 촉진시켜 준다[2].

시맨틱 웹과 수많은 다른 어플리케이션 환경에서 온톨로지를 구성하기 위해 기계 학습 기술을

적용한 분류로써 온톨로지를 구성하는 분야도 있다[3]. 기계 학습의 분류 체계를 이용하여 개념 사이의 유사성의 의미를 측정하고, 유사성의 정의를 이용하여 개념의 공동 분포를 계산함으로써 온톨로지를 구성하고, 이를 응용 프로그램 제공하고 있다.

본 논문에서는 이러한 기계학습 기법을 통하여 빅 데이터에서 중요 요소들을 추출하여 온톨로지를 생성하는 기법에 대해 연구한다. 이를 위해 2장에서는 온톨로지를 구성하는 개념관계를 제시하고, 3장에서는 이를 위한 시스템을 구성하고, 4장 결론은 연구의 방향에 대해 서술한다.

II. 기계학습을 통한 온톨로지의 개념

본 장에서는 빅 데이터를 분석함으로써 생성될

온톨로지가 가질 기본적인 개념과 개념들의 관계에 대해 기술한다.

상위 개념을 기반으로 하위 개념으로 분할하고 각 개념의 의미에 따라 하위의 개념으로 상세화시킨다. 개념과 개념 사이에서 연관관계를 가지고 있어 개념사이의 관계를 표현해야 한다. 이에 대한 개념을 표현한 개념도가 그림 1과 같다.

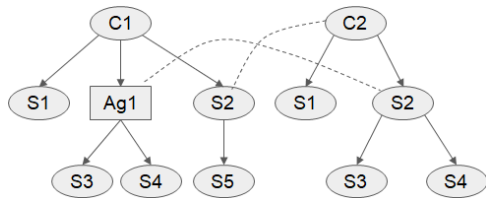


그림 1. 온톨로지 구성을 위한 개념도

그림 1은 특정 개념(C1, C2), 개념을 구성하거나 의미론적 하위의 속성을 가지는 하위개념(S1, S2, S3,...), 개념의 형식을 위한 분류(Ag) 그리고 개념들 간의 관계(점선: 관계에 대한 기술 포함)를 나타낸 것이다. 이는 개념 간에는 서로 연관성을 가지고 있으며, 개념들 간의 의미적 유사성을 표현할 수 있다.

III. 온톨로지를 위한 시스템 구성

본 장은 기계 학습을 통한 온톨로지의 생성을 위한 시스템을 구성한다. 기계 학습은 빅 데이터로부터 추출된 정보를 기계학습으로 추출 및 분류된 학습 결과를 산출한다. 이 결과는 온톨로지가 되고, 이를 통해 요구되는 어플리케이션에 제공한다. 이에 대한 시스템의 설계는 그림 2와 같다.

그림 2는 SNS와 Web에서 발생하는 빅 데이터를 수집하여 기계학습을 통하여 개념 학습을 하고 이 결과를 온톨로지에 저장하고, 이후 학습에 이 온톨로지를 빅 데이터와 같이 사용함으로써 온톨로지를 정제 할 수 있다.

이에 따른 시스템의 각 구성요소의 역할은 다음과 같다.

- **Concept Extractor** :빅 데이터로부터 요구된 도메인에 적합한 개념을 추출하여 기계학습을 위한 입력이 된다.
- **Concept Learner** :기계학습 기법으로 개념 간의 관계성을 산출한다.
- **Classification** :개념 학습으로 산출된 결과를 분류기법을 통하여 개념의 관계를 정립한다.
- **Repeater** : 다음 학습을 위해 생성된 온톨로지를 개념 학습의 입력으로 사용한다.
- **Ontology Accumulator** :학습을 통해 생성된 개념을 온톨로지에 축적한다.
- **Ontology Manager** :온톨로지의 관리와 어플리케이션의 요구를 충족을 위해 온톨로지를

관리한다.

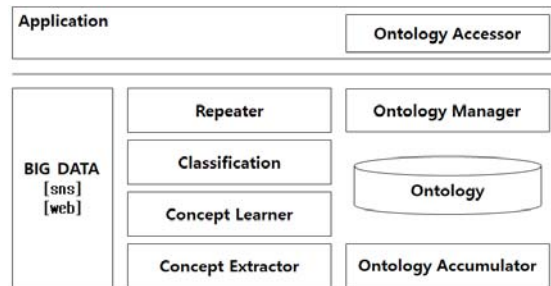


그림 2. 기계학습을 통한 온톨로지 구성을 위한 시스템 설계

IV. 결론

온톨로지는 개념 간의 연관관계를 바탕으로 하여 의미론적 관계성을 제공하여 데이터 간에 발생하는 유사성과 이질적 문제를 해결한다. 빅 데이터에서 이를 제공하기 위한 방법론으로 기계학습을 통해 온톨로지를 생성할 수 있는 방법론을 제시했다. 향후는 생성된 온톨로지를 특정 도메인에 적용하여 어플리케이션에 적용할 수 있도록 한다.

References

- [1] D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, B. Parsia, T. Payne, E. Sirin, N. Srinivasan and K. Sycara, "OWL-S: Semantic Markup for Web Services," *W3C Member Submission*, 2004.
- [2] Klusch, Matthias, Benedikt Fries, Katia Sycara. "OWLS-MX: A hybrid Semantic Web service matchmaker for OWL-S services," *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 7, No. 2, pp.121-133, 2009.
- [3] Doan, A., Madhavan, J., Domingos, P., & Halevy, A. (2004). "Ontology Matching: A Machine Learning Approach," *In Handbook on ontologies*, Springer, Berlin, Heidelberg, pp. 385-403, 2004.