

점진적 학습 기반 모아 콘텐츠 큐레이션 서비스 시스템 설계

Design of Moa Contents Curation Service System Based on Incremental Learning Technology

이 정 원*, 민 병 원**, 오 용 선**
 충남대학교대학원 컴퓨터공학과*,
 목원대학교 정보통신융합공학부**

Lee Jeong-won*, Min Byung-Won**, Oh Yong-Sun**
 Chungnam Univ. *, Mokwon Univ. **

요약

콘텐츠 큐레이션 서비스를 위해서 대용량 데이터를 학습하는 과정에서 발생하는 메모리부족 문제, 학습소요시간 문제 등을 해결하기 위한 『대용량 문서학습을 위한 동적학습 파이프라인 생성기술 중 빅데이터 마이닝을 위한 점진적 학습 모델』 기술이 필요하며, 본 논문에서 제안한 콘텐츠 큐레이션 서비스는 온라인상의 수많은 콘텐츠들 중 개인의 주관이나 관점에 따라 관련 콘텐츠들을 수집, 정리하고 편집하여 이용자와 관련이 있거나 좋아할 만한 콘텐츠를 제공하는 서비스이다. 본 논문에서 설계된 모아 큐레이션 서비스는 대용량의 문서를 학습함에 있어서 메모리 부족 문제, 학습 소요시간 문제 등을 해결하기 위해 학습데이터의 용량 제한이 없는 문서를 자유롭게 학습하고 부분적인 자질추가/변경 시에 변경요소만을 추가 반영할 수 있는 범용적이고 일반적인 분류기의 구조설계 방법 등을 제시하였다.

I. 서론

‘큐레이션(Curation)’은 미술관·박물관 등에 전시되는 작품을 기획하고 설명해 주는 ‘큐레이터’에서 파생된 말이다[1]. 최근에는 디지털 콘텐츠의 양이 기하급수적으로 증가한 상황에서, 선별된 양질의 정보에 대한 사용자 요구를 충족시키기 위해 “디지털 큐레이션”이라는 개념으로 확장하여 사용하고 있다. 큐레이션이라는 용어는 디지털 큐레이션, 데이터 큐레이션, 소셜 큐레이션, 콘텐츠 큐레이션로 대체로 네 가지로 사용되고 있다. 오늘날 큐레이션은 디지털 및 인터넷 환경을 전제로 한 것이므로 모두 디지털 큐레이션에 해당한다고도 볼 수 있다. 또 텍스트, 이미지, 영상, 음악 등 모든 종류의 데이터가 ‘콘텐츠’라는 용어로 수렴되는 지금은, 콘텐츠 큐레이션이 가장 포괄적으로 큐레이션을 대표하는 표현이라고 볼 수 있을 것이다.

정보 과잉의 시대에 의미 있고 가치 있는 정보의 획득과 공유가 중요해짐에 따라 온라인에서의 큐레이션 서비스 산업이 주목받기 시작하였고, 다양하게 존재하는 소셜 정보/지식 자원(Youtube, Google, Blog, Facebook, Twitter, Instagram 등)의 검색을 통해 유용한 정보를 큐레이팅하고 기관 내부 정보와의 융합을 통해 통합된 정보전달 경험을 제공하는 큐레이션 서비스 등장하였다.

전통적인 마이닝 기법은 다양한 디지털 매체와 센서 등에서 생산되는 빅데이터를 처리하기 어려울 뿐 아니라 신규 데이터 누적 시 전체 데이터를 재분석해야하므로 메모리부족 문제, 학습소요시간 문제 등 효율성이 떨어지는 문제점들이 발생한다. 이러한 문제점을 해결하기 위해서는 대용량 데이터가 수시로 추가되는 환경에서 데

이터 마이닝을 위해 데이터 전체를 반복하지 않고 증분 데이터만을 학습하는 기술인 점진적 학습(Incremental learning) 기술이 필수적이다.

본 논문에서 제안한 콘텐츠 큐레이션 서비스에서는 기존의 빅데이터 학습에 있어 메모리부족 문제, 학습소요시간 문제 등을 해결하기 위해 막대한 정보를 『빅데이터 마이닝을 위한 점진적 학습 모델』 기반으로 분류하고 유용한 정보를 골라내어 수집하여 다른 사람에게 새로운 가치를 생성하여 배포할 수 있는 LOD(Linked Open Data) 클라우드 서비스 플랫폼 제공으로 콘텐츠 큐레이션 서비스 산업의 영역을 확장하고자 하였다.

II. 본론

콘텐츠 큐레이션 서비스는 다양한 데이터에서 발생하는 정형·반정형·비정형 데이터를 실시간 수집하는 것부터 저장·분석·시각화까지 전체적인 작업을 통합할 수 있는 빅데이터 분석 서비스로, 방송·통신·금융·공공 등 분야에서 사용되고 있다.

콘텐츠 큐레이션 서비스를 위해서 대용량 데이터를 학습하는 과정에서 발생하는 메모리부족 문제, 학습소요시간 문제 등을 해결하기 위한 『대용량 문서학습을 위한 동적학습 파이프라인 생성기술 중 빅데이터 마이닝을 위한 점진적 학습 모델』 기술이 필요하며, 본 논문에서 제안한 콘텐츠 큐레이션 서비스는 (그림1)에서와 같이 온라인상의 수많은 콘텐츠들 중 개인의 주관이나 관점에 따라 관련 콘텐츠들을 수집, 정리하고 편집하여 이용자와 관련이 있거나 좋아할 만한 콘텐츠를 제공하는 서비스로 설계하였다.



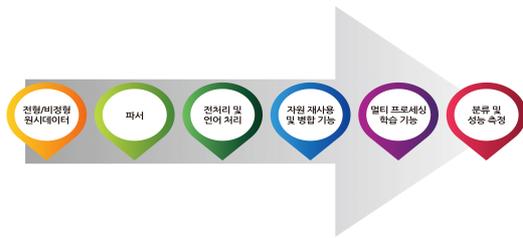
▶▶ 그림 1. 콘텐츠 큐레이션 서비스 개념

(그림2)와 같이 콘텐츠 큐레이션 서비스를 위한 인공지능형 LOD 클라우드 플랫폼의 요소 기술은 ①오픈소스 기반의 빅데이터 수집 및 전처리 기술, ②LOD 클라우드 서비스를 위한 트리플 변환 및 저장기술, ③점진적 학습 모델 기반 빅데이터 마이닝 기술, ④콘텐츠 큐레이션 LOD 클라우드 서비스 개발 및 배포 기술 등으로 구성된다.



▶▶ 그림 2. 콘텐츠 큐레이션 서비스시스템 구성도

본 논문에서 제안하는 점진적 학습 점진적 학습은 데이터가 동적으로 추가되는 환경에서 전체 데이터를 반복적으로 학습하지 않고 증분된 데이터만을 학습하여 기존의 학습 결과와 통합할 수 있는 학습 방법이다. 모델에 기반한 빅데이터 학습기술은 대용량의 문서를 학습함에 있어 메모리부족 문제, 학습 소요시간 문제 등을 해결하기 위해 널리 사용하는 자질축소 기법에 의존하지 않고 대량의 문서를 자유롭게 학습하고 부분적인 자질추가 변경 시에 변경요소만을 추가 반영할 수 있는, 범용적이고 일반적인 분류기의 구조설계 방법에 관한 것이다. 이러한 점진적인 학습을 수행하기 위해 제안 분류기의 단위 컴포넌트 생성방법과 이를 동적으로 결합하는 방법에 대한 핵심사항은 (그림3)과 같다.



▶▶ 그림 3. 점진적 학습 작업관리 워크플로우

III. 결론

4차 산업혁명에는 IoT(Internet of Things) 환경에서 생성되는 다양한 데이터를 처리하기 위한 클라우드 컴퓨팅 및 빅데이터 산업이 발달하고 일련의 혁신적인 기술들은 인공지능(AI)이 더해지며 삶의 변화를 이끌어내고 있으며, 빅데이터를 활용한 산업이 4차 산업혁명의 한 축으로 성장할 것이란 전망이 제기된 가운데, 방대한 양의 데이터를 수집하고 분석해 가치 있는 정보를 활용하는 인공지능형 빅데이터는 4차 산업혁명의 핵심 기술이다.

콘텐츠 큐레이션 서비스를 위해서 대용량 데이터를 학습하는 과정에서 발생하는 메모리부족 문제, 학습소요시간 문제 등을 해결하기 위한 『대용량 문서학습을 위한 동적학습 파이프라인 생성기술 중 빅데이터 마이닝을 위한 점진적 학습 모델』 기술이 필요하며, 본 논문에서 제안한 콘텐츠 큐레이션 서비스는 온라인상의 수많은 콘텐츠들 중 개인의 주관이나 관점에 따라 관련 콘텐츠들을 수집, 정리하고 편집하여 이용자와 관련이 있거나 좋아할 만한 콘텐츠를 제공하는 서비스이다.

큐레이션 서비스에서는 개인비서, 금융 분야의 투자, 자율주행, 저널리즘, 효율적인 업무 지시/감독, 제조업의 자동화 공정, 교육, 콘텐츠 유통, 학습정보 등에서 컴퓨터가 방대한 양의 데이터로 부터 학습하여 사람의 일을 대신 처리하거나 의사결정에 도움을 줌으로써 업무의 효율을 높여주는 서비스 산업에 활용이 가능하다.

참고 문헌

- [1] <http://terms.naver.com/entry.nhn?docId=1691524&cid=42171&categoryId=42190>
- [2] R. Burke, "Hybrid Recommender Systems:Survey and Experiments," User Modeling and User-Adapted Interaction, Vol.12, No.4, pp.331-370,2002(11).
- [3] Francesco Ricci, Recommender Systems Handbook, Springer, 2011.
- [4] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.6, pp734-749,2005.
- [5] R. Plutchik, "The Nature of Emotions," American Scientist, Vol.89, No.4, pp.344-350, 2001.
- [6] J. A. Russell, J. M. Fernandez-Dols, A. S. R. Masted, and J. C. Wellenkamp, Everyday Conceptions of Emotion: An introduction to the Psychology, Anthropology and Linguistics of Emotion, Kluwer Academic Publishers, 1995.