

불리언 질의어 재구성에서 의사결정 나무의 점진적 학습 성능 분석

Performance Analysis of Boolean Query Reformulation using Decision Tree with Incremental Learning

윤정미, 권영식(동국대학교) 김남호(선문대학교)

Abstract

인터넷등에 의한 정보검색환경의 변화로 많은 양의 정보 취득이 용이해짐에 따라 사용자가 원하는 정보를 효과적으로 검색하는 것이 중요한 문제로 제기되고 있다. 정보검색시스템을 사용하는 사용자는 불명확한 정보요구만을 갖고 여러번의 시행착오를 거쳐 질의를 만들어 나간다. 그러나 이러한 모든 과정에는 사용자가 질의를 집적 재구성해야하는 어려움, 사용자와 시스템간의 용이 불일치 등 많은 문제점이 내재되어 있다.

기존연구로는 시소리스를 이용하여 질의를 확장하거나 문헌이나 질의어에 가중치를 부여하여 보다 적합한 문헌을 검출하는 방법이 제기되어 있다. 이런 방법은 보다 적합한 용어를 선택할 수 있지만 사용자가 직접 질의어를 구성해야 하므로 초보자가 초기에 적절한 질의어를 구성해야 하므로 초보자가 초기에 적절한 질의어를 구성하는 것이 어려운 단점이 있다. 최근에는 하이퍼텍스트를 이용한 네비게이션을 통해 사용자가 원하는 정보를 찾는 연구도 행해지고 있지만, 네비게이션 과정에서 발생하는 인지적 과부담과 정보에 접근하는 과정이 길다는 단점이 있다.

이런 문제를 해결하기 위해서는 사용자가 확인하는 바람직한 혹은 바람직하지 않은 정보들을 기초로 질의 부분을 학습하여 사용자에게 편리하고 효율적인 지능적 질의어 재구성 기법을 제공하는 것이 바람직하다. 본 연구에서는 질의어 재구성을 바람직한(바람직하지 않은) 문서를 분류하는 분류문제로 보고 귀납적 학습 알고리즘인 ID5R은 기본적으로 ID3와 비슷하나 처리할 예제가 순차적으로 입력될 경우 그들을 단계적으로 학습할 수 있는 점진적 학습능력을 갖고 있다. 정보검색문제에서는 원하는 정보를 단 한번에 검출하기 어렵기 때문에 여러번에 걸쳐 정보를 검색하게 된다. 이 경우 사용자의 피드백을 점진적으로 학습할 수 있는 ID5R 알고리즘이 효과적이다.

실험은 불리언 정보검색 시스템 하에서 사용자의 피드백에 의한 예제를 중심으로 ID5R 알고리즘을 이용하여 질의어를 자동으로 재구성하고 사용자의 피드백인 새로운 예제를 대표하는 용어가 기존의 예제에서 존재하지 않을 경우 그 용어를 처리하는 방법에 대해 행해지고, 그 결과를 다른 귀납적 학습알고리즘과 비교했다. 이 연구를 통하여 사용자의 적합성 피드백 과정을 학습하는 적응적 지능형 정보검색시스템을 구현함으로써 시스템을 이용하는데 있어서 그 편의성의 증가와 보다 적합한 문서 검출을 기대할 수 있다.