



즉시 보도(2016. 12. 20.)

문의 : 담당자(정진홍) 연락처(02-880-7263)
 연구책임자 : 강 유 교수(02-880-7254) / 교신저자
 연구진 : 정진홍, 박남용, 한국뉴욕주립대 이슬 교수 / 공동저자

대용량 그래프에서 랜덤 워크 기반 고속 랭킹 기법 개발 - 실세계 그래프의 구조적 특징을 활용 -

- 서울대학교 컴퓨터공학부 강 유 교수 연구진은 소셜 네트워크, 웹, 그래프 기반 추천 시스템 등의 랭킹에 널리 활용되는 RWR (Random Walk with Restart) 알고리즘이 대용량 그래프를 빠르게 처리하지 못하는 문제를 해결하였다.
- RWR은 그래프에서 랜덤 워크 (Random Walk)를 통해 한 정점을 기준으로 다른 정점에 대한 중요도를 계산하는 랭킹 기법으로 개인화된 추천에 적합하여 정점 랭킹, 추천 시스템, 링크 예측 등의 다양한 그래프 마이닝 응용에 활발하게 활용되고 있다. 예를 들어 소셜 네트워크에서 특정 사용자와 다른 모든 사용자간의 유사도를 계산하여 유사도가 높지만 아직 연결이 안 된 사용자들을 찾음으로써 그 사용자에게 새로운 친구를 추천하는데 쓰이고 있다.
- 기존의 RWR을 계산하는 알고리즘은 반복적 기법과 전처리 기법으로 분류되는데 반복적 기법은 메모리를 적게 사용하여 대규모 그래프를 처리할 수는 있으나, RWR의 계산 속도가 느리다는 단점이 있다. RWR 계산 속도를 빠르게 하기 위해 여러 전처리 기법이 제안되었지만 전처리 기법은 메모리를 과도하게 사용하기 때문에 대용량 그래프를 처리하지 못한다.
- 강 유 교수 연구진은 실세계 그래프의 구조적 특징을 이용하고 반복적 기법과 전처리 기법의 조합으로 기존 방법들의 장점을 취해 대용량 그래프에서 빠르고 메모리 효율적인 RWR 알고리즘을 개발하였다. 논문에서 제안한 알고리즘은 기존 알고리즘보다 100배 이상 큰 그래프를 처리할 수 있고 130배 이상 메모리를 적게 사용하

면서 실행 시간은 9배 빠르되, 같은 정확도를 제공하도록 설계되었다.

- 본 연구결과를 통해 RWR을 기반으로 하는 응용의 계산 확장성과 속도를 크게 향상시킬 수 있기 때문에 그래프를 바탕으로 RWR을 활용하는 그래프 마이닝, 정보 검색, 인공지능, 생물정보학 등 다양한 분야에 도움이 될 것으로 기대한다.

- 본 연구는 서울대학교 컴퓨터공학부 강유 교수 연구진이 주도하였으며 한국뉴욕주립대 이슬 교수도 참여했다.

- 이번 연구 결과는 데이터베이스와 빅데이터 분야에서 세계 최고 학회인 ACM SIGMOD 2017 (ACM International Conference on Management of Data)에 채택되어 2017년 5월 중순 미국에서 발표될 예정이다. (<http://sigmod2017.org/>)

- [붙임] 1. 연구결과 2. 용어설명 3. 그림설명
4. 연구진 이력사항

연구 결과

BePI: Fast and Memory Efficient Method for Billion-Scale Random Walk with Restart

Jinhong Jung, Namyong Park, Lee Sael, and U Kang

(ACM International Conference on Management of Data (SIGMOD) 2017)

본 논문은 소셜 네트워크, 웹과 같은 대용량 그래프에서 특정 정점과 다른 정점들간의 유사도를 계산하는 랭킹 기법인 RWR (Random Walk with Restart)의 계산 확장성을 높이고 속도를 빠르게 하는 알고리즘을 제안한다. RWR은 널리 쓰이는 랭킹 기법으로 개인화된 추천에 적합하여 친구 추천, 아이템 추천 등의 다양한 응용에 활발하게 활용되고 있다. 제안하는 알고리즘은 실세계 그래프 노드 차수의 비대칭성과 같은 구조적 특징을 활용하고, 행렬 연산 최적화를 통해 대용량 그래프에서 RWR을 효율적으로 계산한다. 논문에서 제안하는 알고리즘은 기존 알고리즘보다 100배 이상 큰 그래프를 처리하고 130배 이상 메모리를 적게 사용하면서 실행 시간은 9배 빠르되 같은 정확도를 제공하도록 설계되었다. 본 연구 결과는 그래프 마이닝, 정보 검색, 인공지능, 생물정보학 등 그래프를 바탕으로 RWR을 활용하는 다양한 분야에 도움이 될 것으로 기대한다.

용 어 설 명

1. Random Walk with Restart (RWR)

- RWR은 그래프에서 한 정점과 다른 정점 간의 근접도(중요도, 유사도)를 계산하는 그래프 랭킹 모델로 그래프에서 개인화된 랭킹을 구하기 위해 널리 사용된다. RWR에서는 랜덤 서퍼가 시작 정점에서 출발하여 무작위로 이동하면서 근접도를 계산하며 현재 시점에서 임의의 이웃 정점으로 이동(Random Walk) 하거나 시작 정점으로 다시 돌아가는 행위(Restart)를 반복한다. RWR을 반복적으로 수행하면 랜덤 서퍼가 각 정점을 방문할 확률이 정해지며 이 확률이 시작 정점과 다른 정점 간의 근접도가 된다.

그림 설명

실세계 그래프의 구조적 특징을 활용

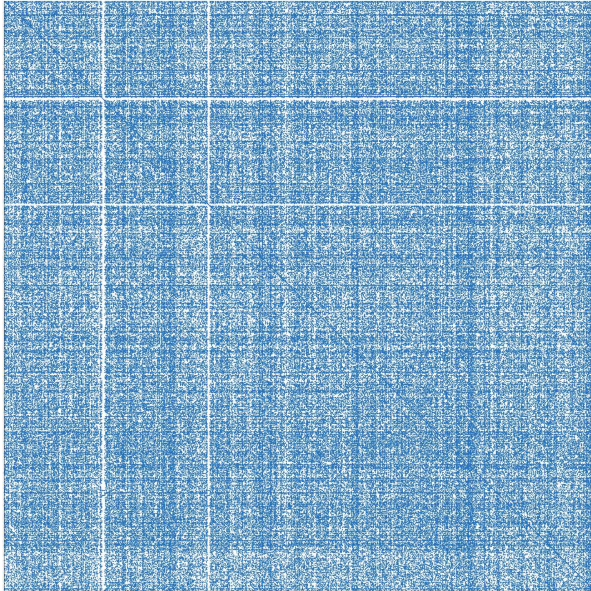


그림 1. 입력 그래프의 정점 재정렬 전의 행렬

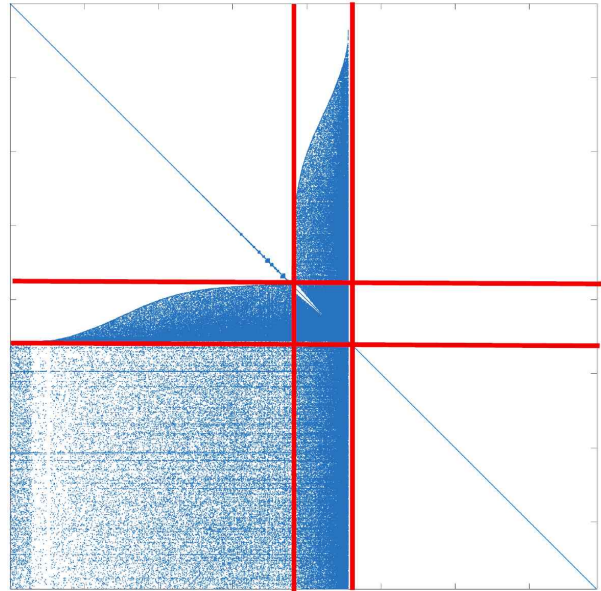


그림 2. Deadend 구조와 hub-and-spoke 구조를 이용하여 정점 재정렬 후의 행렬

본 논문에서는 대용량 그래프에서 효율적인 랭킹 계산을 위해 실세계 그래프의 구조적 특징으로 deadend 구조와 hub-and-spoke 구조를 활용하였다. deadend란 웹 문서 네트워크와 같은 그래프에서 파일이나 이미지 같이 나가는 간선이 없는 정점을 의미하며 실세계 그래프가 deadend 노드를 많이 가지는 특징을 deadend 구조라 한다. Hub-and-spoke 구조는 실세계 그래프가 소수의 연결이 많은 정점들(hub)과 다수의 연결이 적은 정점들(spoke)으로 구성되는 구조를 일컫는다. 이 두 가지 실세계 그래프의 구조적 특징을 활용하여 입력 그래프의 인접 행렬을 기반으로 한 행렬 (그림 1)에 정점 재정렬을 수행하면 그림 2와 같은 모양을 띄게 된다. 본 논문에서 제안하는 알고리즘은 효율적인 RWR 계산을 위해 그림 2와 같은 행렬을 부분 행렬로 나누어 RWR 계산에 필요한 전처리를 수행한다.

연구자 이력사항

1. 인적사항

- 소 속 : 서울대학교 컴퓨터공학부 조교수
- 전 화 : 02-880-7254
- E-mail : ukang@snu.ac.kr



2. 학력

- 1996 - 2003 서울대학교 학사
- 2007 - 2009 Carnegie Mellon University 석사
- 2009 - 2012 Carnegie Mellon University 박사

3. 경력사항

- 2004.1 - 2007.7 KT 연구소 연구원
- 2012.6 - 2012.12 Carnegie Mellon University Postdoctoral Fellow
- 2013.1 - 2015.8 KAIST 전산학과 조교수
- 2015.9 - 현재 서울대학교 컴퓨터공학부 조교수

4. 기타 정보

- SIGKDD Doctorial Dissertation Award (2013)
- New Faculty Award from Microsoft Research Asia (2013)