<Chater 1 : Collabortative recommendation Summary>

2011-11703

컴퓨터공학과 박태헌

1. summary

해당 chapter에서는 제목과 같이 collaborative recommendation의 종류와 적용에 대하여 설명하고 있다. user의 similar preference를 중심으로 설계한 user-based recommendation은 Pearson’s correlation coeffient를 통해 수식적으로 보일 수 있다. 하지만 user가 늘어나게 되면 현실적으로 computing이 힘들어져 즉각적으로 보여주기가 힘들다. 따라서 threshold를 두는 등 각종 해결 방안이 있지만 이에도 분명 한계가 존재한다.

Item-based recommendation은 상대적으로 item은 한정되어 있기 때문에 user-based recoomendation보다 수월한 계산이 가능하다. Cosine similarity를 통해 이를 보일 수 있으며, precomputing 외에 다양한 방식으로 real-world에서 적용의 가능성을 확인할 수 있다. 참고로 memory상에서 즉각적으로 computing하는 것은 memory-based, precomputing 혹은 learning 방식은 model-based라 부른다. user-based recommendation은 memory-based에, item-based recommendation은 model-based에 포함된다.

Data mining에 있어 항상 문제가 되는 것 중 일부가 바로 data sparsity와 cold-start이다. 이를 해결하기 위한 대책으로 n-step path를 활용한다. 연역적 방법으로 보면 된다.

더 나은 model-based로 matrix를 이용하는 방식이 있다. SVD의 경우 n\*n matrix를 n\*1, 1\*n, 그리고 singular values matrix로 나누어 predition에 접근한다. 나누어진 matrix를 2-demension에 표현하여 시각적으로 classify가 가능하다. Eigenvector를 이용한 PCA방식도 존재한다.

위의 방식과 더불어 association rule이 있다. 관계를 분석하여 하나의 rule(혹은 함수로 봐도 좋을 것 같다.)로 지정하여 input을 넣으면 output을 뱉는 방식으로 rule을 정하는 단계는 precomputing, input을 넣고 output을 뱉는 단계는 실시간으로 처리할 수 있다. rule을 정하는데 지표가 되는 것으로 support와 confidence가 있다.

또한 확률론적으로 접근하는 방식이 있다.Bayes’ Theorem과 조건부 확률을 사용하여 계산한다. 좀 더 나은 설계를 위해 expectation maximization algorithm등을 사용하기도 한다.

이전까지 소개한 방식은 이전부터 많이 연구되어 온 방식으로 대부분 accuracy와 complexity는 비례 관계라면, 최근에 소개된 방식인 slope one predictors는 굉장히 단간하면서 accuracy 또한 훌륭하다.

이러한 recommendation의 완성체로 google news personalization engine을 소개한다.

2. 질문

내용 이해에는 전반적으로 문제가 없었으나 각종 수식에 대한 자세한 설명이 없어 이해할 수가 없습니다. 이에 대한 추가적인 설명을 원합니다. 특히 SVD의 경우 U와 V의 앞 두 column을 자르는데 대충 느끼기엔 computing cost를 줄이기 위함인것 같지만, 과연 column을 절반이라 잘라낸 matrix로의 표현이 정확성 면에서 문제가 없는지 알고 싶습니다.

association rule에서 confidence는 어느 정도 이해가 가는 반면, support는 감이 잘 안 옵니다.