

협력적 추천을 위한 효율적인 통합 방법

도영아*, 김종수, 류정우, 김명원
송실대학교 컴퓨터학과

npng1@naver.com, kjongsu@orgio.net, ryu0914@orgio.net, mkim@computing.soongsil.ac.kr

Efficient Combining Methods for a Collaborative Recommendation

YoungA Do, JongSu Kim, JungWoo Ryu, Myungwon Kim
School of Computing, Soongsil University

요약

신경망을 이용한 추천 기술은 항목이나 사용자간의 가중치를 학습할 수 있고, 자료 유형에 상관없이 데이터 처리가 용이하다. 또한 최근 연구를 통해서 그 우수성이 입증되고 있다. 그러나 사용자간의 상관관계로 추천하는 사용자 신경망 모델과 항목간의 상관관계로 추천하는 항목 신경망 모델이 서로 다른 관점으로 다른 선호도를 제시할 경우에 선택한 모델의 선호도에 따라 시스템의 성능이 좌우된다. 그러므로 효율적이고 성능이 우수한 추천 시스템을 위해 사용자와 항목 신경망 모델의 통합 방법을 제안한다. 두 모델 사이에 우선 순위를 결정하여 통합하는 순차적 통합 방법과 두 모델을 동시에 고려하는 병렬적 통합 방법을 제안한다. 그러나 두 통합 방법은 선호도 예측 기준에 있어서 정적이고, 문제에 대한 적용성이 없다. 그러므로 신경망(퍼셉트론, 다층 퍼셉트론)을 이용한 통합 방법을 제안한다. 또한 퍼지의 소속함수를 이용하여 퍼지 추론을 적용한 통합 방법을 제안하고, 패턴 인식 분야에서 사용하는 BKS 방법을 적용하여 두 신경망 모델을 통합하여 실험한다. 본 논문에서는 사용자와 항목 신경망 모델을 통합함으로써 기존의 추천 기술인 연관 규칙과 단일 신경망 모델을 이용한 추천보다 우수함을 보이고 있다.

1. 서론

최근에는 사용자 개인의 취향이나 특성에 맞는 정보를 자동으로 찾거나 추천해 주는 개인화(personalization) 기술이 크게 요구되고 있다[1][2]. 추천 시스템은 정보나 서비스의 개인화 된 추천을 위한 지식을 발견하고, 개인의 특성에 맞춘 서비스를 통해서 현재의 인터넷이 갖고 있는 정보 탐색의 비효율성 문제를 해결할 수 있는 새로운 해결책으로 주목받고 있다. 그러므로 개인화 된 효율적인 추천을 위해서 [4]에서는 연관 규칙을 이용한 추천 기술[3]과 신경망을 이용한 추천 기술의 성능을 비교함으로써, 신경망을 이용한 추천 기술이 더욱 향상된 추천 기술임을 제시한다. 그러나 사용자 또는 항목 신경망 모델만 추천함으로써 두 신경망 모델이 서로 다른 선호도를 제시할 경우, 어느 신경망 모델의 선호도 결과로 추천해야 하는지에 따라서 추천 성능이 좌우된다. 본 논문에서는 두 신경망 모델이 각각의 다른 특징과 다른 관점으로 예측한 선호도를 통합하는 방법들을 제안한다. 제안하는 통합 방법으로는 두 신경망 모델 중 우선 순위를 적용하는 순차적 통합 방법, 동시에 고려하는 병렬적 통합 방법, 두 모델의 출력값을 입력으로 하는 신경망(퍼셉트론, 다층 퍼셉트론)을 이용한 통합 방법, 퍼지 추론을 이용한 통합 방법을 제안한다. 또한 패턴 인식에서 쓰이는 CME 방법 중의 하나인 BKS방법을 통합 방법으로 적용한다.

본 논문은 다음과 같은 순서로 구성되어 있다. 2절에서는 패턴 인식 분야에서 쓰이는 기존의 통합 방법과 추천 기술로써 연관 규칙에서 사용하는 통합 방법을 기술한다. 3절에서는 사용자와 항목 신경망 모델에 대해 기술하고, 4절에서는 두 신경망 모델을 통합하는 방법들을 기술한다. 5절에서는 실험 데이터와 결과를 제시하고, 6절에서는 결론을 맺고, 향후연구를 제시한다.

2. 기존의 통합 방법

CME(Combination of Multiple Experts)는 패턴 인식 분야에서 사용하는 방법으로 다른 방법론이나 다른 특성을 갖고 있는 전문가들

(classifiers 혹은 experts)이 서로의 의견을 통합하기 위해 사용하는 방법이다. CME 방법을 이용함으로써 집단의 의사 결정이나 문제 해결에 있어서 높은 성능 향상과 더불어 오차를 급속히 줄일 수 있다[5].

2.1 Voting [6]

다른 전문가들의 의견을 통합하기 위해 사용되는 대표적이고, 간단한 방법으로 voting by majority 방법이 있다. 이 방법은 반수 이상의 전문가들이 같은 의견으로 동의하면 그 의견으로 집단의 전체 의사를 결정하는 방법이다. 또한 conservative voting rule 방법은 같은 의견으로 모든 전문가들이 동의할 경우, 그 의견으로 의사 결정을 하는 방법이다. 또다른 방법인 comparative majority-voting 방법은 가장 많이 동의한 의견과 두 번째로 많이 동의한 의견의 전문가 수 사이의 값이 주어진 임계치보다 클경우에 가장 많이 동의한 의견이 집단의 전체 의사 결정으로 반영되는 방법이다.

2.2 The Borda Count Method [7]

이 방법은 1770년에 Jean-Charles de Borda에 의해 제시된 방법으로 전문가들의 의견을 통합하는 문제에 쉽게 적용되는 방법중의 하나이다 [8]. 전문가들이 의견을 제시할 때 우선 순위를 매겨서 제시하는 방법으로 가장 높은 우선 순위를 갖는 의견은 가장 높은 가중치를 줌으로써 집단의 의사를 결정한다.

2.3 Behavior Knowledge Space(BKS) [5]

2.1과 2.2에서 기술한 CME 방법은 각 전문가들을 동등하게 취급하고, 한 전문가가 잘못된 의견을 제시해도 모든 전문가의 의견을 통합하는 데에는 큰 영향을 미치지 않는다. 그 이유는 CME방법이 전문가의 독립성을 가정하고 있기 때문이다. BKS에서는 이러한 가정을 두지 않기 위해 모든 전문가들의 의견을 동시에 수용하는 지식 공간(knowledge space)으로부터 의사 결정을 도출한다. 지식 공간에 전문가의 모든 의사 결정을 수용하기 때문에 이러한 방법을 행동 지식 공간(Behavior-Knowledge Space)이라고 부른다. 간단하게 말해서 BKS는 행동 지식 공간으로부터 그 집단의 최종 결정을 내린다. 자세한 방법은 4.5에서 다루기로 한다.

본 연구는 뇌 연구 개발사업
(과제번호 : 98-J04-01-01-A-04)의 지원을 받았다.

2.4 순차적 통합 방법(Sequential Method)

순차적 통합 방법은 전문가들에게 우선 순위를 주어 의견을 순차적으로 적용하는 방법이다. [3]에서는 사용자 연관 규칙과 항목 연관 규칙을 통합하는 방법으로 임계치를 이용한 통합 방법을 기술한다. 사용자 연관 규칙의 최소 지지도가 임계치보다 크면 사용자 연관 규칙을 사용하고, 작으면 항목 연관 규칙을 사용하는 방법이다. 본 논문에서 적용하는 순차적 통합 방법은 4.1에서 자세하게 다루기로 한다.

Voting방법과 The Borda Count방법은 둘 이상의 전문가가 필요하므로 본 논문에서 통합하려는 방법에 적용할 수가 없다. 그러므로 사용자와 항목 신경망 모델을 통합하는데 있어서 기존의 통합 방법 중에서는 BKS와 순차적 통합 방법을 적용해 본다.

3. 신경망을 이용한 추천 [4]

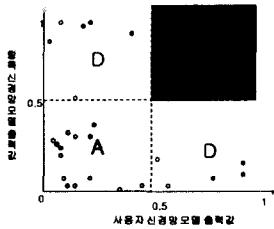
[4]에서는 항목들간 또는 사용자들간의 선호 상관관계를 신경망으로 학습시킴으로써 모델을 생성하고, 그 모델을 사용하여 선호도를 예측한다.

3.1 사용자 신경망 모델

사용자 신경망 모델은 사용자들간의 연관성으로 항목의 선호도를 예측한다.

3.2 항목 신경망 모델

항목 신경망 모델은 항목들간의 연관성으로 사용자에게 추천할 항목의 선호도를 예측한다.



<그림 1> 사용자와 항목신경망 모델의 선호도 분포

4. 사용자와 항목 신경망 모델의 통합 방법

[4]에서는 단일 신경망 모델로만 추천함으로써, 다음과 같은 문제가 발생한다. 사용자와 항목 신경망 모델이 한 항목에 대해 동시에 같은 선호도로 동의하면 즉, <그림 1>의 A영역인 경우에는 어느 모델의 선호도 결과에 따라도 상관없다. 그러나 두 모델이 서로 다른 선호도로 동의하지 않을 때 <그림 1>의 D영역일 경우, 선택한 모델의 결과에 따라 성능이 좌우된다. 그러므로 D영역은 문제에 따라서 어떻게 선호도 영역으로 분할하느냐에 따라서 추천 성능에 영향을 미친다. 따라서, 서로 다른 관점을 통해 예측된 선호도를 통합하여 항목에 대한 추천이 이루어지면, 단일 신경망 모델의 추천이나 기존의 추천 기술보다 유연성과 효율성 모두 향상될 것이다.

본 논문에서는 이러한 점을 이용하여 사용자와 항목 신경망 모델을 통합하는 방법들을 제안한다.

4.1 순차적 통합 방법(Sequential Method)

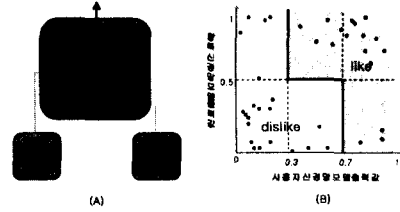
순차적 통합 방법은 어떤 한 모델을 우선적으로 채택하여 선호도를 결정하되 확실한 결론을 내리기 어려울 때에는 제 2 모델의 결과로 선호도를 판정하는 방법이다. 본 실험에서는 사용자 신경망 모델의 실제 출력값이 0.3보다 작거나 0.7보다 크면 사용자 신경망 모델의 결과로 선호도를 판정하고, 그렇지 않으면 항목 신경망 모델의 결과로 선호도를 판정한다. <그림 2>의 (B)는 순차적 통합 방법을 사용할 경우 입력 공간에서 선호도의 분할 영역을 보여주고 있다.

4.2 병렬적 통합 방법(Parallel Method)

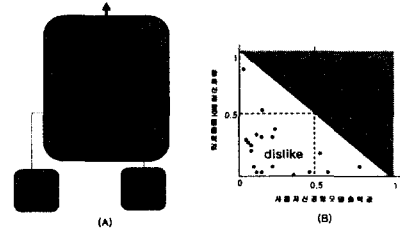
병렬적 통합 방법은 두 모델의 결과를 동시에 고려하여 최종 결론을 내리는 방법이다. 예를 들어, 어떤 항목에 대해 사용자 신경망 모델의 선호도가 0.9이고, 항목 신경망 모델의 선호도가 0.2일 경우 사용자 신

경망 모델이 긍정적인 선호도에 더 가깝기 때문에 그 항목을 사용자에게 추천한다.

<그림 3>의 (B)는 병렬적 통합 방법을 사용할 경우 입력 공간에서 선호도의 분할 영역을 보여주고 있다.



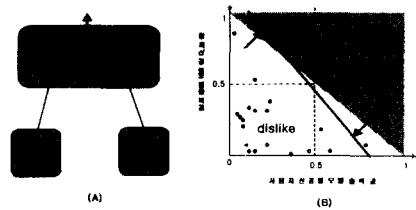
<그림 3> 순차적 통합 방법과 선호도 경계선



<그림 4> 병렬적 통합 방법과 선호도 경계선

4.3 신경망을 이용한 통합 방법

4.1과 4.2에서 제시한 순차적 통합 방법과 병렬적 통합 방법은 선호도 예측 기준에 있어서 정적이고, 문제에 대해 적응성이 없다. 따라서 사용자와 항목 신경망 모델의 출력을 입력으로 하는 신경망 모델의 학습 과정을 적용하여 보다 유연하고 효율적인 통합 방법을 제안한다.



<그림 5> 퍼셉트론을 이용한 통합 방법과 선호도 경계선

4.3.1 퍼셉트론(Perceptron)을 이용한 통합 방법

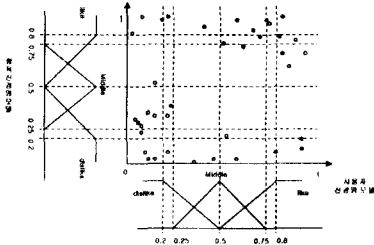
퍼셉트론은 선형 분리 가능한 패턴들만을 분류할 수가 있다. 선형 분리 가능한 데이터 클래스가 하나의 직선에 의해 두 개의 영역으로 나뉘어지는 것을 말한다. <그림 4>의 (B)는 입력 공간에서 선호도의 분할 영역이 퍼셉트론에 의해 입력 데이터를 선형 분리하는 것을 보여주고 있다. 다시 말하면, 두 모델의 출력값에 따라 적절히 선형 분리함으로써 통합 모델에 의한 정확도를 향상시킬 수 있다. 그러나 선형 분리 가능한 데이터만을 분리할 수 있으므로 복잡한 경계선은 분리할 수 없다.

4.3.2 다층 퍼셉트론(MLP)을 이용한 통합 방법

문제에 대한 적응성을 높이기 위해 사용자와 항목 신경망 모델의 출력을 입력으로 하는 다층 퍼셉트론(Multi-Layer Perceptron)을 이용한 통합 방법을 제안한다. 다층 퍼셉트론은 임의의 복잡한 경계선을 근사할 수 있으므로 보다 성능이 향상된 통합 효과를 기대할 수 있다.

4.4 퍼지 추론을 이용한 통합 방법

4.3에서 제안하는 신경망을 이용한 통합 방법은 두 모델의 결과 값을 바탕으로 학습을 통해 최종 결과 값의 정확도를 향상시키는 반면, 퍼지 규칙 방법은 퍼지 추론을 통해 정확도를 향상시킬 수 있다. 본 논문에서는 퍼지 소속 함수로 각 신경망 모델의 출력에 대해 세 개의 퍼지 집합(Dislike, Middle, Like)으로 나타내며, 각 모델의 출력값의 영역이 <그림 5>과 같이 분할됨을 알 수 있다. 각 영역에 대응하는 퍼지 규칙을 생성하여 퍼지 추론을 통하여 최종 결과 값을 산출함으로써 정확도를 향상시킬 수 있다.



<그림 6> 퍼지규칙에 의한 분할 영역

4.5 BKS를 이용한 통합 방법

BKS는 행동 지식 공간인 K 차원의 공간을 형성한다. 각 차원은 전문가로 대응되고, 각 차원은 다시 전문가가 결정할 수 있는 의견(클래스)들로 분할되어 작은 셀(cell 또는 unit)들이 형성된다[5][9]. 각 셀들은 다음과 같은 세 가지 형태의 데이터 정보를 가지게 된다. 셀에 들어온 입력 데이터 수, 그 셀에서 대표되는 의견과 셀의 입력 데이터 중에서 각 클래스에 속하는 입력 데이터의 수에 대한 정보를 갖게 된다. 테스트 데이터가 들어오면 전문가들은 이러한 셀의 데이터 정보를 이용하여 즉, 행동 지식 공간의 정보를 이용하여 의견을 통합 도출하게 된다.

(식 1)에서 $R_{e(1), \dots, e(K)}$ 는 한 셀에서의 대표 의견을 나타낸다. $e(K)$ 는 K 전문가를 나타내고, 한 셀에서 $n_{e(1), \dots, e(K)}(m)$ 는 의견 m에 속하는 입력 데이터 수를 나타낸다. 즉, 그 셀에서 가장 많이 차지하는 의견 j가 그 셀을 대표하는 의견이 된다.

$$R_{e(1), \dots, e(K)} = \left\{ j \mid n_{e(1), \dots, e(K)}(j) = \max_{1 \leq m \leq M} n_{e(1), \dots, e(K)}(m) \right\} \quad (식 1)$$

5. 실험

EachMovie[10] 데이터는 72,916명의 사용자와 1,628편의 영화로 구성되어 있다. 사용자가 본 영화에 대한 선호도(rating) 정보는 0 (Dislike), 0.2, 0.4, 0.6, 0.8, 1.0 (Like)의 수치로 표현되어져 있고, 각 사용자의 나이와 성별, 각 영화의 장르 정보가 포함되어 있다. 본 실험을 위해서 영화 100편 이상 본 사용자 1000명은 학습 데이터로, 사용자 아이디가 70000번 이상인 사용자중 영화 100편 이상 본 사용자 100명은 테스트 데이터로 사용한다. 실험 데이터 중에서 사용자 아이디가 70000번 이상인 사용자 30명과 영화 30편으로 신경망 모델을 생성한 후에, 통합 모델을 생성한다. 통합 모델으로써 다층 퍼셉트론은 두 개의 은닉 노드, initial weight는 -0.5에서 0.5, 학습율은 0.05로 하여 선호도를 판정한다.

<표 1>에서 보는 바와 같이 단일 신경망 모델만을 사용하는 추천 기술보다 사용자와 항목 신경망 모델을 통합한 추천 기술의 성능이 높음을 알 수 있다. 그 이유는 서로 다른 관점에서 선호도를 예측하기 때문에 두 신경망 모델의 선호도를 통합함으로써 단일 신경망 모델보다 성능이 우수하다. 통합 방법 중에서도 유연성과 적응성이 높은 신경망을 이용한 통합 방법이 가장 성능이 좋음을 알 수 있다. 사용자와 항목을 혼합한 연관 규칙 추천 기술과 비교하여도 본 논문에서 제안하고 있는 통합 방법이 추천 기술에 있어서 우수함을 알 수 있다.

6. 결론 및 향후 연구

추천 시스템에 있어서 신경망용 이용한 기술은 최근에 그 성능의 우수함이 입증되어지고 있다[4][11]. 본 논문에서는 사용자와 항목 신경망

모델을 통합함으로써 단일 신경망 모델이나 연관 규칙을 이용한 추천

<표 1> 각 통합 방법의 성능 비교

	연관 규칙	단일 신경망		사용자와 항목 신경망 모델의 통합					
		사용자와 항목의 혼합	사용자	순차적	병렬적	퍼셉트론	다층 퍼셉트론	BKS	퍼지 추론
Accuracy (%)	71.7	78.9	84.2	84.2	88.1	88.1	88.1	86.8	85.5

기술보다 높은 성능을 보이고 있다. 순차적 통합 방법과 병렬적 통합 방법은 선호도 예측 기준에 있어서 정적이고 적응성이 없으므로, 사용자와 항목 신경망 모델의 출력값을 입력으로 하는 퍼셉트론을 이용한 통합 방법도 제안한다. 그러나 선형 분리 가능 데이터만을 분리 할 수 있으므로 복잡한 경계선은 분리 할 수 없는 단점을 가지고 있다. 그러므로 다층 퍼셉트론을 이용한 통합 방법을 제안한다. 또한 두 모델의 선호도 소속 정도를 이용하기 위해 퍼지 추론을 이용한 통합 방법도 제안한다. 그리고, 패턴 인식에서 쓰이는 BKS 방법으로도 실험하였다. 사용자와 항목 신경망 모델이 서로 다른 특징과 관점으로 선호도를 예측하기 때문에 두 모델을 통합함으로써 추천 성능의 우수함을 보이고 있다.

향후 연구로는 기존의 CME 방법중 Fuzzy를 이용한 CME방법을 적용해보고자 한다. 두 신경망 모델의 실제 출력값은 Like 혹은 Dislike의 선호 정도이기 때문에 이러한 선호도는 Fuzzy를 이용한 CME 방법을 적용함으로써 성능을 향상시킬 수 있다. 그리고 사용자와 항목 신경망 모델뿐만 아니라 다른 입력값, 즉 다른 관점을 갖는 여러 신경망 모델을 생성하여 통합하는 방법을 적용해 보고자 한다.

7. 참고문헌

- [1] Michal J. Pazzani, A Framework for Collaborative, Content-Based and Demographic Filtering, Artificial Intelligence Review 13(5-6): pages 393-408, 1999
- [2] Badrul Sarwar, George Karypis, Joseph Konstan and John Riedl, Item-based Collaborative Filtering Recommendation Algorithms, WWW10 Conference, May, 2001
- [3] Weiyang Lin, Sergio A. Alvarez, Carolina Ruiz, Collaborative Recommendation via Adaptive Association Rule Mining, International Workshop on Web Mining for E-Commerce held in conjunction with the Sixth International Conference on Knowledge Discovery and Data Mining(KDD2000)
- [4] 김종수, 류정우, 도영아, 김명원, 신경망을 이용한 추천시스템, 한국 뇌학회 학술대회, p110~111, 6월, 2001
- [5] Y. S. Huang and C. Y. Suen, A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals, IEEE, 1995
- [6] C.A.O. Vieira and P.M.Mather, A Comparative Study of Multiple Classifier Combination Methods in Remote Sensing, In Proceedings of the IC-AI'2000, June, 2000
- [7] Tin Kam Ho, Jonathan J. Hull and Sargur N. Srihari, On Multiple Classifier Systems for Pattern Recognition, IEEE, 1992
- [8] Merijn Van Erp and Lambert Schomaker, Variants of The Borda Count Method for Combining Ranked Classifier Hypotheses, Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition, 2000
- [9] A. Khotanzad and C. Chung, Hand Written Digit Recognition using BKS Combination of Neural Network Classifiers, IEEE, 1994
- [10] P.McJoes, Eachmovie collaborative filtering data set, http://www.reach.digital.com/SRC/eachmovie, 1997, DEC Systems Research Center
- [11] D. Billsus and M. J. Pazzan, Learning collaborative information filters, In Proceedings of the Fifteenth International Conference on Machine Learning, pages 6-54, Madison, WI, 1998