김일성종합대학학보
정보과학

주체109(2020)년
제66권 제4호

프로그람작성실기체계에서 3층협동려과모형 실현의 한가지 방법

김경림, 김창수

협동려과(Collaborative Filtering)기술[1]은 많은 사람들의 경험과 의견으로부터 정보를 이끌어내는 기술로서 임의의 사람의 지난 시기 리력자료를 다른 사람들과 비교해보는 방식으로 어떤 항목을 선택하는가를 예측하는 기술이다. 이 기술은 사용자들사이 혹은 항목들사이의 류사도를 평가한 사용자정보들의 평가자료기지를 해석하고 사용자와 항목사이의 련관성을 평가하는데 리용된다.

협동려과기술에는 기억기에 기초한 방법과 모형에 기초한 방법이 있다.

① 기억기에 의한 방법은 사용자기반형식과 항목기반형식으로 구분된다. 이 방식들은 사용자들이 진행한 이전의 리력정보들을 해석하고 그것으로부터 얻어진 효률값에 대한 사용자평가를 통하여 류사도와 예측된 무게값을 리용한다.

② 모형에 기초한 방법[2]은 기계학습이나 통계적인 방법들을 리용하여 기초정보들로부터 모형을 학습하여 평가를 예측한다.

선행연구들은 지난 시기의 리력자료로부터 사용자평가를 예측하는 방법들이다.

론문에서는 직결심사체계의 호상작용특성을 리용하여 구축된 리력자료로부터 협동려과모형을 설계하고 프로그람작성실기체계에서 문제자동권고기능을 실현하는 방법을 제안하였다.

1. 3층협동려과모형설계

협동려과방법에서 많이 리용하고있는 잠재적디리클레할당(Latent Dirichlet Allocation)모형을 리용하여 사용자－수준－항목으로 된 협동려과모형에 대하여 보자. 여기서 매 사용자의 지난 기간 문제풀이리력자료를 가지고 매 항목에 대한 수준층을 생성할수 있다.

모형에서 사용자－항목선택과정은 두가지 단계로 진행된다.

① 매 사용자는 어떤 잠재적인 수준에 각이한 확률을 가지고 할당된다.

② 매 수준은 각이한 확률을 가지고 여러 항목들을 선택한다.

두 단계들에서는 모두 확률값들을 리용하므로 이 값을 근사적으로 계산하기 위하여 깁즈표본화(Gibbs Sampling)를 리용한다.

사용자의 항목선택과정은 확률적사건이다. 이때 어떤 사용자는 여러 수준에 속하며 매 수준에 속한 사용자들은 각이한 항목들에 대한 선택을 진행한다. 즉 이 방법에서 항목선택과정은 두가지 확률과정의 결합으로서 매 사용자는 어떤 잠재적인 수준의 어떤 확률값으로 할당되고 매 수준은 각이한 확률로 여러 항목들을 선택하게 된다.

프로그람작성실기체계에서 구현된 3층협동려과모형을 다음의 그림에 보여주었다.

수준층 $T=\{T_1, T_2, \cdots, T_K\}$ 는 사용자층 $U=\{U_1, U_2, \cdots, U_M\}$ 과 항목층 $I=\{I_1, I_2, \cdots, I_N\}$ 으로부터 생성된다. $k=1, \cdots, K,\ m=1, \cdots, M, n=1, \cdots, N$ 일 때 사용자 U_m 이 수준 T_k 에 할당될 확률은 $p(T_k|U_m)=\theta_{mk}$ 로, 수준 T_k 가 할당된 사용자가 항목 I_n 을 선택할 확률은

$p(I_n|T_k)=\phi_{nk}$ 로 표현된다.

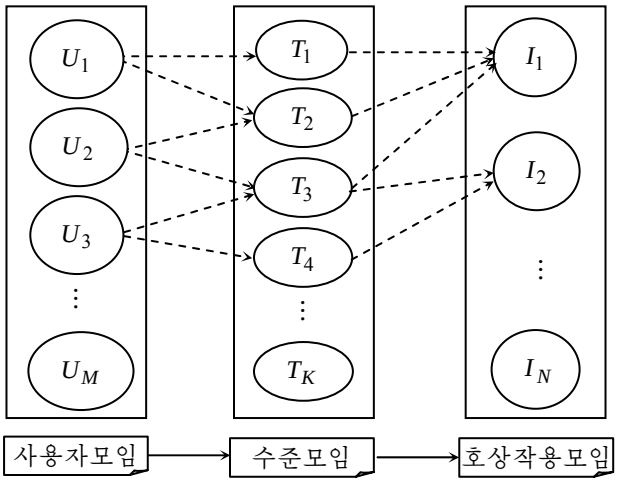

그림. 3층협동려과모형

사용자 U_m이 항목 I_n을 선택할 확률은 다음과 같다.

$$p(I_n|U_m)=\sum_k p(I_n|T_k)p(T_k|U_m)=\sum_k \phi_{nk}\theta_{mk} \tag{1}$$

두 분포 ϕ_{nk}와 θ_{mk}를 계산하기 위해 생성확률모형인 잠재적디리클레할당을 리용한다.

$$\theta_{mk}=p(T_k|U_m)=\frac{Count_{mk}^{MK}+\alpha}{\sum_{q=1}^{k} Count_{mq}^{MK}+K\alpha} \tag{2}$$

$$\phi_{nk}=p(I_n|T_k)=\frac{Count_{nk}^{NK}+\beta}{\sum_{q=1}^{k} Count_{qk}^{NK}+N\beta} \tag{3}$$

식에서 $Count_{nk}^{NK}$는 항목 I_n이 수준 T_k로부터 표본화된 개수이고 $Count_{mk}^{MK}$는 사용자 U_m이 항목에 할당되는 수준 T_k의 개수이다. 그리고 2개 파라메터들은 $\alpha=50/K, \beta=0.01$로서 경험적으로 설정된다.

$$p(I_n|U_m)=\sum_{k=1}^{K} p(I_n|t=k)p(t=k|U_m)=\sum_{k=1}^{K}\phi_{nk}\theta_{mk} \tag{4}$$

여기서 확률값이 클수록 사용자가 해당 문제를 선택할 확률이 크다는것을 의미하며 이 순위에 따라 문제권고목록을 얻을수 있다.

2. 협동려과모형에 의한 문제자동권고방법

프로그람작성실기체계의 호상작용방식에서 리용되는 요소들을 다음과 같이 정의한다.

정의 호상작용항목

사용자모임을 $U = \{U_1, U_2, \cdots, U_m\}$, 문제모임을 $P = \{P_1, P_2, \cdots, P_l\}$, 귀환상태모임을 $S = \{S_1, S_2, \cdots, S_j\}$라고 할 때 매개 호상작용은 하나의 항목을 이루며 그 항목모임은

$$I = \{[P_l, S_j] \mid l = 1, 2, \cdots, L, \quad j = 1, 2, \cdots, J\}$$

로 정의된다. 여기서 귀환상태모임에는 프로그람작성실기체계의 특성으로부터 콤파일오유(CE: Compile Error), 실행오유(RE: Runtime Error), 제한시간초과(TLE: Time Limit Exceeded), 틀린답(WA: Wrong Anwser), 접수(AC: Accepted) 등 5가지 상태값들이 속할수 있으며 또는 성적등급이 속할수도 있다.

사용자 U_m이 P_l을 선택할 확률이 $p(P_l \mid U_m)$이라고 할 때 $p(P_l \mid U_m)$이 클수록 사용자 U_m이 P_l을 더 잘 선택한다는것을 의미한다.

프로그람작성실기체계에서 항목은 2항요소이다. 즉

$$I_n = [P_l, U_m]$$

이다.

문제 P_l이 사용자 U_m에 제공된다고 가정하면 잠재적인 수준 T_k와 사용자 U_m은 문제 P_l을 포함하는 항목들에 할당된다. 다시말하여 가능한 큰 확률을 가진 $I_{p(l)}$이라는 것이다.

호상작용항목 $I_{p(l)}$에 대한 수준 T_k의 분포 $\phi_{p(l)k}$는 식 (2)에 의하여 다음과 같이 표시된다.

$$\phi_{nk} = p(I_{p(l)} \mid T_k) = \frac{Count_{p(l)k}^{NK} + \beta}{\sum_{q=1}^{L} Count_{p(q)k}^{LK} + N\beta} \tag{5}$$

식에서 $Count_{p(l)k}^{NK}$는 항목 $I_{p(l)}$이 수준 T_k에 할당되는 개수를 표시한다. 따라서 사용자의 수준에 따르는 문제모임은

$$p(P_l \mid U_m) = p(I_{p(l)} \mid U_m) = \sum_{k=1}^{K} p(I_{p(l)} \mid t = k)\, p(t' = k \mid U_m) = \sum_{k=1}^{K} \phi_{p(l)k}\theta_{mk} \cdots \tag{6}$$

와 같이 표시된다.

깁즈표본화로 다음의 알고리듬과 같이 최종적인 $\phi_{p(l)k}$를 얻을수 있다.

① 매개 항목

$$I = \{[P_l, S_j] \mid l = 1, 2, \cdots, L, \quad j = 1, 2, \cdots, J\}$$

를 수준모임 $T = \{T_1, T_2, \cdots, T_k\}$에서 수준 T_k에 우연적으로 할당하는데 이것은 특정한 마르꼬브사슬의 시작요구에 따라 초기화되며 순환주기는 $w=1$로 설정한다.

② w번째의 순환에서 다음의 조건을 만족시켜야 한다.

if eval(Assignment(T, I)) > eval (BestAssignment)

BestAssignment ← Assignment(T, I)

여기서 eval(·)는 평가함수이며 Assignment(T, I)는 현재항목 $I_{p(l)}$에 할당된 수준 T_k를 나타낸다. 그것은 항목 $I_{\neg p(l)}$과 $I_{p(l)}$에 마지막으로 할당된 T_k에 따라 선택되며 그것들의 분포는

$$p^{(w)} = p(t_{p(l)} = k \mid (t_{p(l)} = k, t_{\neg p(l)} = k, S_{p(l)}, S_{\neg p(l)}))^{(w-1)}$$

와 같은 귀환상태모임 S에 따른다. 여기서 $k=1,\cdots,K,\ l=1,\cdots,L$이다.

③ $w=w+1$로 설정하고 ②로 간다. 충분히 순환된 후 $\phi_{p(l)k}\ (k=1,\cdots,K,\ l=1,\cdots,L)$는 최종적으로 안정화된 BestAssignment에 따라 식 (5)에 의하여 계산된다.

②단계에서 $p^{(w)}$와 eval(Assignment(T, I))는 통계학적으로 다음의 식으로 평가된다.

$$p^{(w)}=p(t_{p(l)}=k\,|\,(t_{p(l)}=k,\ t_{\neg p(l)}=k,\ S_{p(l)},\ S_{\neg p(l)}))^{(w-1)}\propto$$

$$\propto\frac{a\times Count^{LKJ}_{\neg p(l)kS(t)}+b\times Count^{LKJ}_{\neg p(l)kS(t)}+\gamma}{\sum_{k=1}^{K}\left(a\times Count^{LKJ}_{\neg p(l)kS(t)}+b\times a\times Count^{LKJ}_{p(l)kS(t)}\right)+K\times\gamma} \tag{7}$$

$$\text{eval(Assignment(T, I))}\ \propto\sum_{l=1}^{L}\frac{\sum_{k=1}^{K}Count^{LKJ}_{\neg p(l)kS(t)}}{\sum_{k=1}^{K}Count^{LKJ}_{\neg p(l)kS(t)}} \tag{8}$$

식 (7)과 (8)에서 $Count^{LKJ}_{\neg p(l)k}$는 T_k가 할당된 문제 P_l을 포함하지 않는 항목 즉 $I_{\neg p(l)}$의 개수를 표시하고 $Count^{LKJ}_{\neg p(l)kS(t)}$는 T_k가 할당된 $I_{\neg p(l)}$들중에서 T_k의 할당을 지원하는 상태를 가지는 항목개수이며 $Count^{LKJ}_{p(l)kS(t)}$는 T_k가 할당된 $I_{p(l)}$들중에서 T_k의 할당을 지원하는 상태를 가지는 항목개수이다. 또한 γ는 분자가 령이 되는 경우의 동조파라메터이고 $a,\ b$는 어느 부분을 더 고려하여야 하는가를 결정하는 평형파라메터이다. 이때 $a,\ b$는 $a+b=1$을 만족시켜야 한다.

$\phi_{mk}\ (m=1,\cdots,M,\ k=1,\cdots,K)$에 대해서는 알고리듬에서 생성된 수준에 대한 사용자 항목의 분포로부터 직접 얻을수 있다.

최종적으로 $p(P_l\,|\,U_m)$은 식 (6)에 의하여 θ_{mk}와 $\phi_{p(l)k}$로부터 얻을수 있다. 이 확률값이 크다는것은 사용자가 문제를 선택할 확률이 보다 높다는것을 의미하며 이 확률값을 순서화하여 권고목록을 작성할수 있다. 따라서 우의 알고리듬에 의하여 얻어지는 문제항목들을 목록화하여 학생들의 수준에 따르는 문제목록을 자동으로 제공하도록 한다.

맺 는 말

프로그람작성실기체계에서 3층협동려과모형을 설계하고 문제자료기지에서 학생들의 수준에 알맞는 문제목록자동권고방법을 제안하였다.

참 고 문 헌

[1] Q. Liu et al.; IEEE Trans. Sys. Man., 42, 1, 218, 2012.

[2] W. Wang et al.; International Journal of Intelligent Systems, 30, 8, 854, 2015.

주체109(2020)년 8월 5일 원고접수

A Method of Three-layer Collaborative Filtering Recommendation Implementation in Practical System for Programming

Kim Kyong Rim, Kim Chang Su

In this paper, we designed the three-layer collaborative filtering model in practical system for programming and proposed the method of automatic recommendation of problem list using this model.

Keywords: collaborative filtering, recommendation system, online judge