

DATA SCIENCE LAB – 04/05/2019

MÔ HÌNH PHÂN RÃ MA TRẬN POISSON KẾT HỢP BỘ TRI THỨC TIỀN NGHIỆM TRONG HỆ GỢI Ý

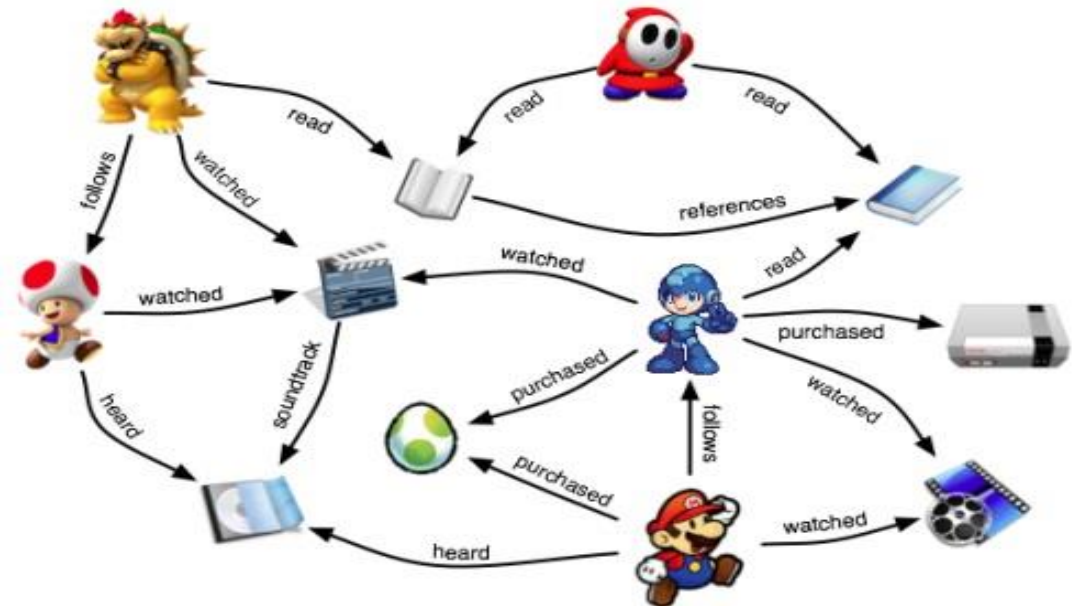
NGUYỄN VĂN TÚC

NỘI DUNG TRÌNH BÀY

- I. Tổng quan hệ gợi ý
- II. Tiếp cận bài toán
- III. Tri thức tiên nghiệm, kỹ thuật dropout
- IV. Mô hình đề xuất
- V. Thử nghiệm và đánh giá
- VI. Kết luận

I. Tổng quan hệ gợi ý (recommender system)

- ❖ Hệ thống có khả năng gợi ý sản phẩm phù hợp có liên quan đến người dùng
- ❖ Được ứng dụng phổ biến trong các website tin tức, thương mại điện tử, quảng cáo hoặc xem phim trực tuyến.
- ❖ Ví dụ:
 - Hệ thống gợi ý video của Youtube, MyClip.vn
 - Trang web thương mại điện tử: Amazon.com



I. Tổng quan hệ gợi ý (recommender system)

❖ Bài toán hệ gợi ý: ước lượng đánh giá độ phù hợp của các sản phẩm chưa được người dùng xem xét, từ đó đưa ra gợi ý một cách phù hợp.

❖ Mô tả bài toán:

- U người dùng, I sản phẩm
- Ma trận tương tác user – item
- Các thông tin khác: user profile, item,...
- **Output:** Mức độ phù hợp user – item

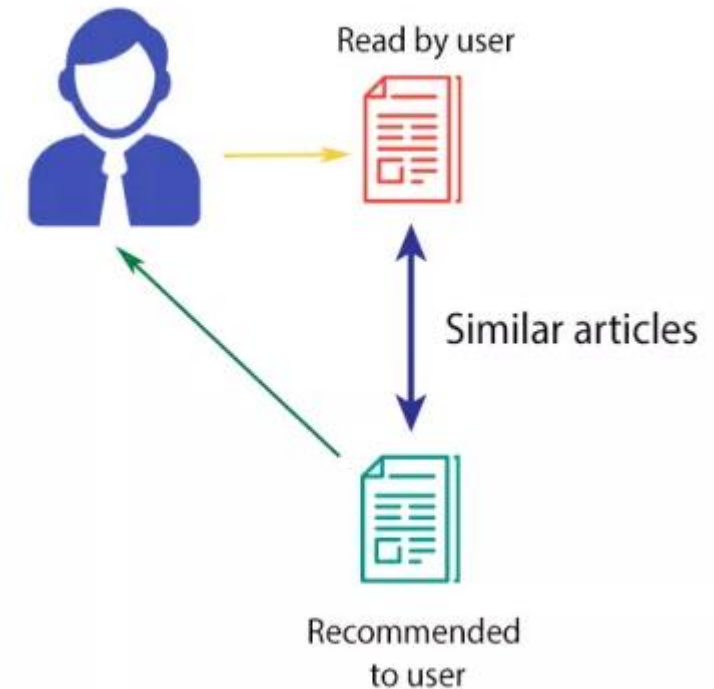
❖ Các phương pháp tiếp cận:

- Content-based filtering
- Collaborative filtering

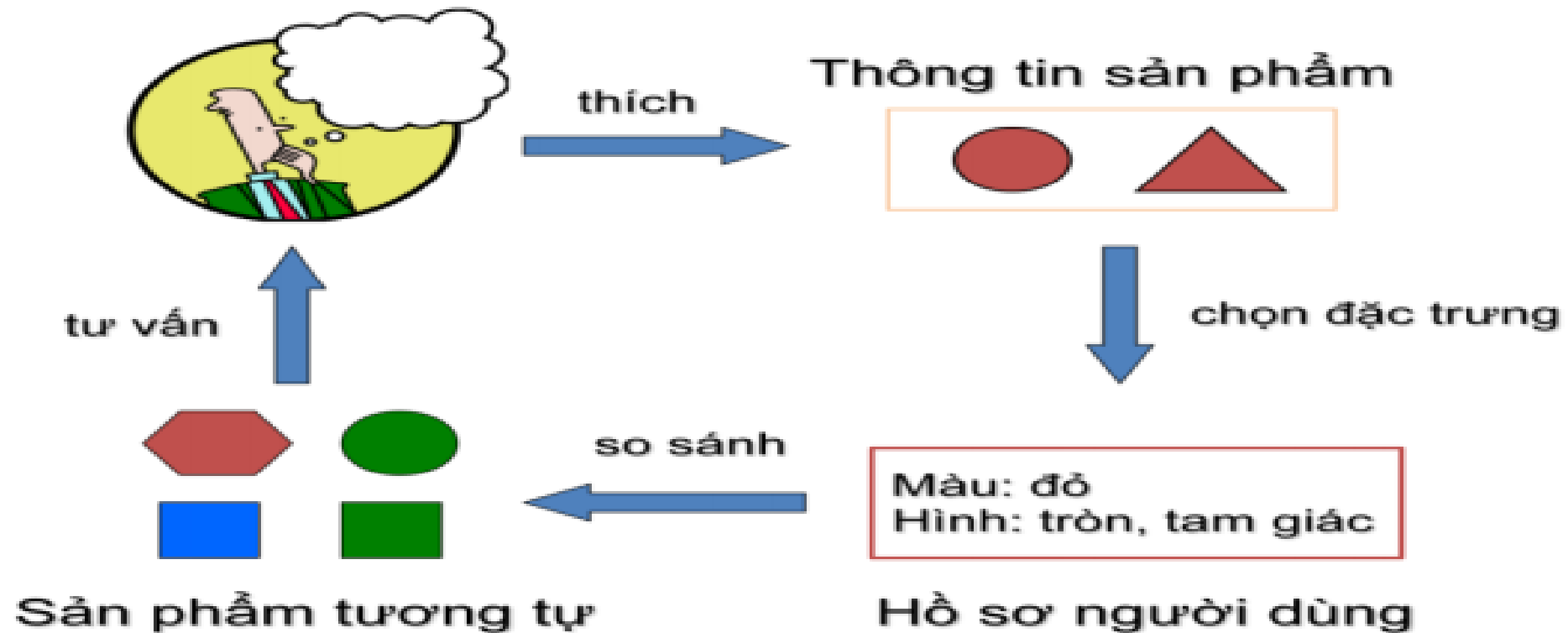
	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0	3	0	3	0
User 2	4	0	0	2	0
User 3	0	0	3	0	0
User 4	3	0	4	0	3
User 5	4	3	0	4	0

Tiếp cận dựa trên lọc nội dung (content-based filtering)

- ❖ Xây dựng vector đặc trưng cho mỗi item.
- ❖ Dựa trên đặc điểm của các item từng được user đánh giá cao.
- ❖ Đề xuất item tương đồng với sở thích của user.

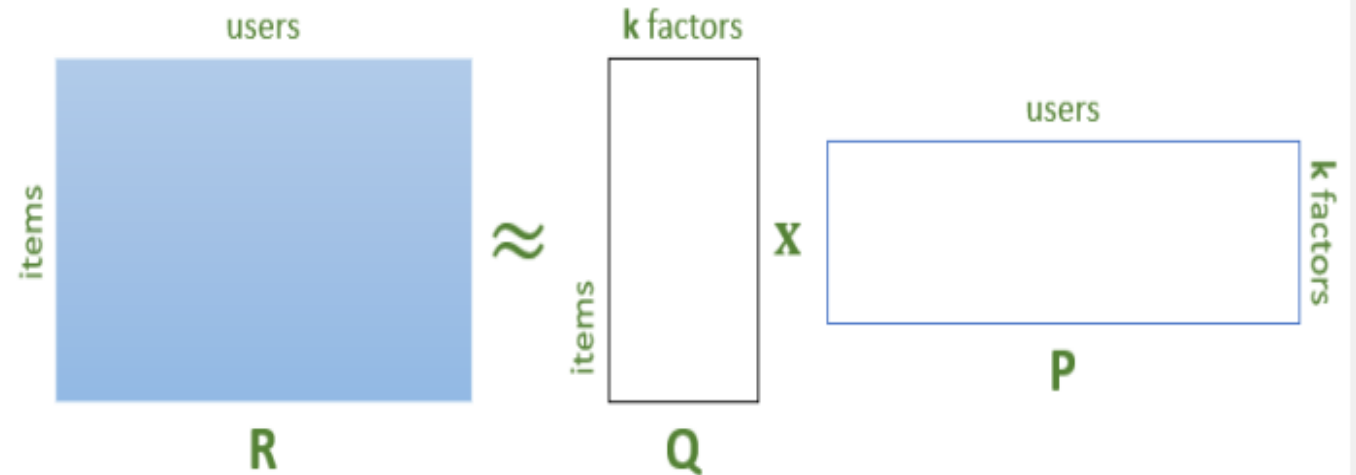
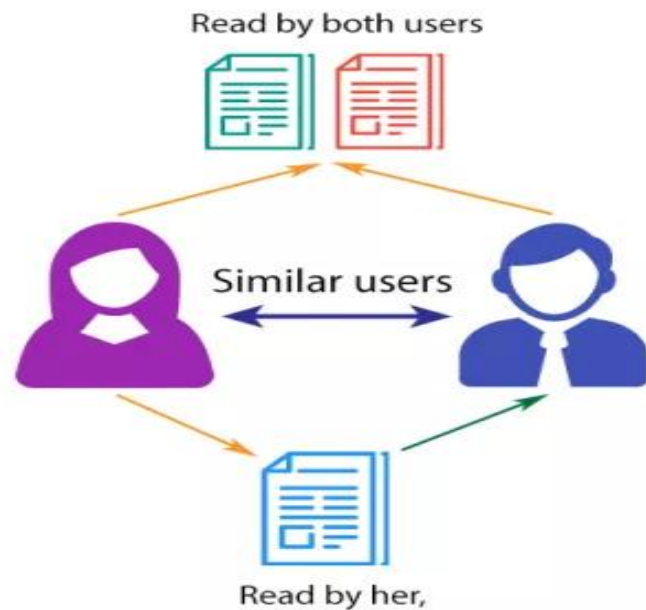


Tiếp cận dựa trên lọc nội dung (content-based filtering)



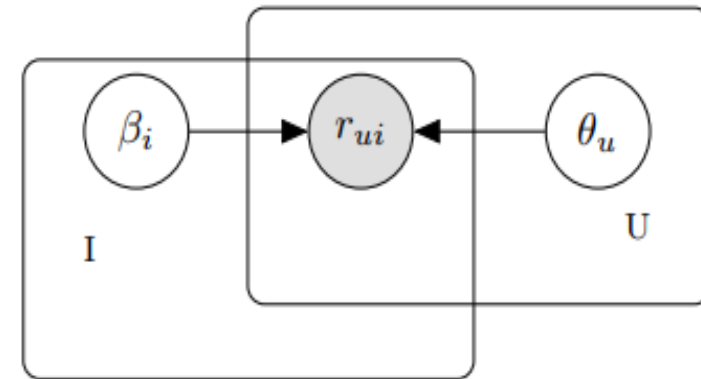
Tiếp cận dựa trên lọc cộng tác (collaborative filtering)

- ❖ Neighborhood-Based.
- ❖ Matrix Factorization.



II. Tiếp cận bài toán.

- ❖ Xét một hệ thống có:
 - U người dùng và I sản phẩm
 - Ma trận tương tác $R = \{ r_{ui} \}_{u=1, i=1}^{U, I}$
 - Tìm β , θ sao cho $\beta^T \times \theta$ gần với R nhất
- ❖ Mô hình đồ thị xác suất cho phân rã ma trận.
 - β ma trận item features, θ ma trận user features.
 - Tương tác user u, item i:
$$r_{ui} = \beta_i^T \cdot \theta_u$$



Hình: Mô hình đồ thị xác suất cho phân rã ma trận

II. Tiếp cận bài toán.

❖ Đặt $\bar{r} = \beta_i^T \theta_u$ là kỳ vọng của giá trị quan sát thực tế r_{ui}

và tuân theo một hàm phân phối:

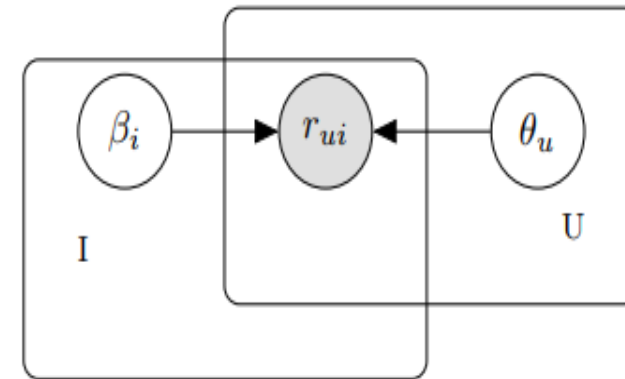
$$p_{ui}(r_{ui} | \bar{r}_{ui})$$

❖ Ta có hàm mục tiêu:

$$P(R|\theta, \beta) = \prod_{u=1, i=1}^{U, I} p_{ui} = \prod_{u, i} p_{ui}$$

❖ Đưa về dạng log:

$$L = \sum_{u=1, i=1}^{U, I} \log p_{ui} = \sum_{u, i} \log p_{ui}$$



Hình: Mô hình đồ thị xác suất cho phân rã ma trận

1. Sử dụng phân phối Gauss.

- ❖ Hàm xác suất P tuân theo phân phối Gaussian:

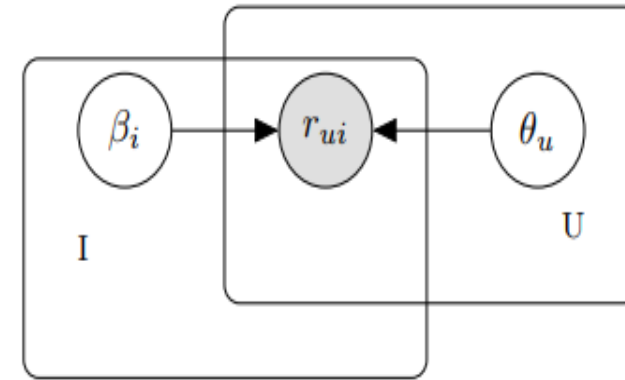
$$p(r_{ui}|\bar{r}_{ui}) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{r_{ui}-\bar{r}_{ui}}{\sigma}\right)^2}$$

- ❖ Với σ thể hiện sự biến động của tương tác quan sát được r_{ui} quanh kỳ vọng \bar{r}_{ui}

- ❖ Đặt $\frac{1}{\sigma^2} = c$, c : độ tin cậy của giá trị quan sát r_{ui} .

- ❖ Hàm mục tiêu:

$$\begin{aligned} L &= - \sum_{u=1, i=1}^{U, I} \frac{c}{2} (r_{ui} - \bar{r}_{ui})^2 + const \\ &= - \sum_{u=1, i=1}^{U, I} \frac{c}{2} (r_{ui} - \beta_i^T \theta_u)^2 + const \end{aligned}$$



Hình: Mô hình đồ thị xác suất cho phân rã ma trận

1. Sử dụng phân phối Gauss.

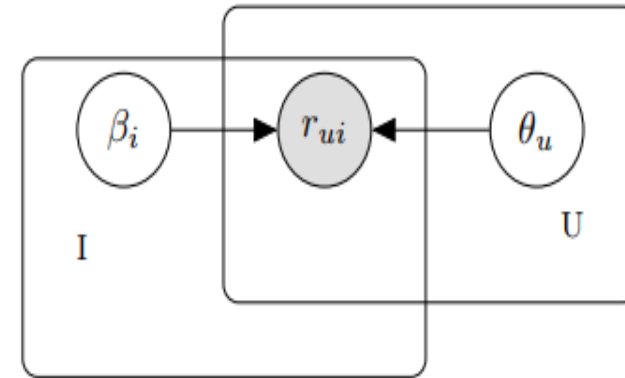
❖ Hàm mục tiêu:

$$\begin{aligned} L &= - \sum_{u=1, i=1}^{U, I} \frac{c}{2} (r_{ui} - \bar{r}_{ui})^2 + const \\ &= - \sum_{u=1, i=1}^{U, I} \frac{c}{2} (r_{ui} - \beta_i^T \theta_u)^2 + const \end{aligned}$$

❖ r_u vector $I \times 1$ biểu diễn tương tác của người dùng u với I sản phẩm:

r_i vector $U \times 1$ biểu diễn tương tác của U người dùng với sản phẩm I

❖ Đạo hàm L theo θ_u, β_i được công thức cập nhật: $\theta_u = (\beta \beta^T)^{-1} \beta r_u$
 $\beta_i = (\theta \theta^T)^{-1} \theta r_i$



Hình: Mô hình đồ thị xác suất cho phân rã ma trận

1. Sử dụng phân phối Gauss.

- ❖ Mô hình phân rã Gauss với ràng buộc biến:

$$p(\theta_u | \lambda_U) \sim N(0, \lambda_U^{-1} I_K)$$

$$p(\beta_i | \lambda_I) \sim N(0, \lambda_I^{-1} I_K)$$

- ❖ Từ đó: $P(R | \theta, \beta, \lambda_U, \lambda_I) = p(R | \theta, \beta) p(\theta | \lambda_U) p(\beta | \lambda_I)$

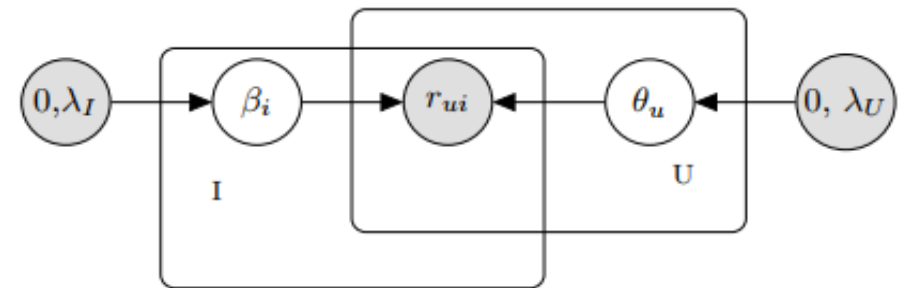
- ❖ Hàm mục tiêu:

$$L = -\frac{c}{2} \sum_{u=1, i=1}^{U, I} (r_{ui} - \beta_i^T \theta_u)^2 - \frac{\lambda_U}{2} \sum_{u=1}^U \theta_u^T \theta_u - \frac{\lambda_I}{2} \sum_{i=1}^I \beta_i^T \beta_i + const$$

- ❖ Công thức cập nhật:

$$\theta_u = (\beta \beta^T + \lambda_U I_K)^{-1} \beta r_u$$

$$\beta_i = (\theta \theta^T + \lambda_I I_K)^{-1} \theta r_i$$



Hình: Mô hình đồ thị xác suất cho phân rã ma trận Gaussian

2. Sử dụng phân phối Poisson.

- ❖ Giả thiết r_{ui} là các giá trị rời rạc tuân theo phân phối Poisson với kỳ vọng $\beta_i^T \theta_u$

$$\begin{aligned} p(r_{ui} | \beta_i^T \theta_u) &= \text{Poisson}(\beta_i^T \theta_u) \\ &= (\beta_i^T \theta_u)^{r_{ui}} \frac{\exp^{-\beta_i^T \theta_u}}{r_{ui}!} \end{aligned}$$

- ❖ Khi đó hàm mục tiêu:

$$\log p(R | \theta, \beta) = \left(\sum_{i,u,r_{ui}>0}^{U,I} r_{ui} \log(\beta_i^T \theta_u) - \log r_{ui}! \right) - \left(\sum_{i=1}^I \beta_i^T \right) \left(\sum_{u=1}^U \theta_u \right)$$

2. Sử dụng phân phối Poisson.

❖ Sử dụng biến phụ và phương pháp suy diễn biến phân

❖ Với điều kiện: $z_{uik} \sim \text{Poisson}(\beta_{ik}\theta_{uk})$

$$\sum_{k=1}^K z_{uik} = r_{ui}$$

❖ Sử dụng suy diễn biến phân với z_{ui}

$$q(z_{ui}|\phi_{ui}) \sim \text{Multinomial}(\phi_{ui})$$

$$\sum_k \phi_{uik} = 1$$

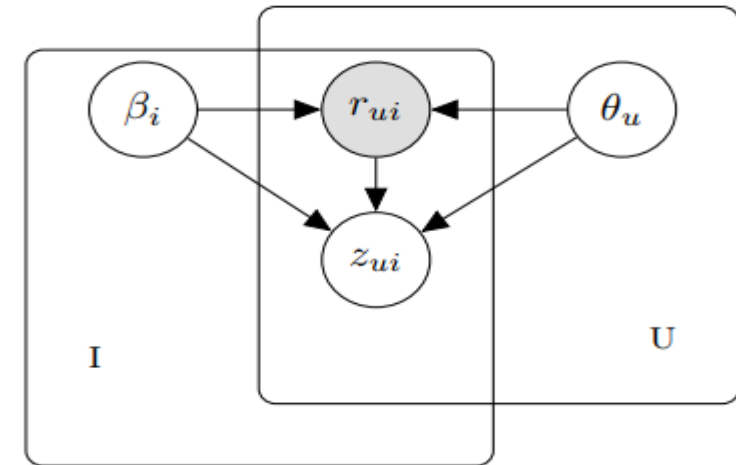
$$E_{q(z)} z_{uik} = r_{ui} \phi_{uik}$$

❖ Sử dụng suy diễn biến phân thu hàm biên dưới (lower bound):

$$l = E_{q(z)} \log p(R, \beta, \theta, Z) - E_{q(z)} \log q(z)$$

❖ Công thức cập nhật

$$\phi_{uik} \propto \beta_{ik} \theta_{uk} \quad \beta_{ik} = \frac{\sum_u r_{ui} \phi_{uik}}{\sum_u \theta_u} \quad \theta_{uk} = \frac{\sum_i r_{ui} \phi_{uik}}{\sum_i \beta_i}$$



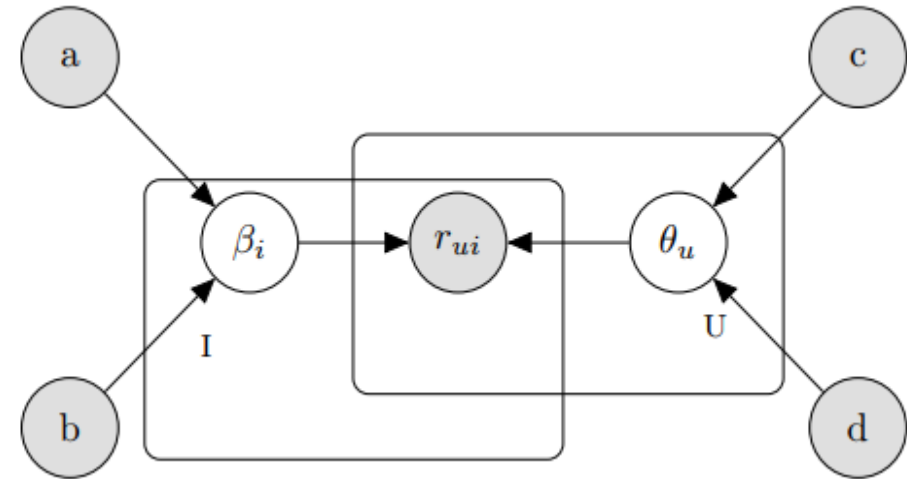
Hình: Mô hình đồ thị xác suất cho phân rã Poisson

2. Mô hình gợi ý sử dụng phân rã Poisson phân cấp - HPF.

- ❖ Sử dụng các mức phân cấp cho θ và β
- ❖ $p(\beta_{ik}|a, b) \sim \text{Gamma}(a, b)$
 $p(\theta_{uk}|c, d) \sim \text{Gamma}(c, d)$
- ❖ Mức phân cấp thứ 2 các tham số b và d được giả thiết tuân theo phân phối Gamma:

$$b_i \sim \text{Gamma}(a', a'/b')$$

$$d_u \sim \text{Gamma}(c', c'/d')$$



Hình: Mô hình đồ thị xác suất cho phân rã Poisson cấp 1

2. Mô hình gợi ý sử dụng phân rã Poisson phân cấp - HPF.

❖ Mô hình sinh:

❖ Mỗi sản phẩm i :

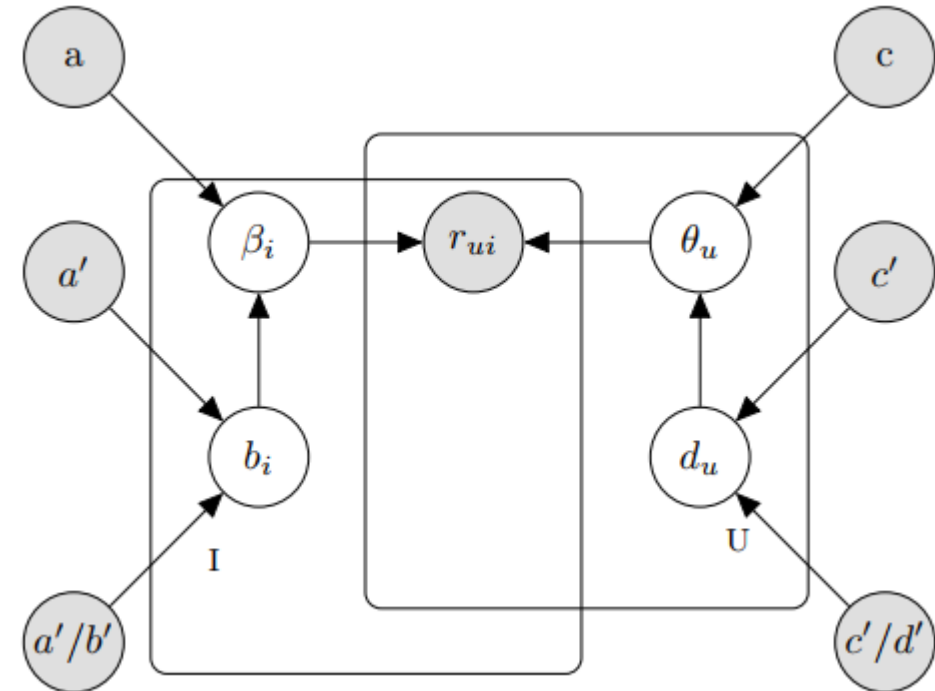
- Lấy mẫu $b_i \sim \text{Gamma}(a', a'/b')$
- Mỗi thành phần k , lấy mẫu: $\beta_{ik} \sim \text{Gamma}(a, b_i)$

❖ Mỗi người dùng u :

- $d_u \sim \text{Gamma}(c', c'/d')$
- $\theta_{uk} \sim \text{Gamma}(c, d_u)$

❖ Mỗi cặp người dùng u , item i

$$r_{ui} \sim \text{Poisson}(\beta_i^T \theta_u)$$



Hình: Mô hình đồ thị xác suất cho phân rã Poisson cấp 2

2. Mô hình gợi ý sử dụng phân rã Poisson phân cấp - HPF.

- ❖ $q(b_i|\kappa_i) \sim \text{Gamma}(\kappa_i^{shp}, \kappa_i^{rte})$
 $q(d_u|\tau_u) \sim \text{Gamma}(\tau_u^{shp}, \tau_u^{rte})$
- ❖ $q(\beta, \theta, b, d, z) = \prod_{i,k} q(\beta_{ik}|\zeta_{ik}) \prod_{u,k} q(\theta_{uk}|\gamma_{uk}) \prod_i q(b_i|\kappa_i) \prod_u q(d_u|\tau_u) \prod_{u,i} q(z_{ui}|\phi_{ui})$

❖ Với các thành phần:

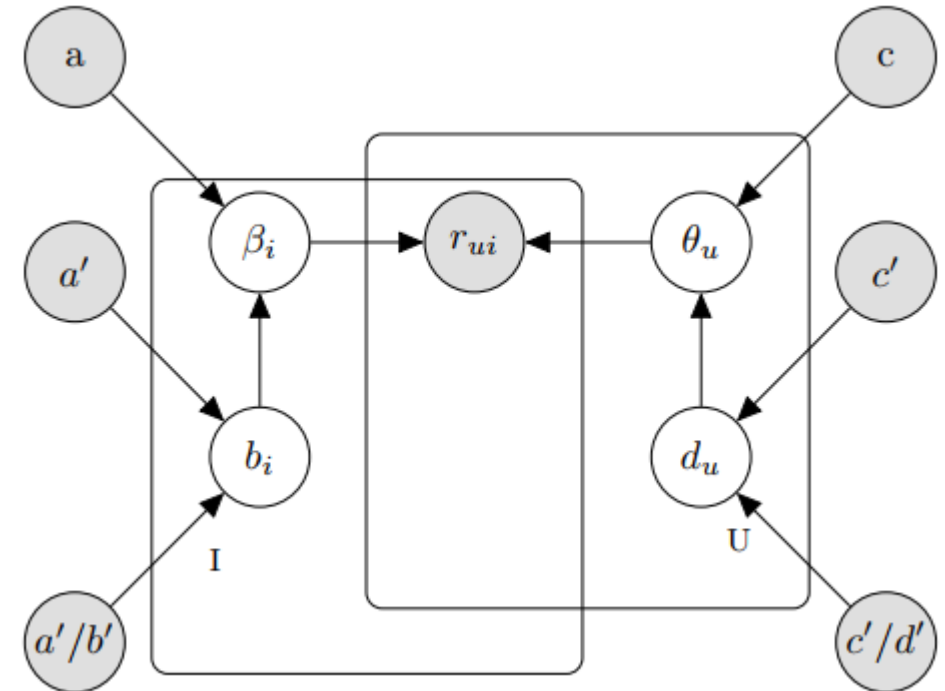
$$q(\beta_{ik}|\zeta_{ik}) \sim \text{Gamma}(\zeta_{ik}^{shp}, \zeta_{ik}^{rte})$$

$$q(\theta_{uk}|\gamma_{uk}) \sim \text{Gamma}(\gamma_{uk}^{shp}, \gamma_{uk}^{rte})$$

$$q(z_{ui}|\phi_{ui}) \sim \text{Multinomial}(\phi_{ui})$$

$$q(b_i|\kappa_i) \sim \text{Gamma}(\kappa_i^{shp}, \kappa_i^{rte})$$

$$q(d_u|\tau_u) \sim \text{Gamma}(\tau_u^{shp}, \tau_u^{rte})$$



2. Mô hình gợi ý sử dụng phân rã Poisson phân cấp - HPF.

$$\phi_{ui} \propto \exp\{\Psi(\gamma_{uk}^{shp}) - \log \gamma_{uk}^{rte} + \Psi(\zeta_{ik}^{shp}) - \log(\zeta_{ik}^{rte})\}$$

$$\gamma_{uk}^{shp} = c + \sum_{i=1}^I r_{ui} \phi_{uik}$$

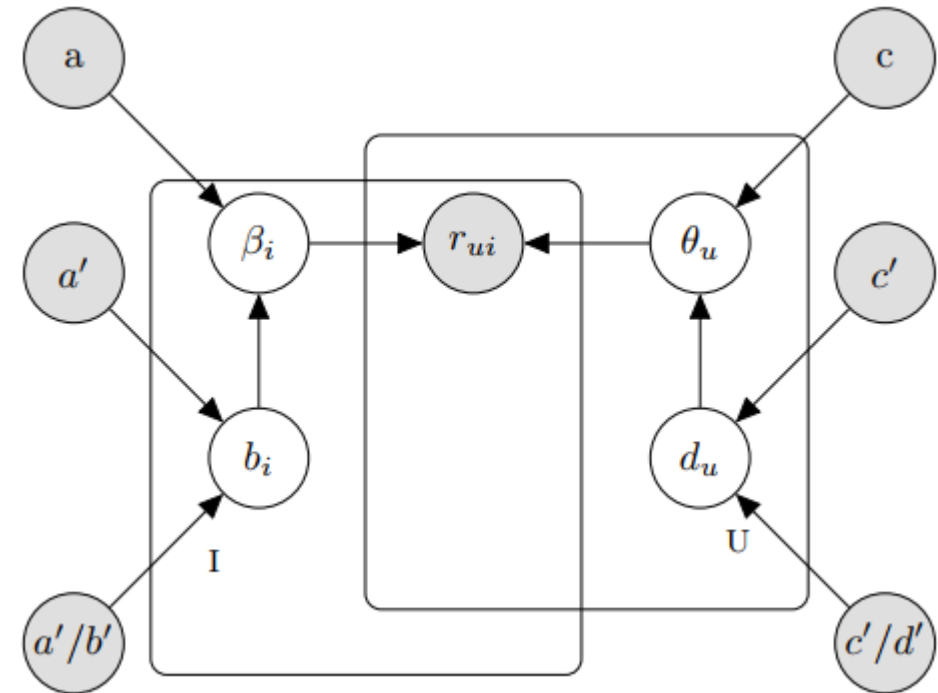
$$\gamma_{uk}^{rte} = \frac{\tau_u^{shp}}{\tau_u^{rte}} + \sum_{i=1}^I \zeta_{ik}^{shp} / \zeta_{ik}^{rte}$$

$$\tau_u^{rte} = \frac{c'}{d'} + \sum_k \frac{\gamma_{uk}^{shp}}{\gamma_{uk}^{rte}}$$

$$\zeta_{ik}^{shp} = a + \sum_{u=1}^U r_{ui} \phi_{uik}$$

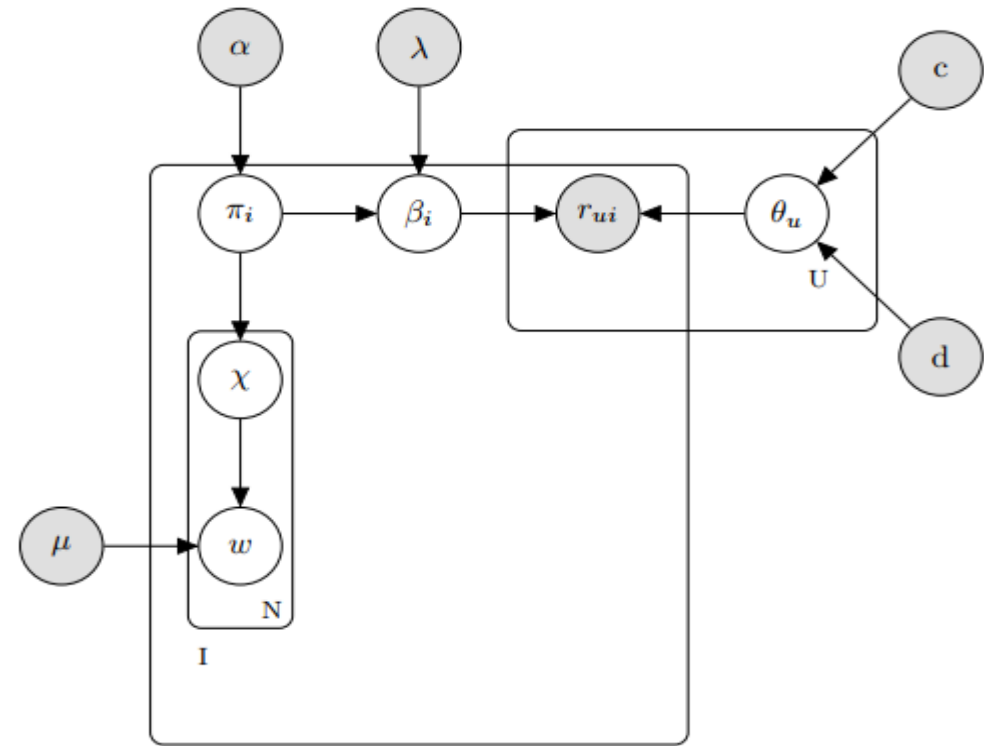
$$\zeta_{ik}^{rte} = \frac{\kappa_i^{shp}}{\kappa_i^{rte}} + \sum_{u=1}^U \gamma_{uk}^{shp} / \gamma_{uk}^{rte}$$

$$\kappa_i^{rte} = \frac{a'}{b'} + \sum_k \frac{\zeta_{ik}^{shp}}{\zeta_{ik}^{rte}}$$



3. Mô hình gợi ý sử dụng phân rã Poisson kết hợp LDA.

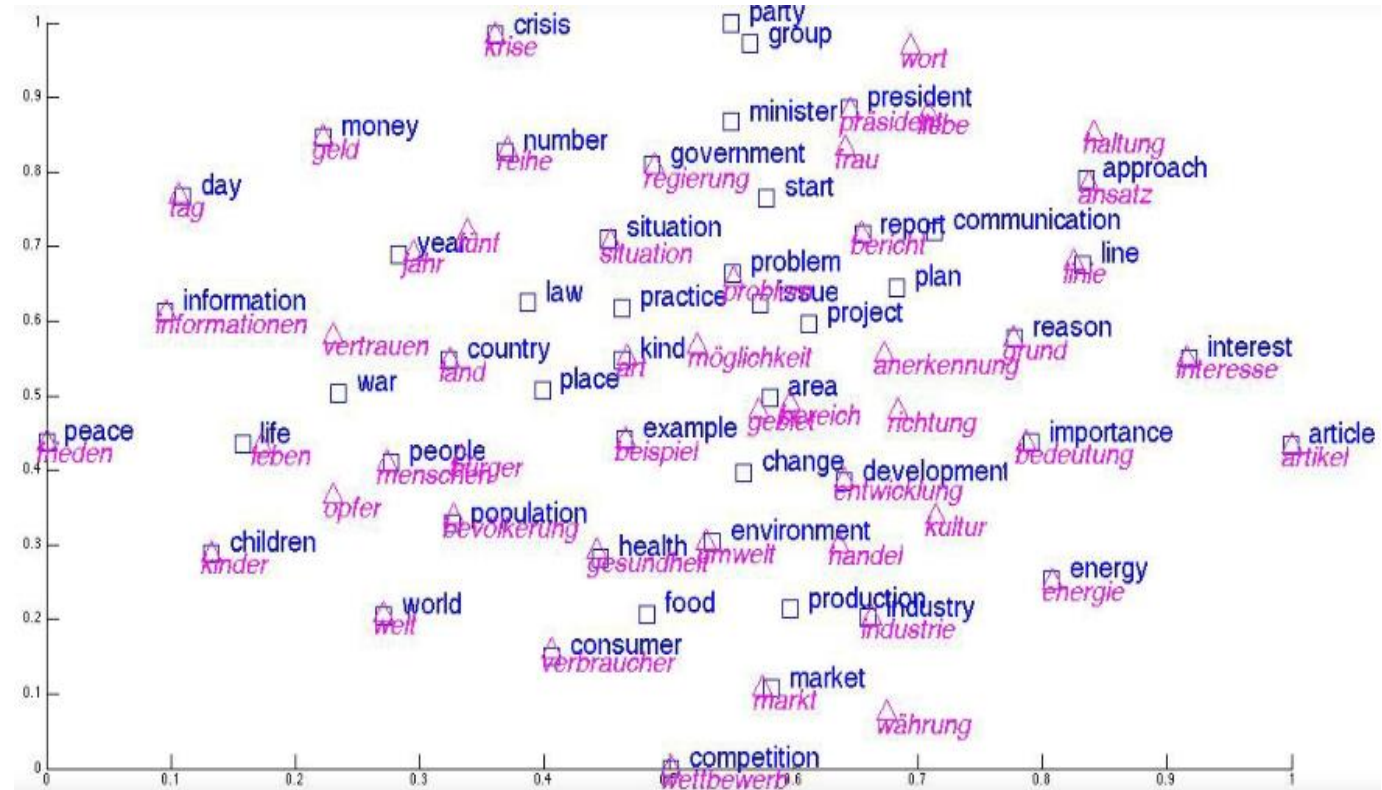
- ❖ Khai thác nội dung sản phẩm qua mô hình LDA.
- ❖ Sử dụng các thông tin từ các item có trong hệ thống.
- ❖ Mô hình sinh:
- ❖ Với người dùng u : $\theta_u \sim \text{Gamma}(c, d)$
- ❖ Với mỗi sản phẩm i :
 - Lấy phân phối chủ đề: $\pi_i \sim \text{Dirichlet}(\alpha)$
 - Từ thứ n trong mô tả của sản phẩm i :
 - Lấy chủ đề $X_{in} \sim \text{Categorical}(\pi_i)$
 - Lấy từ $w_{in} \sim \text{Categorical}(\mu_{X_{in}})$
 - Item features: $\beta_i \sim N(\pi_i, \lambda^{-1}I_K)$
- ❖ Mỗi cặp user-item: $r_{ui} \sim \text{Poisson}(\beta_i^T \theta_u)$



Hình: Mô hình đồ thị xác suất cho phân rã Poisson - LDA

III. Tri thức tiên nghiệm, kỹ thuật dropout.

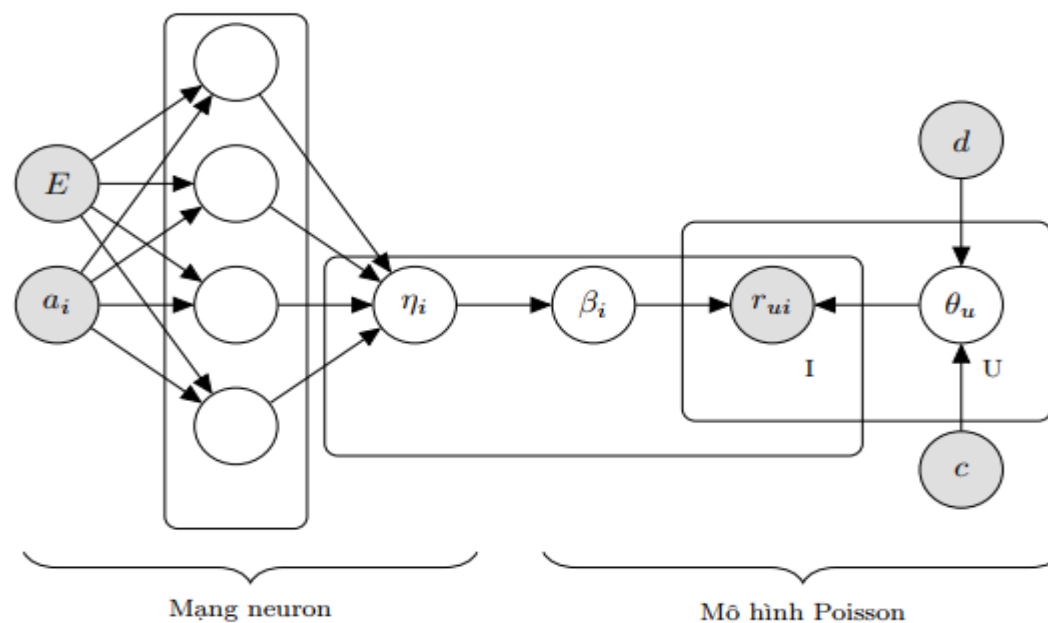
- ❖ Tri thức có trước về đối tượng.
- ❖ Kỹ thuật dropout: loại bỏ đi ngẫu nhiên đầu vào, các tương tác bị loại bỏ sẽ không được dùng trong việc học mô hình.
- ❖ Tri thức tiên nghiệm Glove.



Ảnh : Minh họa tri thức từ nhúng

IV. Mô hình đề xuất.

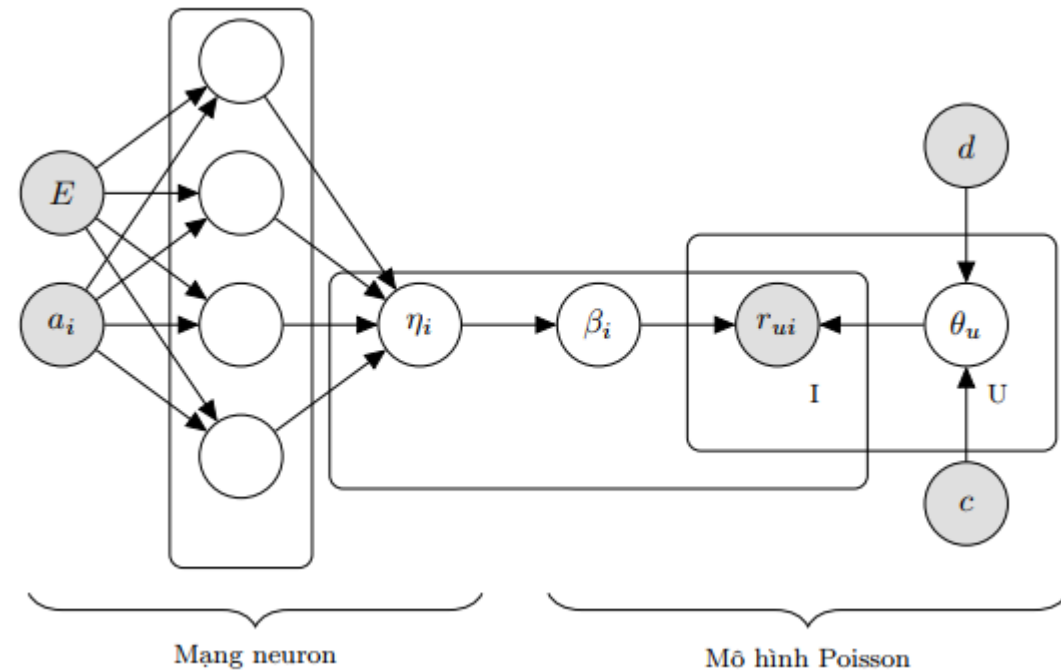
- ❖ Mô hình sử dụng thêm tri thức biểu diễn nhúng có trước của từ để tăng cường thêm thông tin từ sản phẩm.
- ❖ Việc học mô hình gồm 2 thành phần:
 - Học biểu diễn sản phẩm thông qua mạng neural W
 - Mô hình phân rã ma trận Poisson



Hình: Mô hình đồ thị xác suất cho phân rã Poisson kết hợp mạng neural

Mô hình sinh.

- ❖ Mỗi sản phẩm i :
 - Vector biểu diễn thông tin sản phẩm:
 $n_i = f(E, a_i; W)$
 - Vector biểu diễn thuộc tính của sản phẩm:
 $\beta_i = N(n_i, \lambda^{-1} I_K)$
- ❖ Mỗi người dùng u :
 - Vector thuộc tính người dùng:
 $\theta_{uk} \sim \text{Gamma}(c, d)$
- ❖ Mỗi cặp user-item:
 $r_{ui} \sim \text{Poisson}(\beta_i^T \theta_u)$



Hình: Mô hình đồ thị xác suất cho phân rã Poisson kết hợp mạng neural

Học mô hình.

❖ Hàm mục tiêu:

$$\begin{aligned} L &= \log P(\beta, R | E, a, c, d) = \log P(\beta | E, a) + \log P(R | \beta, c, d) \\ &= - \sum_{i=1}^I \frac{\lambda}{2} \|\beta_i - f(E, a_i; W)\|_2^2 + \log \int \sum_z P(R, z, \theta | \beta, c, d) d\theta \end{aligned}$$

❖ Sử dụng kỹ thuật suy diễn biến phân với:

$$Q(\theta, z) = \prod_{u=1, i=1}^{U, I} q(z_{ui} | r_{ui}, \phi_{ui}) \prod_{u=1}^U \prod_{k=1}^K q(\theta_{uk} | \text{shp}_{uk}, \text{rte}_{uk})$$

❖ Cập nhật W bằng stochastic gradient ascent:

$$l(W) = - \sum_{i=1}^I \frac{\lambda}{2} \|\beta_i - f(E, a_i; W)\|_2^2$$

Học mô hình.

❖ Cập nhật β :

$$\beta_{ik} = \frac{-\sum_u \frac{\text{shp}_{uk}}{\text{rte}_{uk}} + \lambda \eta_{ik} + \sqrt{\Delta}}{2\lambda} \quad (1)$$

với $\Delta = \left(-\sum_u \frac{\text{shp}_{uk}}{\text{rte}_{uk}} + \lambda \eta_{ik}\right)^2 + 4\lambda \sum_k r_{ui} \phi_{uik}$

❖ Cập nhật ϕ :

$$\phi_{uik} = \frac{\exp \{ \log \beta_{ik} + \psi(\text{shp}_{uk}) - \log(\text{rte}_{uk}) \}}{\sum_{k=1}^K \exp \{ \log \beta_{ik} + \psi(\text{shp}_{uk}) - \log(\text{rte}_{uk}) \}} \text{ if } r_{ui} > 0 \text{ else } = 0 \quad (2)$$

❖ Cập nhật shp, rte:

$$\text{shp}_{uk} = c + \sum_{i=1}^I r_{ui} \phi_{uik} \quad \text{và} \quad \text{rte}_{uk} = d + \sum_{i=1}^I \beta_{ik} \quad (3)$$

Học mô hình.

❖ Thuật toán học kết hợp với dropout(học loại bỏ).

Input: Dữ liệu quan sát E , a , R , các siêu tham số c, d, λ

Output: Ước lượng $W, \beta, \emptyset, shp, rte, \theta$

Repeat

for $u = 1 : U$ **do**

 Loại bỏ ngẫu nhiên tương tác của người dùng u với tỷ lệ loại bỏ dr

endfor

for $i = 1 : I$ **do**

 Cập nhật β_i bằng công thức (1)

endfor

for $u = 1 : U, k = 1 : K$ **do**

 Cập nhật \emptyset_{uik} bằng công thức (2)

 Cập nhật shp_{uk} và rte_{uk} bằng công thức (3)

endfor

 Cập nhật W bằng stochastic gradient ascent.

Until hội tụ

V. Thử nghiệm, đánh giá.

❖ Bộ dữ liệu thử nghiệm.

Bộ dữ liệu	#người dùng	#sản phẩm	#độ dài mô tả
Citeulike	5551	16980	66.6
Movielens – 1M	6040	3681	4.7
Movielens – 10M	69878	10681	5.3
Movielens – 20M	138493	26744	5.5

V. Thử nghiệm, đánh giá.

❖ **Độ đo:** Sử dụng độ chính xác (precision) và độ phủ (recall) theo top-M sản phẩm được dự đoán điểm cao nhất.

$$\text{prec@M} = \frac{1}{U} \sum_u \frac{M_u^c}{M}$$

$$\text{rec@M} = \frac{1}{U} \sum_u \frac{M_u^c}{M_u}$$

- ❖ M_u^c : số lượng sản phẩm đúng trong top – M sản phẩm được gợi ý cho người dùng.
- ❖ M_u : số lượng sản phẩm được tương tác bởi người dùng u trong tập kiểm thử.

V. Thử nghiệm, đánh giá.

- ❖ **Các mô hình so sánh:**

- ❖ **WMF** (Weighted matrix factorization) và **HPF** (hierarchical Poisson factorization) dựa trên phân rã ma trận Gaussian và Poisson nhưng không sử dụng thông tin sản phẩm.

- ❖ **CTMP** (collaborate topic model for Poisson) dựa trên phân rã ma trận Poisson kết hợp với LDA để học biểu diễn nội dung sản phẩm.

- ❖ **Thiết lập tham số:**

- ❖ **WMF** : hệ số cho phân rã Gauss: 0.01

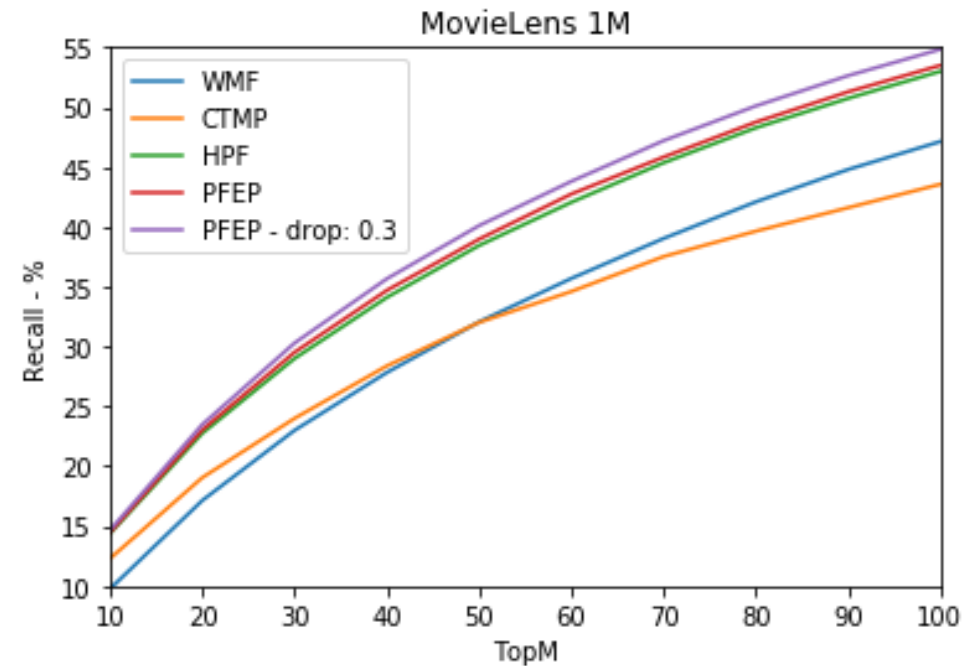
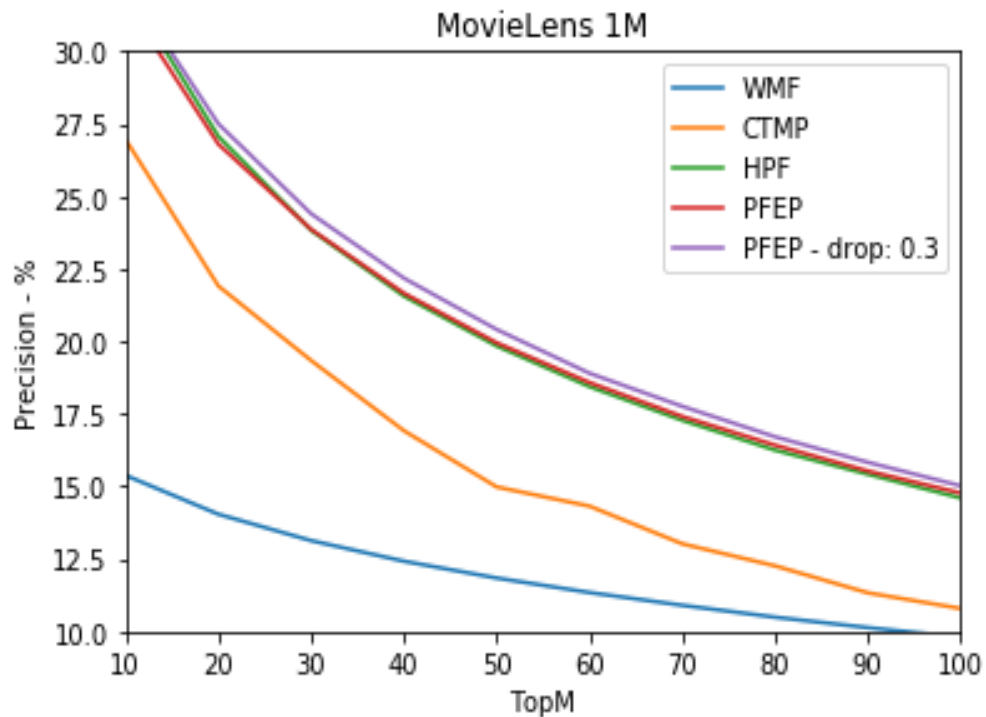
- ❖ **CTMP** : tham số tiên nghiệm cho phân phối Poisson $c = d = 0.3$.

hệ số Gauss cho phần nội dung: 1

- ❖ **HPF** : các hệ số tiên nghiệm đặt $a = a' = c = c' = 0.3$ và $b' = d' = 1$

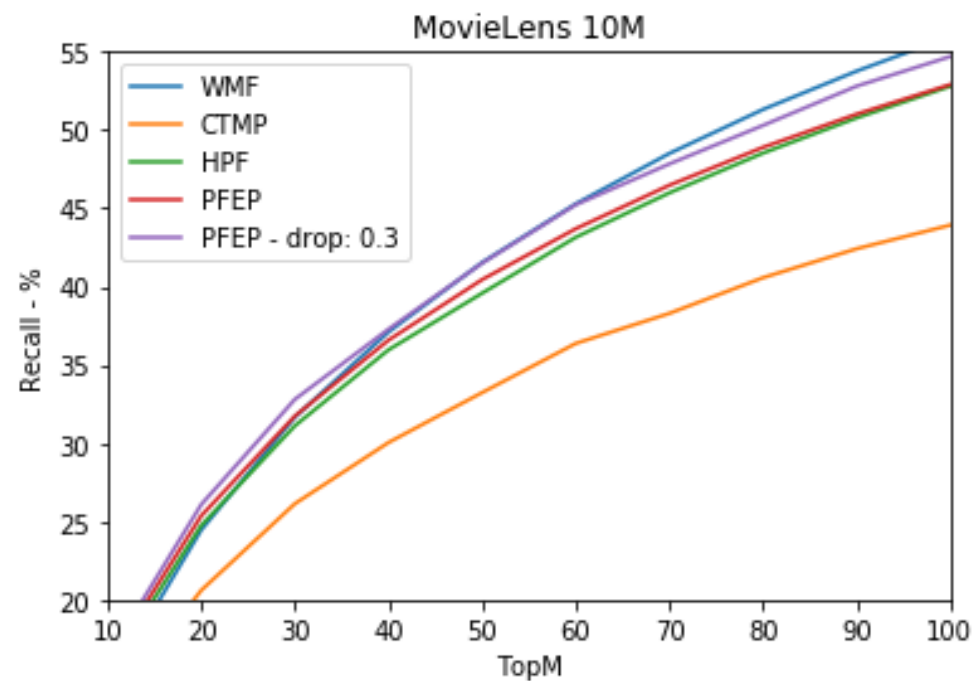
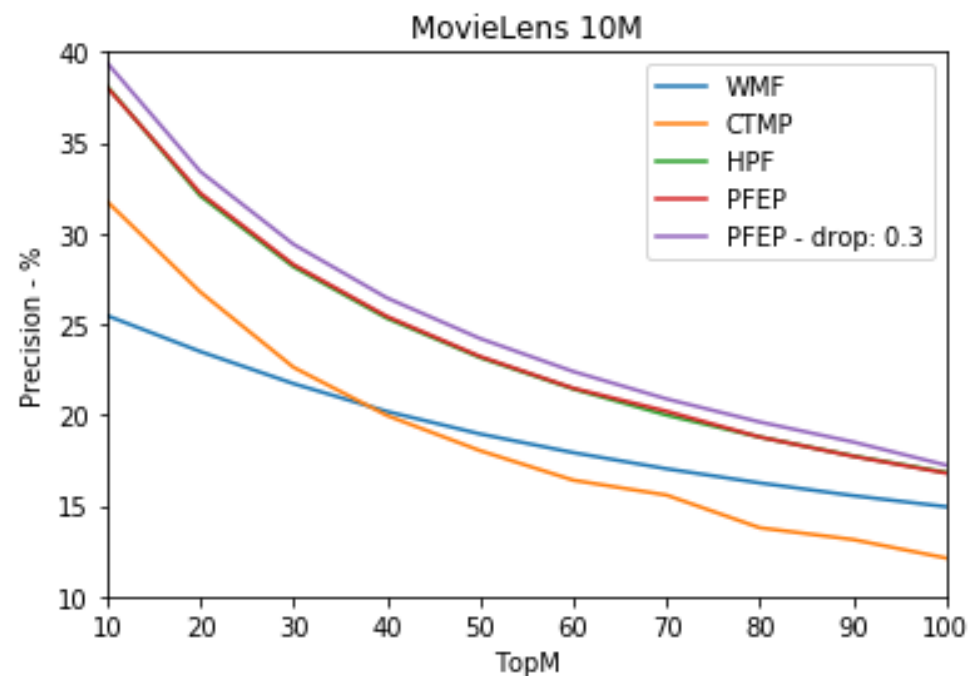
Kết quả với mô tả ngắn

❖ Movielens 1M.



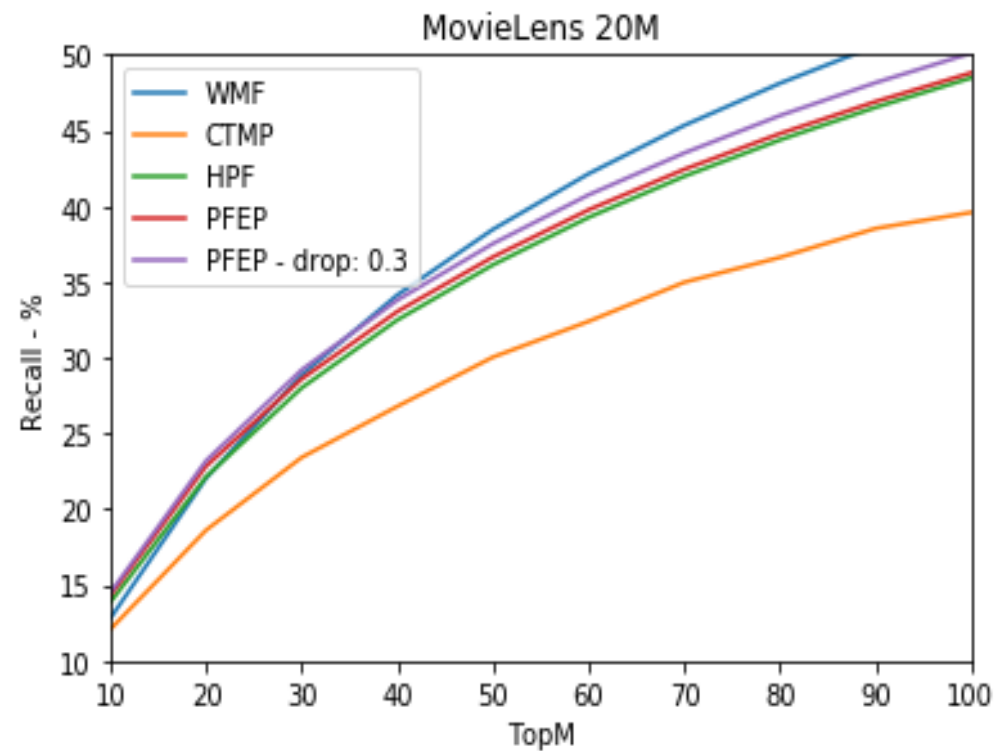
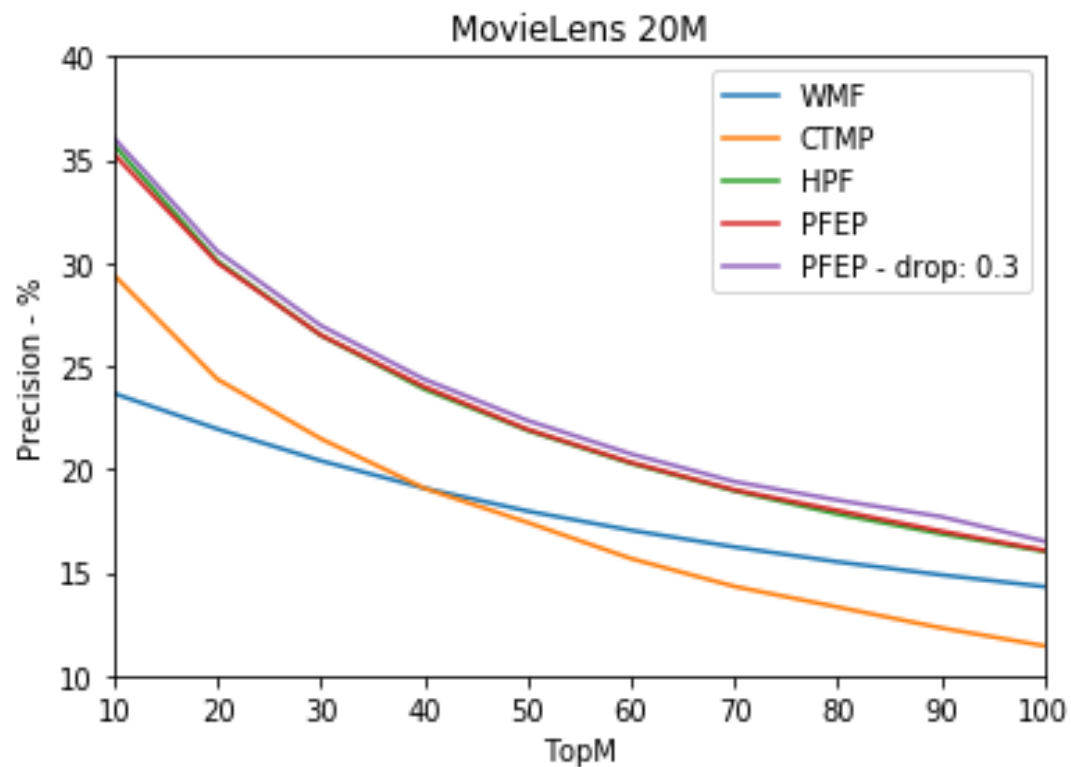
Kết quả với mô tả ngắn

❖ Movielens 10M.



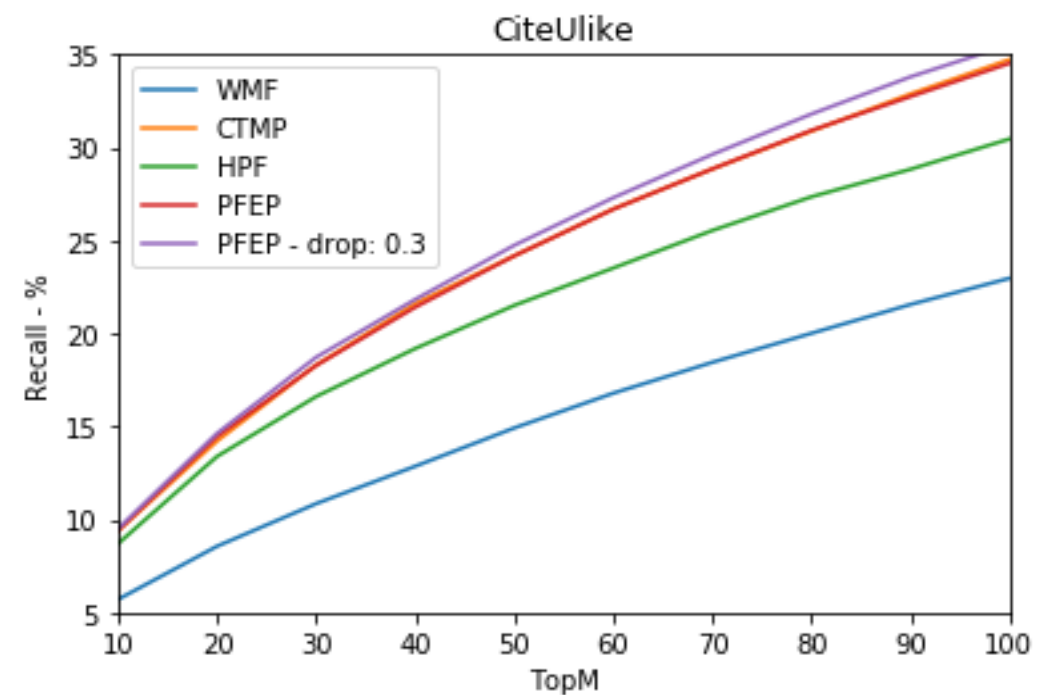
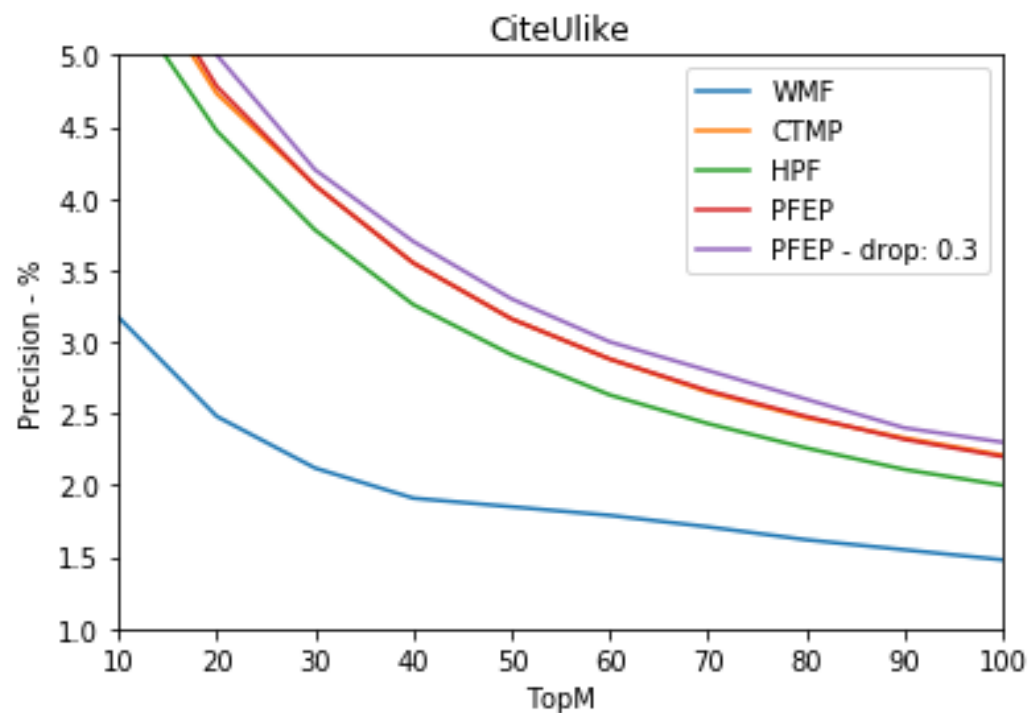
Kết quả với mô tả ngắn

❖ Movielens 20M.



Kết quả với mô tả thông thường.

❖ Citeulike.



PFEP với tỷ lệ dr thay đổi.

❖ Độ chính xác của gợi ý từ top -10 đến top – 100 trên bộ CiteUlike.

	Top 10	20	30	40	50	60	70	80	90	100
PFEP	5.89	4.78	4.09	3.55	3.16	2.88	2.66	2.48	2.32	2.2
PFEP – 0.1	5.84	4.75	4.09	3.55	3.22	2.94	2.71	2.54	2.38	2.25
PFEP – 0.2	5.97	4.91	4.19	3.61	3.24	2.96	2.73	2.55	2.4	2.27
PFEP – 0.3	5.87	4.77	4.11	3.59	3.19	2.9	2.69	2.52	2.38	2.25

PFEP với tỷ lệ dr thay đổi.

❖ Độ bao phủ của gợi ý từ top -10 đến top – 100 trên bộ CiteUlike.

	Top 10	20	30	40	50	60	70	80	90	100
PFEP	9.41	14.47	18.28	21.41	24.17	26.68	28.85	30.92	32.74	34.53
PFEP – 0.1	9.35	14.52	18.51	21.79	24.62	27.09	29.42	31.5	33.44	35.2
PFEP – 0.2	9.45	14.88	18.76	21.72	24.67	27.19	29.4	31.55	33.58	35.19
PFEP – 0.3	9.5	14.63	18.72	21.83	24.74	27.3	29.63	31.8	33.8	35.53

PFEP với tỷ lệ dr thay đổi.

❖ Độ chính xác của gợi ý từ top -10 đến top – 100 trên bộ Movielens 1M.

	Top 10	20	30	40	50	60	70	80	90	100
PFEP	14.46	23.04	29.52	34.69	38.97	42.74	45.83	48.75	51.28	53.49
PFEP – 0.1	14.72	23.58	30.15	35.43	39.84	43.51	46.77	49.6	52.1	54.3
PFEP – 0.2	14.8	23.65	30.21	35.5	39.79	43.64	46.96	49.79	52.34	54.64
PFEP – 0.3	14.66	23.48	30.29	35.66	40.1	43.8	47.2	50.1	52.63	54.83
PFEP – 0.4	14.44	23.27	29.73	34.96	39.44	43.14	46.54	49.49	52.1	54.33

PFEP với tỷ lệ dr thay đổi.

❖ Độ bao phủ của gợi ý từ top -10 đến top – 100 trên bộ Movielens 1M.

	Top 10	20	30	40	50	60	70	80	90	100
PFEP	31.73	26.8	23.86	21.66	19.95	18.58	17.41	16.42	15.53	14.77
PFEP – 0.1	32.15	27.29	24.16	21.94	20.02	18.79	17.61	16.59	15.69	14.92
PFEP – 0.2	32.34	27.44	24.31	22	20.28	18.86	17.66	16.64	15.76	15
PFEP – 0.3	32.36	27.5	24.4	22.18	20.41	18.9	17.76	16.71	15.83	15.02
PFEP – 0.4	31.4	26.76	23.79	21.62	19.94	18.5	17.4	16.42	15.56	14.79

PFEP với tỷ lệ dr thay đổi.

❖ Độ chính xác của gợi ý từ top -10 đến top – 100 trên bộ Movielens 10M.

	Top 10	20	30	40	50	60	70	80	90	100
PFEP	38.05	32.22	28.28	25.43	23.22	21.47	20.02	18.78	17.72	16.79
PFEP – 0.1	38.94	32.24	28.33	25.5	23.3	21.55	20.1	18.84	17.8	16.84
PFEP – 0.3	39.4	33.41	29.4	26.45	24.21	22.39	20.88	19.61	18.51	17.23

❖ Độ bao phủ trên bộ Movielens 10M:

	Top 10	20	30	40	50	60	70	80	90	100
PFEP	15.92	25.36	31.73	36.58	40.45	43.7	46.64	49.07	51.23	53.16
PFEP – 0.1	15.84	25.37	31.85	36.7	40.6	43.84	46.64	49.07	51.23	53.16
PFEP – 0.3	16.32	26.08	32.8	37.28	41.5	45.23	47.84	50.3	52.81	54.7

PFEP với tỷ lệ dr thay đổi.

❖ Độ chính xác của gợi ý từ top -10 đến top – 100 trên bộ Movielens 20M.

	Top 10	20	30	40	50	60	70	80	90	100
PFEP	35.25	30	26.5	24	21.93	20.33	19	18	17	16.07
PFEP – 0.1	35.57	30.11	26.6	24.1	22.09	20.44	19.12	18.1	17.06	16.17
PFEP – 0.2	35.87	30.42	26.79	24.18	22.2	20.59	19.25	18.11	17.14	16.28
PFEP – 0.3	36.04	30.56	26.96	24.36	22.36	20.74	19.4	18.25	17.27	16.41

❖ Độ bao phủ:

	Top 10	20	30	40	50	60	70	80	90	100
PFEP	14.31	22.84	28.6	33.05	36.65	39.76	42.43	44.79	46.9	48.8
PFEP – 0.1	14.27	22.68	28.63	33.2	36.86	40	42.7	45	47.7	49.02
PFEP – 0.2	14.39	23.03	28.92	33.48	37.23	40.4	43.11	45.5	47.63	49.54
PFEP – 0.3	14.5	23.15	29.13	33.75	37.53	40.74	43.51	45.94	48.12	50.1

VI. Kết luận.

- ❖ Các mô hình phân rã Poisson đem lại hiệu quả cao hơn so với các mô hình sử dụng phân rã Gaussian.
- ❖ Sử dụng bộ tri thức tiên nghiệm không hợp lý sẽ làm giảm chất lượng của mô hình.
- ❖ PFEP giúp tăng chất lượng gợi ý đối với cả sản phẩm với mô tả ngắn và mô tả trung bình, nhưng đem lại hiệu quả cao hơn với mô tả trung bình.
- ❖ PFEP kết hợp dropout vừa giúp cải thiện thời gian học và hiệu quả của mô hình.

Hết !!!

CẢM ƠN THẦY CÔ VÀ CÁC BẠN ĐÃ LẮNG NGHE.