IP items recommender

T.G. Hwang

July 24, 2024

Keywords

- data mining
- linear regression
- tags scores
- preferences prediction

Contents

1 IPIRecModel

1.1 Series1

Algorithm 1 IPIRecModelSeries1

Require: No. of top n decisioned tags $NT(u)$, threshold of co-occurred items Freq. θ_I

PREPROCESS($NT(u), \theta_I$)
PROCESS( )

function PREPROCESS($NT(u), \theta_I$)
    top_n_decisioned_tags()
    mean_freq_tags()

function PROCESS( )
    item_based_tags_corr() ▷ *in Equation* (1)
    user_based_tags_corr() ▷ *in Equation* (2)
    tags_score() ▷ *in Equation* (3)

$$sim_U(x, y) = \frac{\sum_{u \in U(x) \cap U(y)} (r(u, x) - \bar{r}(x))(r(u, y) - \bar{r}(y))}{\sqrt{\sum_u (r(u, x) - \bar{r}(x))^2 \sum_u (r(u, y) - \bar{r}(y))^2}} \tag{1}$$

$$sim_I(x, y) = \frac{\sum_{i \in I(x) \cap I(y)} (r(i, x) - \bar{r}(x))(r(i, y) - \bar{r}(y))}{\sqrt{\sum_i (r(i, x) - \bar{r}(x))^2 \sum_i (r(i, y) - \bar{r}(y))^2}} \tag{2}$$

$$s(x, y) = \mathrm{CoOccur}(x, y) \cdot \Pr(x, y) \cdot \frac{1}{|\mathcal{S}|} \sum_{d \in \mathcal{S}} sim_d(x, y) \tag{3}$$

$$\mathrm{CoOccur}(x, y) = \frac{\ln(\min(|I(x) \cap I(y)| + 1,\ \theta_I))}{\ln(\max_{x', y' \in T} |I(x') \cap I(y')|)};\ \theta_I = 20 \text{ (optional)}$$

$$\Pr(x, y) = |I(x)|^{-1} |I(x) \cap I(y)|$$

$$\bar{r}(x) = \frac{\sum_u |I(u) \cap I(V)|}{|U(x) \cap U(V)|} + \frac{\sum_i |U(i) \cap U(V)|}{|I(x) \cap I(V)|}$$

1.2 Series2

Algorithm 2 IPIRecModelSeries2

Require: No. of top n decisioned tags $NT(u)$, threshold of co-occurred items Freq. θ_I

PREPROCESS($NT(u), \theta_I$)
PROCESS( )
POSTPROCESS( )

function PREPROCESS($NT(u), \theta_I$)
    top_n_decisioned_tags()
    mean_freq_tags()

function PROCESS( )
    item_based_tags_corr() ▷ *in Equation* (1)
    user_based_tags_corr() ▷ *in Equation* (2)
    tags_score() ▷ *in Equation* (3)

function POSTPROCESS( )
    append_tendencies() ▷ *in Equation* (4)

$$S \leftarrow S + \mathbb{B}(\mathbb{X}(V), S), \tag{4}$$

$$\bar{b}(\mathbb{X}(V), x, y) = \hat{b}^U(\mathbb{X}(V), x, y) + \hat{b}^I(\mathbb{X}(V), x, y) + \frac{b(\mathbb{X}(V), x) + b(\mathbb{X}(V), y)}{2}$$

$$b(\mathbb{X}(V), t) = \frac{|R(t)|}{|R|} - \mu(\mathbb{X}(V))$$

$$\hat{b}^U(\mathbb{X}(V), x, y) = \frac{\sum_{u \in U(x) \cup U(y)} b(u)}{|U(x) \cup U(y)|}$$

$$b(\mathbb{X}(V), u) = \frac{|I(u)|}{|I|} - \mu(\mathbb{X}(V))$$

$$\hat{b}^I(\mathbb{X}(V), x, y) = \frac{\sum_{i \in I(x) \cup I(y)} b(i)}{|I(x) \cup I(y)|}$$

$$b(\mathbb{X}(V), i) = \frac{|U(i)|}{|U|} - \mu(\mathbb{X}(V))$$

1.3 Series3

Algorithm 3 IPIRecModelSeries3

Require: training set $\mathbb{X}$, test set $\mathbb{Y}$

$\rho = \text{corr}(\mathbb{X}(V))$ ▷ *in Equation (5)*

$\mathcal{B} = \text{tendencies}(\mathbb{X}(V), \rho)$ ▷ *in Equation (6)*

$S = \rho + \mathcal{B}$ ▷ *in Equation (7)*

$$\rho(x,y) = J(x,y)\Pr(x,y)\Pi_{d\in\mathcal{S}} sim_d(x,y) \tag{5}$$

$$J(x,y) = \frac{|I(x)\cap I(y)|}{|I(x)\cup I(y)|}$$

$$\Pr(x,y) = \frac{|I(x)\cap I(y)|}{|I(x)|}$$

$$sim_{UB}(x,y) = \frac{\sum_{u\in U(x)\cup U(y)} |NT(u)\cap x||NT(u)\cap y|}{\sqrt{\sum_u |NT(u)\cap x| \sum_u |NT(u)\cap y|}}$$

$$sim_{UF}(x,y) = \frac{\sum_{u\in U(x)\cup U(y)} |I(u)\cap I(x)||I(u)\cap I(y)|}{\sqrt{\sum_u |I(u)\cap I(x)|^2 \sum_u |I(u)\cap I(y)|^2}}$$

$$sim_{IB}(x,y) = \frac{\sum_{i\in I(x)\cup I(y)} |T(i)\cap x||T(i)\cap y|}{\sqrt{\sum_i |T(i)\cap x|^2 \sum_i |T(i)\cap y|^2}}$$

$$sim_{IF}(x,y) = \frac{\sum_{i\in I(x)\cup I(y)} \tau(x)\tau(y)}{\sqrt{\sum_i \tau(x)^2 \sum_i \tau(y)^2}}$$

$$\tau(t) = |I(t)|^{-1} \sum_{i\in I(t)} |T(i)|$$

$$\bar{b}(x,y) = \hat{b}^U(x,y) + \hat{b}^I(x,y) + \frac{b(x)+b(y)}{2} \tag{6}$$

$$b(t) = \frac{|R(t)|}{|R|} - \mu(\mathbb{X})$$

$$\hat{b}^U(x,y) = \frac{\sum_{u\in U(x)\cup U(y)} b(u)}{|U(x)\cup U(y)|}$$

$$b(u) = \frac{|I(u)|}{|I|} - \mu(\mathbb{X})$$

$$\hat{b}^I(x,y) = \frac{\sum_{i\in I(x)\cup I(y)} b(i)}{|I(x)\cup I(y)|}$$

$$b(i) = \frac{|U(i)|}{|U|} - \mu(\mathbb{X})$$

$$s(x,y) \leftarrow \rho(x,y) + \bar{b}(x,y) \tag{7}$$

2 Estimators

2.1 Series1

Algorithm 4 IPIRecEstimatorSeries1

Require: *iter*

```
ADJUST_TAGS_SCORE(𝕏(V))
for a ∈ [V, L, P] do
    for _ in range(iter) do
        TRAINING(𝕏(a))
function ADJUST_TAGS_SCORE(𝕏)
    W = numpy.ones(|T|, |T|)
    for r(u, i) ∈ 𝕏 do
        r̂(u, i) = estimate(u, i)                          ▷ in Equation (8)
        W = feed_back(r(u, i), r̂(u, i), W)                ▷ in Equation (9)
    S = SW
function TRAINING(𝕏)
    for r(u, i) ∈ 𝕏 do
        r̂(u, i) = estimate(u, i)                          ▷ in Equation (8)
        Ω = feed_back(r(u, i), r̂(u, i), Ω)                ▷ in Equation (10)
```

$$\hat{r}(u,i) = \frac{\sum_x \sum_y \omega(u,x,y)s(x,y)\nu(u,y)}{\sum_{x\in NT(u)} \sum_{y\in T(i)} |\omega(u,x,y)|} \tag{8}$$

$$\nu(u,y) = \begin{cases} NT(u) \cap y = \phi & \text{Default voting} \\ \textbf{otherwise} & 1 \end{cases}$$

$$\epsilon(u,i) = r(u,i) - \hat{r}(u,i)$$

$$w(x,y) = \frac{w + \eta\left(w(x,y)\epsilon(u,i)\lambda^S\gamma^S s(x,y)\right)}{\sum_{z\in T(i)} |w(x,z)s(x,z)|} \tag{9}$$

$$\eta(x) = \begin{cases} x \geq 0 & 2^{-1}x \\ \textbf{otherwise} & 2x \end{cases}$$

$$\mathbb{W}(u) = \sum_{x\in NT(u)} \sum_{y\in T(i)} |\omega(u,x,y)|$$

$$\omega(u,x,y) = \omega(u,x,y)\frac{r(u,i) + \lambda^\Omega s(x,y)}{\hat{r}(u,i) + \gamma^\Omega \mathbb{W}(u)} \tag{10}$$

2.2 Series2

Algorithm 5 IPIRecEstimatorSeries2

Require: *iter*

```
Ω = numpy.ones(|U|, |T|, |T|)
ADJUST_TAGS_SCORE(𝕏(V))
for a ∈ [V, L, P] do
    for _ in range(iter) do
        TRAINING(𝕏(a))
function ADJUST_TAGS_SCORE(𝕏)
    W = numpy.ones(|T|, |T|)
    for r(u, i) ∈ 𝕏 do
        for x ∈ NT(u) do
            for y ∈ T(i) do
                r̂(u, i) = estimate(u, i)                      ▷ in Equation (11)
                W = feed_back(r(u, i), r̂(u, i), W)            ▷ in Equation (12)
    S = SW
function TRAINING(𝕏)
    for r(u, i) ∈ 𝕏 do
        r̂(u, i) = estimate(u, i)                              ▷ in Equation (11)
        Ω = fit_weights(r(u, i), r̂(u, i), S, Ω)               ▷ in Equation (13)
        S = fit_scores(r(u, i), r̂(u, i), S, Ω)                ▷ in Equation (14)
```

$$\hat{r}(u,i) = \frac{\sum_x \sum_y \omega(u,x,y)s(x,y)\nu(u,y)}{\sum_{x\in NT(u)} \sum_{y\in T(i)} |\omega(u,x,y)|} \tag{11}$$

$$\nu(u,y) = \begin{cases} NT(u) \cap y = \phi & DV \\ \textbf{otherwise} & 1 \end{cases}$$

$$\epsilon(u,i) = \ r(u,i) - \hat{r}(u,i)$$

$$s(x,y) = \ s(x,y) + \lambda^S \eta\left(\epsilon(u,i)\Big(\frac{s(x,y)}{\|S\|_2}\Big)\right) + \gamma^S \frac{\bar{s}(x) + \bar{s}(y)}{2} \tag{12}$$

$$\eta(x) = \ \begin{cases} x \geq 0, 2^{-1}x \\ \textbf{otherwise}, 2x \end{cases}$$

$$\omega(u,x,y) = \ \omega(u,x,y)\frac{(r(u,i) + \alpha(\hat{r}(u,i))) + \lambda^\Omega s(x,y)}{(\hat{r}(u,i) + \alpha(\hat{r}(u,i))) + \gamma^\Omega \mathbb{W}(u,i)} \tag{13}$$

$$\alpha(x) = \begin{cases} x = 0 & 1 \\ \textbf{otherwise} & 0 \end{cases}$$

$$\mathbb{W}(u,i) = \ \sum_{x\in NT(u)} \sum_{y\in T(i)} |\omega(u,x,y)|$$

$$s(x,y) = s(x,y)\frac{(r(u,i) + \alpha(\hat{r}(u,i))) + \lambda^\Omega \omega(u,x,y)}{(\hat{r}(u,i) + \alpha(\hat{r}(u,i))) + \gamma^\Omega \mathbb{S}(u,i)} \tag{14}$$

$$\mathbb{S}(u,i) = \ \sum_{NT(u)} \sum_{y\in T(i)} |s(x,y)|$$

2.3 Series3

Algorithm 6 IPIRecEstimatorSeries3

Require: Train.Set. $\mathbb{X}$, main Seq. $\mathcal{A}^{\text{main}}$, post Seq. $\mathcal{A}^{\text{post}}$

$\Omega \leftarrow$ np.ones(shape=($|U|, |T|, |T|$))
$S = \text{ADJUST\_SCORES}(\mathbb{X}(V), S, \Omega)$
for $a \in \mathcal{A}^{\text{main}}$ **do**
  $\mathcal{L}_{\text{min}} = \infty$
  while True **do**
    $\mathbb{L}^{\Omega}, S, \Omega = \text{FIT\_WEIGHTS}(\mathbb{X}, a, S, \Omega)$
    $\mathbb{L}^{S}, S, \Omega = \text{FIT\_SCORES}(\mathbb{X}, a, S, \Omega)$
    $\mathcal{L} = \frac{2\mathbb{L}^S \mathbb{L}^\Omega}{\mathbb{L}^S + \mathbb{L}^\Omega}$ ▷ *in Euqation (18).*
    if $\mathcal{L}_{\text{min}} \geq \mathcal{L}$ **then**
      $\mathcal{L}_{\text{min}} = \mathcal{L}$
      $S_{\text{min}} = S$
      $\Omega_{\text{min}} = \Omega$
    else
      $\mathcal{L} = \mathcal{L}_{\text{min}}$
      $S = S_{\text{min}}$
      $\Omega = \Omega_{\text{min}}$
      break
for $a \in \mathcal{A}^{\text{post}}$ **do**
  if $a \neq V$ **then**
    $\_, S, \Omega = \text{FIT\_SCORES}(\mathbb{X}(V), S, \Omega)$
  $\_, S, \Omega = \text{FIT\_SCORES}(\mathbb{X}(a), S, \Omega)$

function $\text{ADJUST\_SCORES}(\mathbb{X}, S, \Omega)$
  $\_, S, \_\_ = \text{FIT\_SCORES}(\mathbb{X}, V, S, \Omega)$
  $\mathcal{B} = \text{tendencies}(\mathbb{X}, S)$ ▷ *in Equation (6)*
  $S = S + \mathcal{B}$
  $\_, S, \_\_ = \text{FIT\_SCORES}(\mathbb{X}, V, S, \Omega)$
  return S
function $\text{FIT\_SCORES}(\mathbb{X}, a, S, \Omega)$
  if $a \neq V$ **then**
    $S, \Omega = \arg_{S', \Omega'} \min \mathbb{L}(\mathbb{X}(V), S, \Omega)$
  return $\arg_{\mathbb{L}', S', \Omega'} \min \mathbb{L}(\mathbb{X}(a), S, \Omega)$
function $\text{FIT\_WEIGHTS}(\mathbb{X}, a, S, \Omega)$
  $\mathbb{L}, S, \Omega = \arg_{\mathbb{L}', S', \Omega'} \min \mathbb{L}(\mathbb{X}(a), S, \Omega)$
  $\_, S, \Omega = \text{FIT\_SCORES}(\mathbb{X}, a, S, \Omega)$
  return $\mathbb{L}, S, \Omega$

$$\hat{r}(u,i) = \frac{1}{|T(u)|}\sum_{x\in T(u)}\frac{\Pi_{y\in T(i)}\omega(u,x,y)s(x,y)}{\Pi_{y\in T(i)}|\omega(u,x,y)|} \tag{15}$$

$$\begin{aligned}
\epsilon(u,i) =& \; r(u,i) - \hat{r}(u,i)\\
s(x,y) =& \; s(x,y) + \lambda^S\eta\left(\epsilon(u,i)\Big(\frac{s(x,y)}{\|S\|_2} + \frac{\bar{s}(x)+\bar{s}(y)}{2}\Big)\right)
\end{aligned} \tag{16}$$

$$\begin{aligned}
\eta(x) =& \begin{cases} x\geq 0, 2^{-1}x \\ \text{otherwise}, 2x \end{cases}\\
\omega(u,x,y) =& \; \omega(u,x,y)\frac{r(u,i)+\lambda^\Omega s(x,y)}{\hat{r}(u,i)+\gamma^\Omega\mathbb{W}(u,i)}\\
\mathbb{W}(u,i) =& \sum_{x\in NT(u)}\sum_{y\in T(i)}|\omega(u,x,y)|\\
s(x,y) =& \; s(x,y)\frac{r(u,i)+\lambda^\Omega\omega(u,x,y)}{\hat{r}(u,i)+\gamma^\Omega\mathbb{S}(u,i)}\\
\mathbb{S}(u,i) =& \sum_{NT(u)}\sum_{y\in T(i)}|s(x,y)|
\end{aligned}$$

$$\mathbb{L}(\mathbb{X},S,\Omega) = \text{RMSE}(\mathbb{X},S,\Omega) = \sqrt{\frac{1}{|\mathbb{X}|}\sum_{(u,i)\in\mathbb{X}}(r(u,i)-\hat{r}(u,i))^2} \tag{17}$$

$$\begin{aligned}
\mathbb{L}^S(\mathbb{X}(a),S,\Omega) =& \; \min \text{RMSE}^S(\mathbb{X}(a),S,\Omega)\\
\mathbb{L}^\Omega(\mathbb{X}(a),S,\Omega) =& \; \min \text{RMSE}^\Omega(\mathbb{X}(a),S,\Omega)\\
\mathcal{L}(\mathbb{X}(a),S,\Omega) =& \; \frac{2\mathbb{L}^S(\mathbb{X}(a),S,\Omega)\mathbb{L}^\Omega(\mathbb{X}(a),S,\Omega)}{\mathbb{L}^S(\mathbb{X}(a),S,\Omega)+\mathbb{L}^\Omega(\mathbb{X}(a),S,\Omega)}\\
& \min \mathcal{L}(\mathbb{X},S,\Omega)
\end{aligned} \tag{18}$$

2.4 Series4

식 (19)를 사용한 알고리즘 6으로 모델을 훈련함.

$$\begin{aligned}
\mathcal{F}(\mathbb{X}(a),S,\Omega) =& \; \frac{|\mathbb{X}(a)|}{|U|^2|I|}\sum_{u\in U}\Big(\frac{1}{|T(u)||T|}\sum_{x\in T(u)}\sum_{y\in T}(w(u,x,y)s(x,y))^2\Big)^{0.5}\\
\mathbb{L}(\mathbb{X},S,\Omega) =& \; \text{RMSE}(\mathbb{X},S,\Omega)+\mathcal{F}(\mathbb{X}(a),S,\Omega)
\end{aligned} \tag{19}$$