

Laboratorio di Apprendimento Automatico

Fabio Aiolli

Università di Padova

Sistemi di Raccomandazione

- Se guardando i prodotti di Amazon, ti accorgi che il sistema ti suggerisce altri prodotti "correlati", oppure noti messaggi del tipo "chi ha acquistato questo, ha acquistato anche.."
- Se Facebook e Twitter ti raccomandano nuovi amici o persone da seguire in base amici o following attuali.
- Se Netflix ti suggerisce film che potresti essere interessato a guardare basandosi su cosa hanno guardato altri utenti simili a te o cercando di prevedere in base al genere, al regista, ecc. i film che potrebbero essere di tuo gradimento
- Ecc. ecc.
- Allora sotto sotto c'è un sistema di raccomandazione...

Netflix Prize

- Dal 2006 al 2009, Netflix sponsorizzò una famosa competizione offrendo 1.000.000\$ al team che, basandosi su un dataset di oltre 100M di ratings, fosse capace di migliorare anche solo del 10% le prestazioni dell'algoritmo al tempo utilizzato da Netflix.
- R. Bell, Y. Koren, C. Volinsky (2007). ["The BellKor solution to the Netflix Prize"](#)
- Robert M. Bell, Jim Bennett, Yehuda Koren, and Chris Volinsky (May 2009). ["The Million Dollar Programming Prize"](#)

Tipologie

Explicit Feedback se è disponibile:

- Rate degli item su una scala ordinata (stellette).
- Un ordine degli items dal più favorito al meno favorito.
- Preferenza su coppie di item.

Implicit Feedback se è disponibile:

- L'elenco degli items che l'utente ha visionato/comperato.
- Analisi dei tempi di permanenza di un utente in un sito
- La rete sociale di un utente

Approcci

- Content based (CB)
 - Raccomando gli item più simili a quelli per cui l'utente ha già mostrato interesse. P.e. stesso genere (film), stesso autore (song) ecc.
- Collaborative Filtering (CF)
 - Raccomando a un utente gli item più simili a quelli che piacciono ai suoi simili o, viceversa, gli item più simili a quelli che piacciono a lui.
 - Similarità ITEM-ITEM: due oggetti sono simili se tendono ad ottenere lo stesso rate dagli utenti
 - Similarità USER-USER: due utenti sono simili se tendono a dare lo stesso rate agli oggetti
 - Nota che in questo tipo di approcci non si usa conoscenza specifica su oggetti e utenti ma solo caratteristiche del comportamento sociale determinato dall'interazione utenti-items
- Metodi Ibridi
 - $CB > CF$ quando ho poco storico (cold-start problem)
 - $CF > CB$ quando l'informazione implicita contenuta sulla interazione (rete sociale utenti-items) diventa prevalente rispetto a quella esplicita del contenuto
 - In casi intermedi possiamo utilizzare metodi ibridi che combinano entrambi gli approcci

Valutazione Rating Esplicito

$$RMSE = \sqrt{\frac{1}{|R_{te}|} \sum_{(u,i) \in R_{te}} (r_{ui} - \hat{r}_{ui})^2}$$

Root Mean Square Error

Collaborative Filtering

- Matrix Factorization e regressione

$$r_{ui} = \mathbf{x}_u^\top \mathbf{y}_i$$

$$\min_{\mathbf{x}_u} \sum_{i \in R(u)} |r_{ui} - \mathbf{x}_u^\top \mathbf{y}_i|^2 + \beta_u ||\mathbf{x}_u||^2$$

$$\min_{\mathbf{y}_i} \sum_{u \in R(i)} |r_{ui} - \mathbf{x}_u^\top \mathbf{y}_i|^2 + \beta_i ||\mathbf{y}_u||^2$$

- Nearest Neighbors based CF

$$\hat{r}_{ui} = \frac{\sum_{u \in R(i)} \mathbf{w}_{uv} r_{ui}}{\sum_{u \in R(i)} \mathbf{w}_{uv}}$$

$$\hat{r}_{ui} = \frac{\sum_{j \in R(u)} \mathbf{w}_{ij} r_{uj}}{\sum_{j \in R(u)} \mathbf{w}_{ij}}$$

Discesa Gradiente su $\mathbf{x}_u$

$$\min_{\mathbf{x}_u} \sum_{i \in R(u)} |r_{ui} - \mathbf{x}_u^\top \mathbf{y}_i|^2 + \beta_u \|\mathbf{x}_u\|^2$$

$$\min_{\mathbf{x}_u} Q(\mathbf{x}_u) = \sum_{i \in R(u)} |r_{ui} - \sum_s \mathbf{x}_{us} \mathbf{y}_{is}|^2 + \beta_u \|\mathbf{x}_u\|^2$$

$$\nabla Q(\mathbf{x}_{us}) = -2 \sum_{i \in R(u)} \epsilon_{ui} \mathbf{y}_{is} - \beta_u \mathbf{x}_{us}$$

$$\mathbf{x}_{us} \leftarrow \mathbf{x}_{us} - \eta \nabla Q(\mathbf{x}_{us})$$

Discesa Gradiente su $\mathbf{y}_i$

$$\min_{\mathbf{y}_i} \sum_{u \in R(i)} |r_{ui} - \mathbf{x}_u^\top \mathbf{y}_i|^2 + \beta_i \|\mathbf{y}_i\|^2$$

$$\min_{\mathbf{y}_i} Q(\mathbf{y}_i) = \sum_{u \in R(i)} |r_{ui} - \sum_s \mathbf{x}_{us} \mathbf{y}_{is}|^2 + \beta_i \|\mathbf{y}_i\|^2$$

$$\nabla Q(\mathbf{y}_{is}) = -2 \sum_{u \in R(i)} \epsilon_{ui} \mathbf{x}_{us} - \beta_i \mathbf{y}_{is}$$

$$\mathbf{y}_{is} \leftarrow \mathbf{y}_{is} - \eta \nabla Q(\mathbf{y}_{is})$$

Nearest Neighbors

Similarità tra utenti e tra items

$$\mathbf{w}_{uv} = \frac{|I(u) \cap I(v)|}{|I(u)|^{\frac{1}{2}} |I(v)|^{\frac{1}{2}}}$$

$$\mathbf{w}_{ij} = \frac{|U(i) \cap U(j)|}{|U(i)|^{\frac{1}{2}} |U(j)|^{\frac{1}{2}}}$$

$U(i)$ Insieme di utenti che hanno comperato l'oggetto i

$I(u)$ Insieme di oggetti comperati dall'utente u