4. CONTENT-BASED RECOMMENDER SYSTEMS

4.4. LEARNING USER PROFILES AND FILTERING

4.4.3. Bayes Classifier

4.4.3.2. Example of Bayes Model

Table 4.1: Illustration of the Bayes method for a content-based system

Keyword ⇒ / Song-Id ⇓	Drums	Guitar	Beat	Classical	Symphony	Orchestra	Like or Dislike
1	1	1	1	0	0	0	Dislike
2	1	1	0	0	0	1	Dislike
3	0	1	1	0	0	0	Dislike
4	0	0	0	1	1	1	Like
5	0	1	0	1	0	1	Like
6	0	0	0	1	1	0	Like
Test-1	0	0	0	1	0	0	?
Test-2	1	0	1	0	0	0	?

- The user profile, represented by Table 4.1 clearly seems to suggest a preference for classical music over rock music.
- Test set 두 개 중 첫 번째 item만 유저의 취향과 맞을 것으로 보인다. 이를 Bayes approach로 확인해보자.
- 이 예시에서는 Laplacian smoothing을 사용하지 않았는데 실제로는 이런 smoothing 기법을 적용해주는 것이 매우 중요하다.

$$P(c(\overline{X})=1|x_1\dots x_d)=\frac{P(c(\overline{X})=1)\cdot P(x_1\dots x_d|c(\overline{X})=1)}{P(x_1\dots x_d)}$$
$$\propto P(c(\overline{X})=1)\cdot P(x_1\dots x_d|c(\overline{X})=1)$$
$$=P(c(\overline{X})=1)\cdot\prod_{i=1}^{d}p(x_i|c(\overline{X})=1)\ \text{[Naive Assumption]}$$

- 책에 나오는 식

$$P(\text{Like}|\textit{Test-1})\propto 0.5\prod_{i=1}^{6}P(\text{Like}|x_i)$$
$$=(0.5)\cdot\frac{3}{4}\cdot\frac{2}{2}\cdot\frac{3}{4}\cdot\frac{3}{3}\cdot\frac{1}{4}\cdot\frac{1}{3}$$
$$=\frac{3}{128}$$
$$P(\text{Dislike}|\textit{Test-1})\propto 0.5\prod_{i=1}^{6}P(\text{Dislike}|x_i)$$
$$=(0.5)\cdot\frac{1}{4}\cdot\frac{0}{2}\cdot\frac{1}{4}\cdot\frac{0}{3}\cdot\frac{3}{4}\cdot\frac{2}{3}$$
$$=0$$

- normalizing 을 해주기 위해서 P(Like | test-1) + P(Dislike | test-1) = 1이 될 수 있도록 계산해 주면, P(Like | test-1) = 1, P(Dislike | test-1) = 0 이 된다.
- Laplacian smoothing을 사용하면 특정 클래스가 매우 높은 확률을 가지더라도 1,0 처럼 극단적인 확률 값을 가지지 않게 된다.
- > Laplacian smoothing is advisable because a single 0-value in the product-wise form of the expression on the right-hand side of the Bayes rule can result in a conditional probability value of 0.
- Naive Assumption을 적용해서 풀어본 식

$$P(Like|Test-1)\propto P(Like)\cdot\prod_{i=1}^{6}P(x_i|Like)$$
$$=(0.5)\cdot\frac{0}{3}\cdot\frac{1}{3}\cdot\frac{0}{3}\cdot\frac{3}{3}\cdot\frac{2}{3}\cdot\frac{2}{3}$$

???

- 내 생각 - 틀린생각

$$P(Like|x_4) \propto P(Like) \cdot P(x_4|Like)$$
$$= (0.5) \cdot \frac{3}{3}$$
$$P(Dislike|x_4) \propto P(Dislike) \cdot P(x_4|Dislike)$$
$$= (0.5) \cdot \frac{0}{3}$$
$$P(Dislike|x_1, x_3) \propto P(Dislike) \cdot P(x_1|Dislike) \cdot P(x_3|Dislike)$$
$$= (0.5) \cdot \frac{2}{3} \cdot \frac{2}{3}$$
$$= \frac{4}{18}$$

- Laplacian smoohing parameter β >0 를 적용하면 아래와 같다.

$$P(x_i|c(\overline{X}) = 1) = \frac{q^+(x_i) + \beta}{|\mathcal{D}_L^+| + 2 \cdot \beta}$$

4.4.4 Rule-based Classifiers

- 다양한 Rule-based methods 중 Associative classifiers에 대해 알아본다.
- Ch. 3의 3.3에서 다룬적이 있었는데, association rule의 정의와 *support* 와 *confidence* 와 같은 measures 에 대하여 소개하였다.
- Confidence 는 조건(antecedent)을 만족하는 rows에서 결과(consequent)를 만족하는 비율이다.
- "satisfying"의 의미는 예시와 함께 조금 더 구체적으로 알아보자.
- rule의 조건적 요소가 collaborative filtering에서는 item의 rating이었다. 반면에 content based에서는 item을 설명하는 특정 keyword의 존재 유무 이다.

$$\text{Item contains keyword set A} \Rightarrow \text{Rating} = \text{Like}$$
$$\text{Item contains keyword set B} \Rightarrow \text{Rating} = \text{Dislike}$$

- 만약 모든 keyword가 특정 antecedent에 포함된다면 그 antecedent가 만족(satisfy)되었다고 할 수 있다.
- > A row is said to satisfy the consequent of that rule if the rating value in the consequent matches the dependent variable (rating) of that row.

- Table 4.1을 예로 했을때
  {Classical, Symphony} => Like
  위 rule 은 33%의 support 와 100% confidence를 가진다.
- 기본적인 아이디어는 active user로 부터 모든 가능한 rules를 뽑고, user의 interest를 모르는 item에 대하여 어떤 rule이 *fired* 되는지 찾는 것이다.
- > A rule is fired by a target item description if the former's antecedent keywords are included in the latter.
- (Training phase) : Determine all the relevant rules from the user profile at the desired level of minimum support and confidence from the training data set D_L.
- (Testing phase) : For each item description in D_U , determine the fired rules and an average rating. Rank the items in D_U on the basis of this average rating.
- 한 가지 Rule based system의 큰 장점은 high level의 interpretability이다.

4.4.4.1 Example of Rule-based Methods

- Support : The support of an itemset X ⊆ I is the fraction of transactions in T , of which X is a subset.
- Confidence : The confidence of the rule X ⇒ Y is the conditional probability that a transaction in T contains Y , given that it also contains X. Therefore, the confidence is obtained by dividing the support of X ∪ Y with the support of X .

$$
\begin{aligned}
&\text{Rule 1: } \{\text{Classical}\} \Rightarrow \text{Like} \quad (50\%, 100\%)\\
&\text{Rule 2: } \{\text{Symphony}\} \Rightarrow \text{Like} \quad (33\%, 100\%)\\
&\text{Rule 3: } \{\text{Classical, Symphony}\} \Rightarrow \text{Like} \quad (33\%, 100\%)\\
&\text{Rule 4: } \{\text{Drums, Guitar}\} \Rightarrow \text{Dislike} \quad (33\%, 100\%)\\
&\text{Rule 5: } \{\text{Drums}\} \Rightarrow \text{Dislike} \quad (33\%, 100\%)\\
&\text{Rule 6: } \{\text{Beat}\} \Rightarrow \text{Dislike} \quad (33\%, 100\%)\\
&\text{Rule 7: } \{\text{Guitar}\} \Rightarrow \text{Dislike} \quad (50\%, 75\%)
\end{aligned}
$$

- > The aforementioned rules are primarily sorted in order of decreasing confidence, with ties broken in order of decreasing support.

- > ????
  >
  > It is evident that rule 1 is fired by Test-1, whereas rules 5 and 6 are fired by Test-2. Therefore, Test-1 should be preferred over Test-2 as a recommendation to the active user. Note that the rules fired by Test-1 also provide an understanding of why it should be considered the best recommendation for the active user.

4.4.5. Regression-Based Models

- 다양한 regression models 가 있다. 그 중 linear regression 모델에 대해 알아보자.
- $\overline{y} \approx D_L \overline{W}^T$
- objective function O

  $$\text{Minimize } O = ||D_L \overline{W}^T - \overline{y}||^2 + \lambda ||\overline{W}||^2$$
- 위 문제를 풀기 위해서 W에 대한 gradient 가 0이 되도록 한다.

  $$D_L^T (D_L \overline{W}^T - \overline{y}) + \lambda \overline{W}^T = 0$$
  $$(D_L^T D_L + \lambda I) \overline{W}^T = D_L^T \overline{y}$$
  $$\overline{W}^T = (D_L^T D_L + \lambda I)^{-1} D_L^T \overline{y}$$
- unlabeled set D_U 에서 주어진 document vector (item description) $\overline{X}$ 에 $\overline{W}$ 를 dot product해서 rating을 예측할 수 있다.
- Tikhonov regularization은 L2-regularization term $\lambda \cdot ||W||^2$ 을 사용한다.
- L1-regularization을 사용할 수도 있다. Lasso라고도 불리운다. 이는 feature selection의 효과를 가진다.
- > This is because such methods have the tendency to select sparse coefficient vectors for W, in which most components of W take on the value of 0. Such features can be discarded. Therefore, L1- regularization methods provide highly interpretable insights about important subsets of features for the recommendation process.
- 데이터셋이 적을때는 linear model이 더 성능이 좋을 수 있다 왜냐하면 overfitting에 덜 영향을 받기 때문이다.

Table 4.2: The family of regression models and applicability to various types of ratings

Regression Model	**Nature of Rating (Target Variable)**
Linear Regression	Real
Polynomial Regression	Real
Kernel Regression	Real
Binary Logistic Regression	Unary, Binary
Multiway Logistic regression	Categorical, Ordinal
Probit	Unary, Binary
Multiway Probit	Categorical, Ordinal
Ordered Probit	Ordinal, Interval-based