

# PPT 부록

## 1. 라벨링 태그 점수

- 축제  $f$
- 축제  $f$ 의 소개, 설명의 토큰 수  $L_f$
- 태그 집합  $T$ 
  - $T = \{\text{체험, 아동, 문화, 먹거리, 자연, 야간, 계절}\}$
- 태그  $t$ 에 대한 단어장
  - 강한 단어 집합  $V_t^{\text{Strong}}$
  - 약한 단어 집합  $V_t^{\text{Weak}}$
- 축제  $f$ 의 텍스트에서 단어  $w$ 의 등장 횟수  $C_f(w)$
- 강/약 단어 가중치  $\alpha^{\text{Strong}}, \alpha^{\text{Weak}}$

### 1-1. 태그별 룰베이스 점수 $S_{\text{Rule}}(f, t)$

$$S_{\text{Rule}}(f, t) = \frac{\sum_{w \in V_t^{\text{Strong}}} C_f(w) \cdot \alpha^{\text{Strong}} + \sum_{w \in V_t^{\text{Weak}}} C_f(w) \cdot \alpha^{\text{Weak}}}{L_f}$$

## 2. 태그 예측 모델 (TF-IDF + 로지스틱 회귀 OvR)

- 축제  $f$ 의 TF-IDF 벡터
  - $x_f \in \mathbb{R}^d$
- 태그  $t$ 에 대한 로지스틱 회귀 하이퍼파라미터
  - 가중치(계수) 벡터  $w_t \in \mathbb{R}^d$
  - 절편  $b_t \in \mathbb{R}$
- 시그모이드 함수

$$\circ \sigma(z) = \frac{1}{1 + e^{-z}}$$

## 2-1. 태그별 확률

$$P(f, t) = \sigma(w_t^T x_f + b_t)$$

## 2-2. Top-2 태그와 강도

### 2-2-1. 축제 f에 대해 예측 확률이 큰 상위 2개 태그 집합 $T_f$

$$T_f = \text{Top2}_{t \in T} P(f, t)$$

### 2-2-2. 상위 2개 태그에 대한 태그 확률 $\Theta(f, t)$

$$\Theta(f, t) = \begin{cases} P(f, t) & \text{if } t \in T_f \\ 0 & \text{if } t \notin T_f \end{cases}$$

### 2-2-3. 상위 2개 태그에 대한 정규화된 태그 확률 $\theta(f, t)$ (태그 강도)

$$\theta(f, t) = \begin{cases} \frac{P(f, t)}{\sum_{t' \in T_f} P(f, t')} & \text{if } t \in T_f \\ 0 & \text{if } t \notin T_f \end{cases}$$

## 3. 집단별 태그 방문 경향 지표

- 성별 집합  $S$ 
  - $S = \{\text{남성, 여성}\}$
- 연령 집합  $A$ 
  - $A = \{\text{아동, 10대, 20대, 30대, 40대, 50대, 60대, 70세 이상}\}$
- 성별·연령 집단  $G = \{G(s, a) \mid s \in S, a \in A\}$
- 태그  $t \in T$
- 통계가 제공되는 축제 집합  $F_{\text{Stat}}$

- 축제  $f \in F_{\text{Stat}}$ 에 대한 집단  $g$ 의 방문자 비율  $V(f, g)$
- 축제  $f$ 의 태그  $t$  확률  $\Theta(f, t)$

### 3-1. 집단별 태그 방문 비율

#### 3-1-1. 집단 $g$ 가 태그 $t$ 를 가지는 축제에 방문한 가중 합 $N(g, t)$

$$N(g, t) = \sum_{f \in F_{\text{Stat}}} V(f, g) \cdot \Theta(f, t)$$

#### 3-1-2. 전체 집단 기준 태그 $t$ 의 전체 방문 가중 합 $N(*, t)$

$$N(*, t) = \sum_{g \in G} N(g, t)$$

### 3-2. 집단 내 태그 비율과 전체 태그 비율

#### 3-2-1. 집단 $g$ 에서 태그 $t$ 가 차지하는 비율(조건부 확률)

$$P(t | g) = \frac{N(g, t)}{\sum_{t' \in T} N(g, t')}$$

#### 3-2-2. 전체 방문 비율 기준 태그 $t$ 의 비율(주변 확률)

$$P(t) = \frac{N(*, t)}{\sum_{t' \in T} N(*, t')}$$

### 3-3. 집단별 태그 방문 경향 지표

#### 3-3-1. 집단 $g$ 의 태그 $t$ 에 대한 방문 경향 (향상도)

$$L(g, t) = \frac{P(t | g)}{P(t)}$$

- $L(g, t) > 1$ 
  - 집단  $g$ 가 전체 평균보다 태그  $t$ 가 존재하는 축제를 더 많이 방문

- $L(g, t) < 1$ 
  - 집단  $g$ 가 전체 평균보다 태그  $t$ 가 존재하는 축제를 더 적게 방문

## 4. 회원별 축제 추천 점수

- 회원  $u$
- 회원  $u$ 의 성별·연령 집단  $g(u) \in G$
- 회원  $u$ 의 선호 태그  $t_u \in T$
- 회원  $u$ 의 태그 선호 가중치  $w_u(t)$ 
  - $w_u(t) = \begin{cases} 2 & \text{if } t = t_u \\ 1 & \text{if } t \neq t_u \end{cases}$
  - 집단별 태그 방문 경향 지표의 최저(약 0.7)와 최대(약 1.4)의 비율
- 회원  $u$ 가 선택한 지역, 태그, 진행 여부를 만족하는 후보 축제 집합  $\text{Filter}(u, F_{\text{ing}})$
- 현재 진행중 또는 미래에 진행될 축제 집단  $F_{\text{ing}}$
- 축제  $f \in F_{\text{ing}}$ 의 태그 강도  $\theta(f, t)$
- 집단별 태그 방문 경향  $L(g, t)$

### 4-1. 기본 추천 점수 $\text{Score}(u, f)$

$$\text{Score}(u, f) = \sum_{t \in T_f} \theta(f, t) \cdot L(g(u), t) \cdot w_u(t)$$

### 4-2. 정규화된 추천 점수와 최종 추천 목록

#### 4-2-1. 정규화된 추천 점수 $\text{Score}'(u, f)$

$$\text{Score}'(u, f) = \frac{\text{Score}(u, f) - \min_{f' \in F_{\text{ing}}} \text{Score}(u, f')}{\max_{f' \in F_{\text{ing}}} \text{Score}(u, f') - \min_{f' \in F_{\text{ing}}} \text{Score}(u, f')}$$

#### 4-2-2. 회원의 필터값에 따라 임계값 $\tau(0.8)$ 이상 최대 상위 3개의 축제 집합 $F_{\text{rec}}$ 추천

$$F_{\text{rec}} = \{f \in F_{\text{ing}} \mid \text{Filter}(u, F_{\text{ing}}) \cap \text{Score}'(u, f) \geq \tau\}$$

$$0 \leq n(F_{\text{rec}}) \leq 3$$