

定义 : C : course set
 U : user set

Definition 2. **Enrollment Relation**: Let \mathbb{C} denote the set of courses, \mathbb{U} denote the set of users, and the pair (u, c) denote user $u \in \mathbb{U}$ enrolls the course $c \in \mathbb{C}$. The set of enrolled courses by u is denoted as $\mathbb{C}_u \subset \mathbb{C}$ and the set of users who have enrolled course c is denoted as $\mathbb{U}_c \subset \mathbb{U}$. We use \mathbb{E} to denote the set of all enrollments, i.e., $\{(u, c)\}$

$\mathbb{E}[u, c]$ 含所有的 $[u, c]$ 对

feat_extract.py: $X[u, c]$: 包含一个用户在一门课上的学习行为, 以统计值来记录.

Definition 3. **Learning Activity**: In MOOCs, user u 's learning activities in a course c can be formulated into an m_c -dimensional vector $\mathbf{X}(u, c)$, where each element $x_i(u, c) \in \mathbf{X}(u, c)$ is a continuous feature value associated to u 's learning activity in a course c . Those features are extracted from user historical logs, mainly includes the statistics of users' activities.

preprocess.py: $\mathbf{Z}[u, c]$: 矩阵形式包含用户与课程信息.

one-hot encode 对于分类项

- 性别
- 位置
- 教育程度
- 用户集群 [即上述聚类出的 C_1, C_2, \dots, C_6]
- 课程分类

Value : Age.

Definition 4. **Context Information**: Context information in MOOCs comprises user and course information. User information is represented by user demographics (i.e. gender, age, location, education level) and user cluster. While course information is the course category. The categorical information (e.g. gender, location) is represented by a one-hot vector, while continues information (i.e. age) is represented as the value itself. By concatenating all information representations, the context information of a (u, c) pair is represented by a vector $\mathbf{Z}(u, c)$.

Context-Smoothing. augmentation

$$\hat{\mathbf{X}}_g^{(i)} = [[x_i] \oplus g_u(x_i) \oplus g_c(x_i)]$$