By *Jure Leskovec*

STANFORD
UNIVERSITY

SNAP

SNAP for C++
SNAP for Python
SNAP Datasets
BIOSNAP Datasets
What's new
People
Papers
Projects
Citing SNAP
Links
About
Contact us

Open positions

Open research positions in **SNAP** group are available at **undergraduate, graduate** and **postdoctoral** levels.

Social Network: MOOC User Action Dataset

Dataset information

The MOOC user action dataset represents the actions taken by users on a popular MOOC platform. The actions are represented as a directed, temporal network. The nodes represent users and course activities (targets), and edges represent the actions by users on the targets. The actions have attributes and timestamps. To protect user privacy, we anonimize the users and timestamps are standardized to start from timestamp 0. **The dataset is directed, temporal, and attributed.**

Additionally, each action has a binary label, representing whether the user dropped-out of the course after this action, i.e., whether this is last action of the user.

This dataset serves as a recommender system dataset and a dynamic network dataset.

Project website: The dataset have been generated as part of the research project on advanced user modeling and recommender systems. The details of the project can be found here.

Dataset statistics	
Number of users	7,047
Number of targets	97
Number of actions	411,749
Number of positive action labels	4,066
Timestamp	seconds

Source (citation)

- S. Kumar, X. Zhang, and J. Leskovec. Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks. ACM SIGKDD, 2019.

The following BibTeX citation can be used:

```
@inproceedings{kumar2019predicting,
  title={Predicting dynamic embedding trajectory in temporal interaction networks},
  author={Kumar, Srijan and Zhang, Xikun and Leskovec, Jure},
  booktitle={Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery \& D
  pages={1269--1278},
  year={2019},
  organization={ACM}
}
```

The original (unprocessed) dataset is available at: KDD Cup, 2015.

Files

The dataset can be downloaded at act-mooc.tar.gz. It contains the following files.

File	Description
mooc_actions.tsv	Time-ordered sequence of user actions.
mooc_action_features.tsv	Features associated with each action.
mooc_action_labels.tsv	Binary label associated with each action, indicating whether the student drops-out after the action.

Data format

File: mooc_actions.tsv
The data file is in tab separated format.

```
ACTIONID    USERID    TARGETID    TIMESTAMP
```

where

- `ACTIONID`: a unique id for each action.
- `USERID`: a unique id for each user.
- `TARGETID`: a unique id for each target activity.
- `TIMESTAMP`: timestamp for the action in seconds from the beginning.

File: mooc_action_features.tsv
The data file is in tab separated format.

```
ACTIONID    FEATURE0    FEATURE1    FEATURE2    FEATURE4
```

where

- `ACTIONID`: a unique id for each action.
- `FEATUREx`: a feature value associated with the action. Total four in count, making it a 4-dimensional feature vector.

File: mooc_action_labels.tsv
The data file is in tab separated format.

```
ACTIONID    LABEL
```

where

- `ACTIONID`: a unique id for each action.
- `LABEL`: a binary label indicating whether the student drops-out after the action. The value is 1 for drop-out actions, 0 otherwise.

- `ACTIONID`: a unique id for each action.
- `FEATUREx`: a feature value associated with the action. Total four in count, making it a 4-dimensional feature vector.

File: mooc_action_labels.tsv
The data file is in tab separated format.

```
ACTIONID    LABEL
```