A Pseudo Code for Data Generation

Algorithm 1: Generate Data Samples

Input: Raw interactions, data sample templates for recommendation and auxiliary tasks, data_split $\in \{$*Train*, *Valid*, *Test*$\}$, window size w, candidate pool size c

Output: Data samples $\mathcal{D}$

$\mathcal{I} \leftarrow$ a set of unique items (shuffled and mapped to short IDs)
$\mathcal{S} \leftarrow$ a list of chronologically ordered user purchase sequences
$\mathcal{D} \leftarrow \{\}$
for $s \in \mathcal{S}$ **do**
    if *data_split* = *Train* **then**
        $s_{sub} \leftarrow$ all subsequences of the training split of s, each is of length up to w
    if *data_split* = *Valid* **then**
        $s_{sub} \leftarrow$ a subsequence of s that ends with the validation item, proceeding items beyond w are truncated
    if *data_split* = *Test* **then**
        $s_{sub} \leftarrow$ a subsequence of s that ends with the test item, proceeding items beyond w are truncated
    for $ss \in s_{sub}$ **do**
        for *task* $\in \{$*Retrieval, Ranking, Rating prediction*$\}$ **do**
            if *task* = *Ranking* **then**
                $neg \leftarrow$ sample $c-1$ negative items from $\mathcal{I} \backslash s$
            Generate a data sample d with ss, task template, and neg (for *Ranking* only)
            Add d to $\mathcal{D}$
        if *data_split* = *Train* **then**
            for *task* $\in \{$*MIM, MLM, BPR*$\}$ **do**
                if *task* = *BPR* **then**
                    $neg \leftarrow$ sample 1 negative item from $\mathcal{I} \backslash s$
                Generate a data sample d with ss, task template, and neg (for *BPR* only)
                Add d to $\mathcal{D}$
return $\mathcal{D}$