**Collaboratively Tracking Interests for User Clustering in Streams of Short Texts**

**Abstract:**

In this paper, we aim at tackling the problem of user clustering in the context of their published short text streams. Clustering users by short text streams is more challenging than in the case of long documents associated with them as it is difﬁcult to track users’ dynamic interests in streaming sparse data. To obtain better user clustering performance, we propose two user collaborative interest tracking models that aim at tracking changes of each user’s dynamic topic distributions in collaboration with their followees’ dynamic topic distributions, based both on the content of current short texts and the previously estimated distributions. Our models can be either short-term or long-term dependency topic models. Short-term dependency model collaboratively tracks users’ interests based on users’ topic distributions at the previous time period only, whereas long-term dependency model collaboratively tracks users’ interests based on users’ topic distributions at multiple time periods in the past. We also propose two collapsed Gibbs sampling algorithms for collaboratively inferring users’ dynamic interests for their clustering in our short-term and long-term dependency topic models, respectively. We evaluate our proposed models via a benchmark dataset consisting of Twitter users and their tweets. Experimental results validate the effectiveness of our proposed models that integrate both users’ and their collaborative interests for user clustering by short text streams

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows XP/UBUNTU.
* Implementation : NS2
* NS2 Version : 2.28
* Front End : OTCL (Object Oriented Tool Command  Language)
* Tool : Cygwin (To simulate in Windows OS)