

Cím: Ajánló rendszerek; elosztott és online gépi tanulás

Témavezető (név, intézmény, elérhetőség):

Benczúr András (ELTE Opkut. Tsz., MTA SZTAKI, benczur@sztaki.hu) ,

Pálovics Róbert (MTA SZTAKI, rpalovics@ilab.sztaki.hu)

A Netflix Prize verseny óta az ajánlórendszerekkel kapcsolatos kutatások elsősorban a batch, adatokat egyszer végigolvasó módszereket tekintették a legjobb minőségűeknek. Előzetes méréseim alapján sok esetben az adatoknak olyan erős időbeli heterogenitása van, hogy az adatokat egyszer elolvasó és a modellt azonnal frissítő, ún. online learning módszerek sokkal jobban teljesíthetnek. Cél ezeknek a módszereknek kutatása.

Cél továbbá a kontextus alkalmazása online tanuláson alapuló ajánló rendszerekben. A kontextus lehet a napszak, a használt eszköz (telefon, laptop, TV), az időjárás, a közlekedési mód (mobil eszköz esetén). Ilyen adatokat többek között a portugál NMusic zeneszolgáltatótól kaptunk. Felhasználói viselkedés modellezésére pedig az Angry Bird játékot fejlesztő Rovio cég adatait fogjuk használni.

Elosztott adatfolyam alapú gépi tanulás és ajánló módszerek kutatása során cél elsősorban az elosztott data stream feldolgozó keretrendszerek, architektúráik és az általuk biztosított lehetőségek tanulmányozása a Spark és a Flink Streaming keretrendszerekre fókuszálva.

A jelentkezővel szemben támasztott elvárások:

- Programozási gyakorlat Python, C++ vagy Java programnyelveken
- Algoritmusok alaposabb ismerete
- Angol nyelvismeret.

Hivatkozások:

- Pilászy, István, Dávid Zibriczky, and Domonkos Tikk. "Fast als-based matrix factorization for explicit and implicit feedback datasets." *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 2010.
http://www.researchgate.net/profile/David_Zibriczky2/publication/221140868_Fast_als-based_matrix_factorization_for_explicit_and_implicit_feedback_datasets/links/54ea23eb0cf2f7aa4d56afb0.pdf
- Lohrmann, Björn, Daniel Warneke, and Odej Kao. "Nephele streaming: stream processing under QoS constraints at scale." *Cluster computing* 17.1 (2014): 61-78.
<http://arxiv.org/pdf/1308.1031>
- Pálovics, Róbert, et al. "Exploiting temporal influence in online recommendation." *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 2014.
<https://dms.sztaki.hu/sites/dms.sztaki.hu/files/file/2014/recsys.pdf>
- How to factorize a 700 GB matrix with Apache Flink <http://data-artisans.com/how-to-factorize-a-700-gb-matrix-with-apache-flink/>