ΑΝΑΔΕΙΞΤΟ: Εργαλεία Ανάδειξης Τουριστικής Εμπειρίας

T1ΕΔΚ-02474

4η Συνάντηση Εργασίας, 05 Μαρτίου 2021

Ε.Κ. ΑΘΗΝΑ: Παύλος Εφραιμίδης, Γεώργιος Δροσάτος, Σωτήρης Γυφτόπουλος, Γεώργιος Σταματελάτος, Ελένη Μπριόλα, Ανδρέας Σένδρος, Βασίλης Περηφάνης

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Contents

- Administration
- Papers and Reports
- POI Search API, Support for Dotsoft
- PoiGraph paper
- Work in progress:
  - Social Profile Data Collection Apps
    - Twitter App
    - Facebook App
- Federated Learning for Tourism Recommendations
- Yelp Dataset
- Recommendations: Collaborative Filtering
- Recommendations: Federated Learning

Διαχειριστικά

Τρέχον Κατανομή Προϋπολογισμού

- Αρχικός προϋπολογισμός ΕΚ ΑΘΗΝΑ: 150.000€
- Εισροές:
  - 20/07/2018, 60.000€, Προκαταβολή έργου (40%) από ΕΥΔΕ-ΕΤΑΚ
  - 23/07/2020, 29,649.27€, Ενδιάμεση δόση
- Προϋπολογισμός ομάδας έργου με συμβάσεις: 131,198.91€
  - Καθαρά: 122,625.34€ (+ ΦΠΑ 8,573.57€)
- Προϋπολογισμός λοιπών εξόδων: 10,233.55€
  - Εξοπλισμός: 7,256.45€
  - Αναλώσιμα: 534.44€
  - Ταξίδια: 575.56€
  - ΦΠΑ: 1,867.10€
- Εσωτερικός δανεισμός ΕΚ ΑΘΗΝΑ: 63,516.81€
  - Εκ των οποίων, η προμήθεια του φορέα είναι: 17,107.77€
- Εκτιμώμενο τελικό ποσό απορρόφησης: 149,966.66 (99.98%)

Συμβασιούχοι έργου

- Συμβάσεις προσωπικού (με χρονολογική σειρά):
  - Παύλος Εφραιμίδης, 6/8/2018
  - Σωτήριος Γυφτόπουλος, 14/12/2018
  - Γεώργιος Δροσάτος, 01/01/2019
  - Γεώργιος Σταματελάτος, 01/01/2019
  - Ελένη Μπριόλα, 01/01/2019
  - Ανδρέας Σένδρος, 01/11/2020
  - Βασίλειος Περιφάνης, 01/01/2021

Current Publications and Reports

Giorgos Stamatelatos, George Drosatos, Sotirios Gyftopoulos, Helen Briola and Pavlos S. Efraimidis. Point-of-Interest Lists and their Potential in Recommendation Systems. Information Technology & Tourism, 23(1), Springer, 2021 (Online)

Giorgos Stamatelatos, Pavlos S. Efraimidis. About Weighted Random Sampling in Preferential Attachment Models. arXiv preprint, arXiv:2102.08173v1, 16 Feb 2021 (Online)

Antonios Dimitriadis, George Drosatos and Pavlos S. Efraimidis. How much does a zero-permission Android app know about us?. In proceedings of the 3rd Central European Cybersecurity Conference (CECC 2019), ACM, Munich, Germany, 14-15 Nov 2019 (Online)

PoiGraph

Updated Database Data + API

- ~50K POIs, 10 prefectures
- Potential of 1250 POIs / day (insertion + update)
- Majority of POIs only update
- Links with Twitter, Instagram, Tripadvisor
- New APIs for recommendation based on Foursquare Lists
- Part of the dataset is now published

Research Results

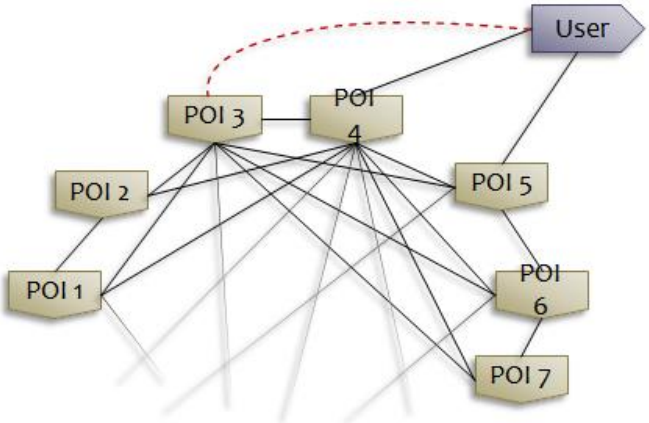
Point-of-Interest lists/Potential recommendation systems

- Novel use of LBSN lists for recommendation
- Lists: collections of semantically related POIs
  - Can capture certain aspects, most obvious temporal, spatial can categorical (e.g. “Best street food in Thessaloniki 2019”)
- Estimate similarities among POIs via the lists
  - Set theoretic, graph theoretic or statistical measures
- Intuition: Recommend the most similar POIs to the POIs that the user finds interesting
- Exact formula is the weighted sum of a POI x with all POIs in the profile

Similarity Measures

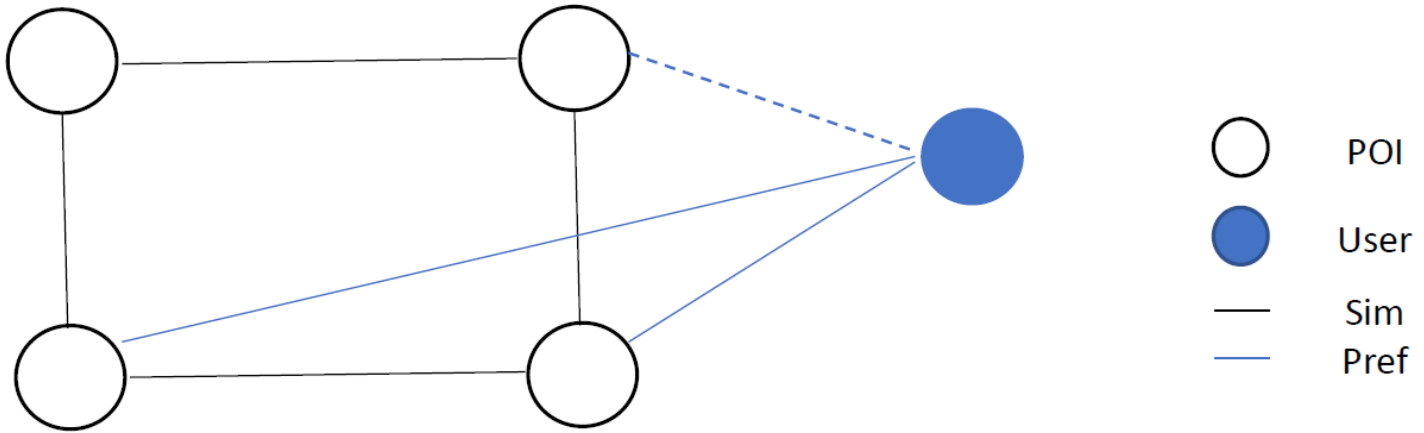
β	Name	Formula	Description
aa	Adamic	$\sum_{t \in N_{11}} \frac{1}{log(n_t')}$	Adamic/Adar index
is	Intersection	n_{11}	Number of common lists
jac	Jaccard	$\frac{n_{11}}{n_{10}+n_{01}+n_{11}}$	Intersection over union
ka	Modified MI	Karagiannis et al. (2015)	Modified MI
ku	Kulczynski-2	$\frac{n_{11}(n_x+n_y)}{2n_x n_y}$	Intersection over harmonic mean
mi	Mutual Information	Manning et al. (2008)	Mutual Information of sets
cos	Ochiai	$\frac{n_{11}}{\sqrt{n_x n_y}}$	Cosine similarity or intersection over geometric mean
ov	Overlap	$\frac{n_{11}}{min(n_x,n_y)}$	Intersection over minimum
ρ	Phi	$\frac{n_{11}n_{00}-n_{10}n_{01}}{\sqrt{n_x*(n-n_x)*(n-n_y)*n_y}}$	Pearson correlation coefficient
sr	SimRank	Jeh and Widom (2002)	Iterative calculation of SimRank
f1	Sorensen	$\frac{2n_{11}}{n_x+n_y}$	Sorensen-Dice index or F1 score or intersection over arithmetic mean

Operation Schema

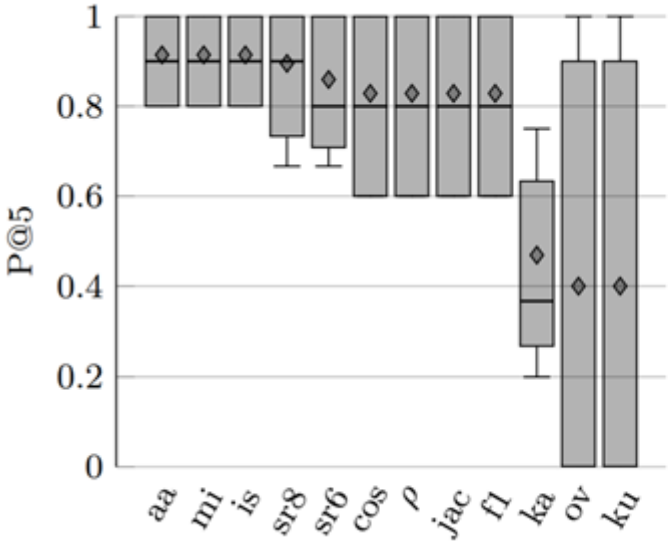


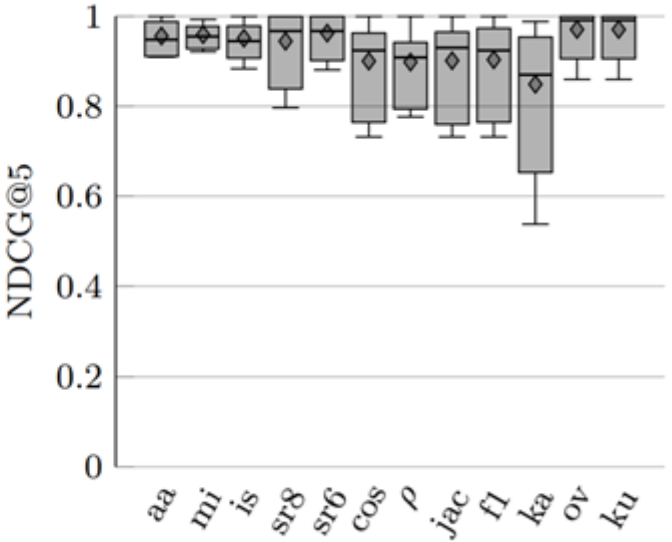
- Pairwise similarities are actually a complete graph.
- Users state their preference and ask the recommender to predict their preferences on other POIs.

Operation Schema



Evaluation


P@5
1
0.8
0.6
0.4
0.2
0
aa
mi
is
sr8
sr6
cos
ρ
jac
f1
ka
ov
ku


NDCG@5
1
0.8
0.6
0.4
0.2
0
aa
mi
is
sr8
sr6
cos
ρ
jac
f1
ka
ov
ku

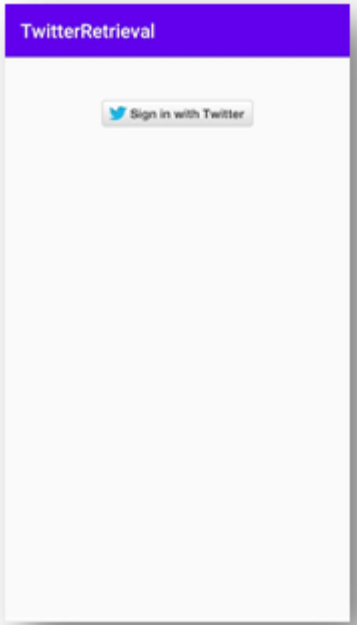
Work in Progress:
Social Profile Data Collection Apps
Federated Collaborative Filtering in Tourism
Yelp dataset

Social Profile Data Collection Apps

Utilizing the User's Social Profile for Recommendations: The TwitterRetrieval App

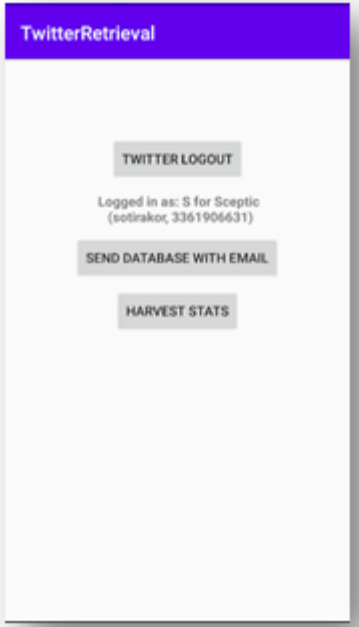
- The aim of the app: to collect data from the user's profile in Twitter in order to be used as input in recommendation algorithms

- The TwitterRetrieval App
  - Main Idea: The user's timeline in Twitter can provide insight about the user's inclinations.
  - A user follows Twitter accounts that reflect his own preferences.



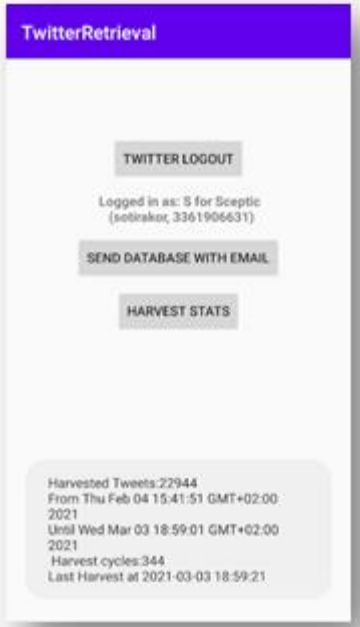
Utilizing the User's Social Profile for Recommendations: The TwitterRetrieval App

- App's features
  - The user logs in to a Twitter account.
  - The app retrieves the recent tweets of his timeline (based on the Twitter API restrictions – 800 tweets)
  - The app is periodically executed to every 15 minutes without the user's intervention to retrieve recent tweets (cumulative database).



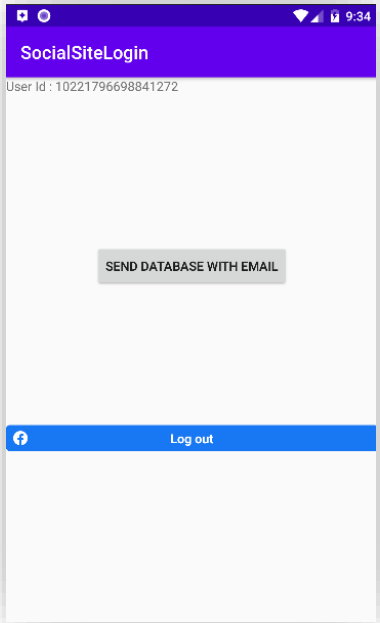
Utilizing the User's Social Profile for Recommendations: The TwitterRetrieval App

- App's features
  - The database is stored only on the user's phone (privacy preserving).
  - Available statistics about the harvest activity of the app.
  - The user can send the collected database via email.
  - Upon logout, the collected data are deleted.



Facebook App

- User Feed
- User Places
- User Likes
- User Events



Database Schema

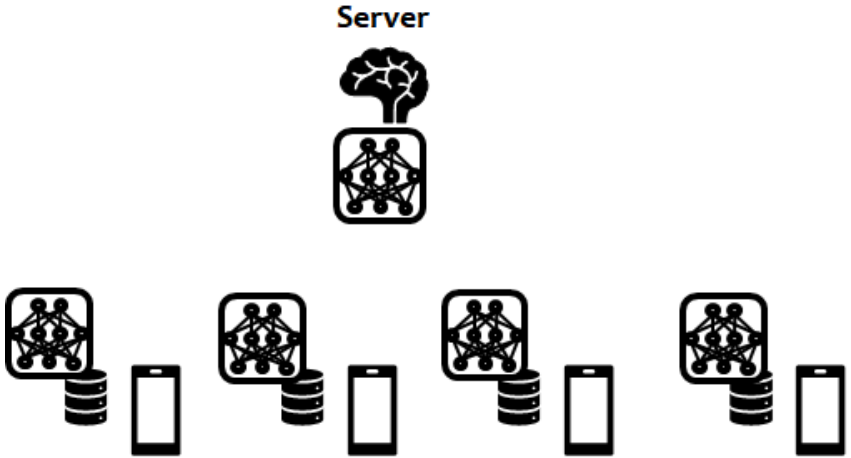
Research

Federated Collaborative Filtering in Tourism

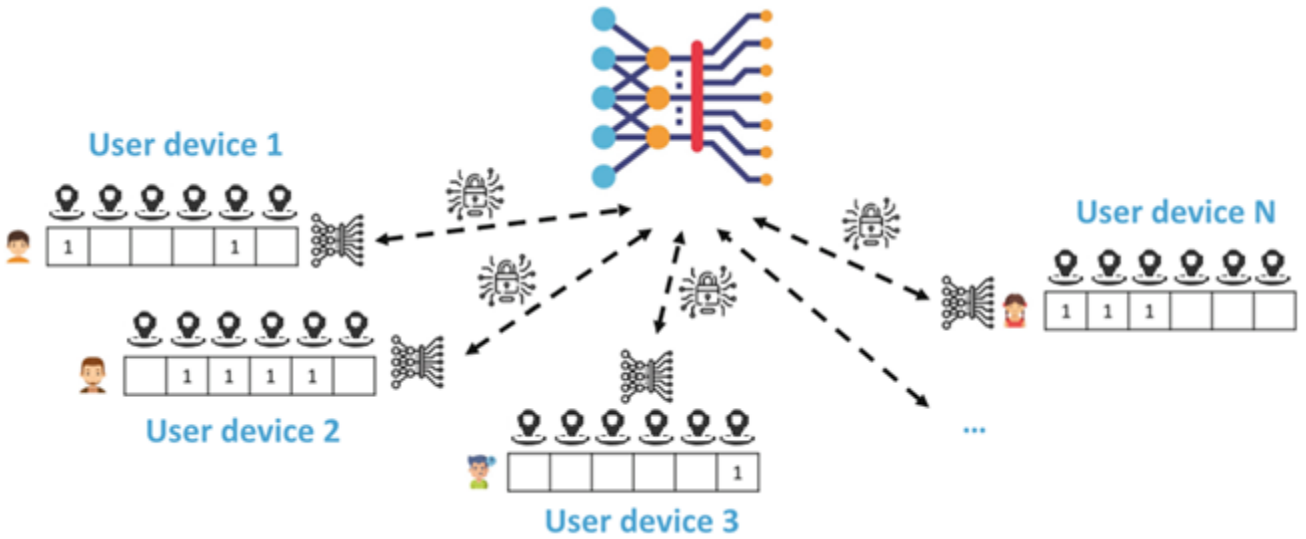
- Distributed Learning using a Federated Learning approach
- Utilize users’ check-ins and friendship from social media
- Personal data of users is stored on their personal devices
- Comparison with centralized approaches to accuracy loss

Federated Learning

- Multiple clients want to collaboratively train a model
- Each client downloads the global model from server
- Clients train the model locally and shares their updates
- The server combines the updates by taking an average



Federated Recommendations Systems



Federated Learning Challenges

- Unbalanced Data
  - Each node may have unequal amount of data
- Limited Communication
  - Devices may go offline unexpectedly
- System Issues
  - Power consumption, data transfer cost
- Privacy Issues
  - Adversary cannot inspect data, but can inspect the global model

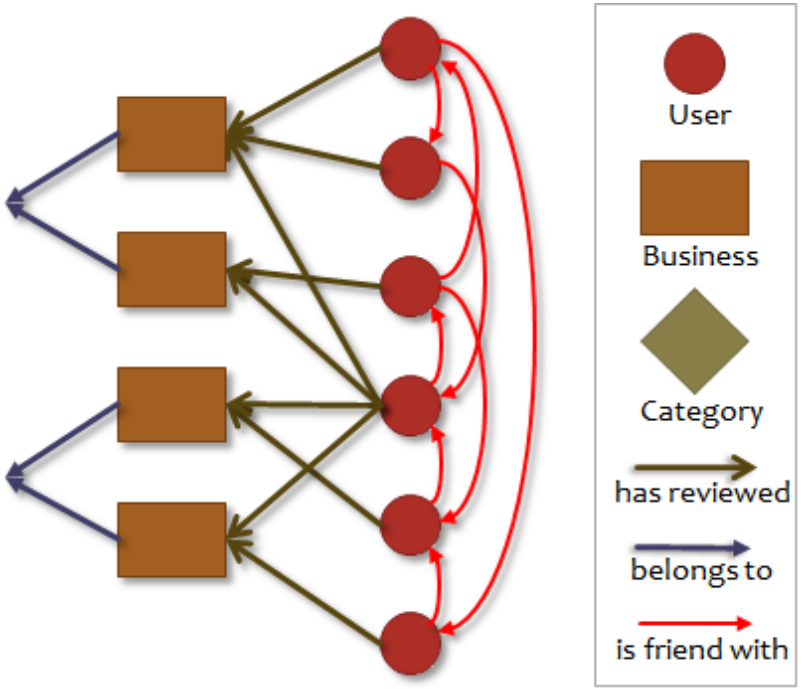
Social Networks as Sources for Touristic Information: The Yelp Case

- Yelp is an online review application for local businesses.
- Worldwide presence in 32 countries and the USA (219 cities)
- Yelp dataset for research:
  - 1.968.703 users
  - 160.585 businesses
  - 8.021.122 reviews
  - 8 metropolitan areas
  - 102.461.451 friendships amongst users
- Our goal: To utilize the dataset in order to develop innovative algorithm for touristic information extraction and recommendation

Social Networks as Sources for Touristic Information: The Yelp Case

- 1st approach:
  - Extract users' preferences from the activity of their friends
    Random Walks

- 2nd approach:
  - Detect business similarities based on users' activity and classification in categories



Επικοινωνία

- Παύλος Εφραιμίδης, Αν. Καθηγητής (pefraimi@athenarc.gr)
- Γεώργιος Δροσάτος, Εντεταλμένος Ερευνητής (gdrosato@athenarc.gr)
- Σωτήρης Γυφτόπουλος, Μεταδιδάκτορας (sgyftopo@athenarc.gr)
- Γεώργιος Σταματελάτος, Υπ. Διδάκτορας (gstamatelat@athenarc.gr)
- Ελένη Μπριόλα, Υπ. Διδάκτορας (elenbrio@athenarc.gr)
- Ανδρέας Σένδρος, Υπ. Διδάκτορας (asendros@ee.duth.gr)
- Βασίλης Περιφάνης, Υπ. Διδάκτορας (vperifan@ee.duth.gr)