

Modelling User Interests in Latent Feature Vector Space based on Document Categorisation

Márius Šajgalík, Supervisor: prof. Mária Bieliková

Overview

Modelling user interests by treating users as categories

Objective is to find discriminative interests

Evaluation based on document categorisation

Evaluated on noisy wild web visit logs

The proposed method of quantitative evaluation

Analogy to text categorisation:

- users represent categories

Several assumptions:

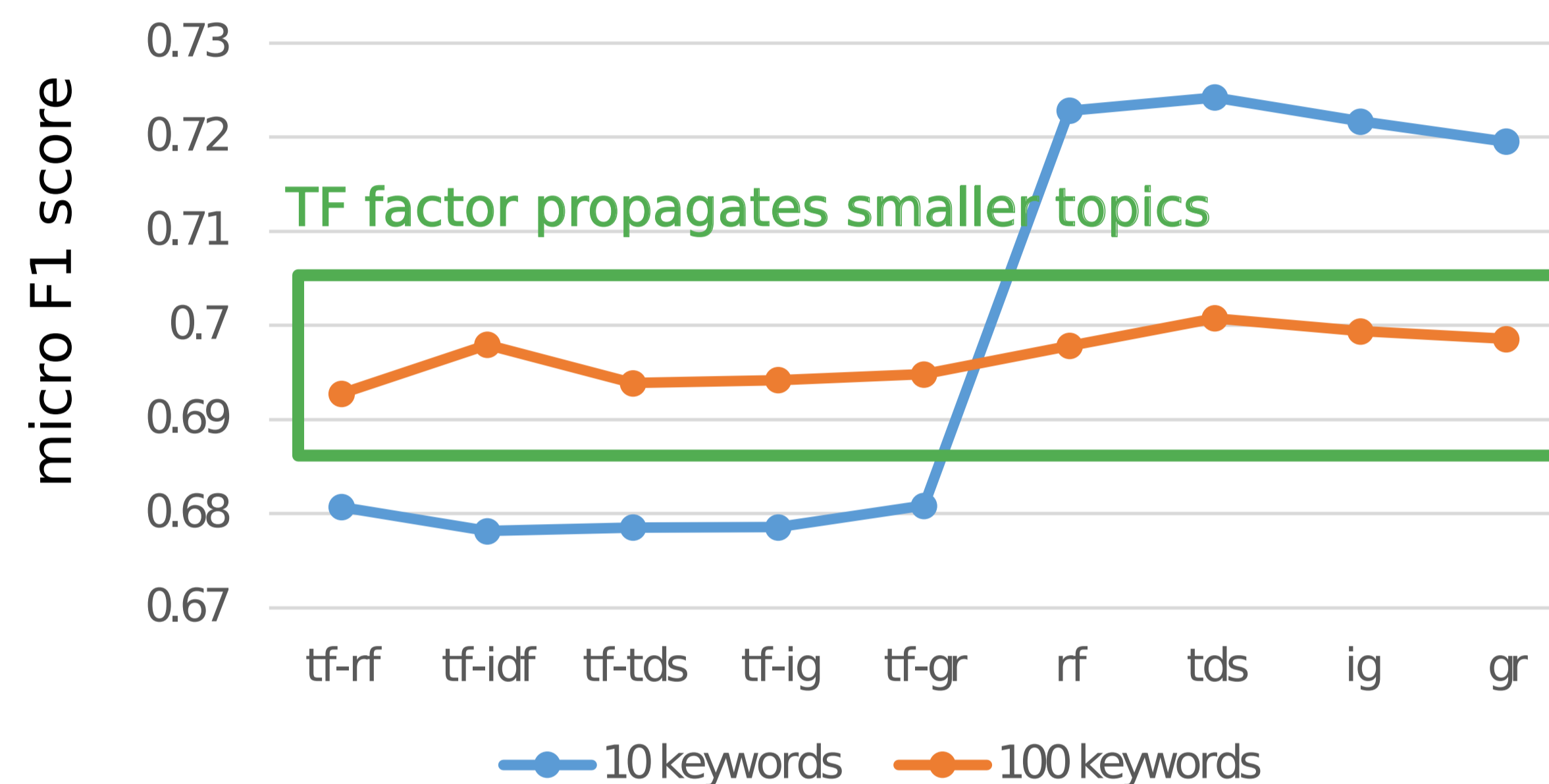
- User is interested in all visited web pages
- We can represent each visited web page by set of keywords, which represent user's interests that caused that page visit

Effect:

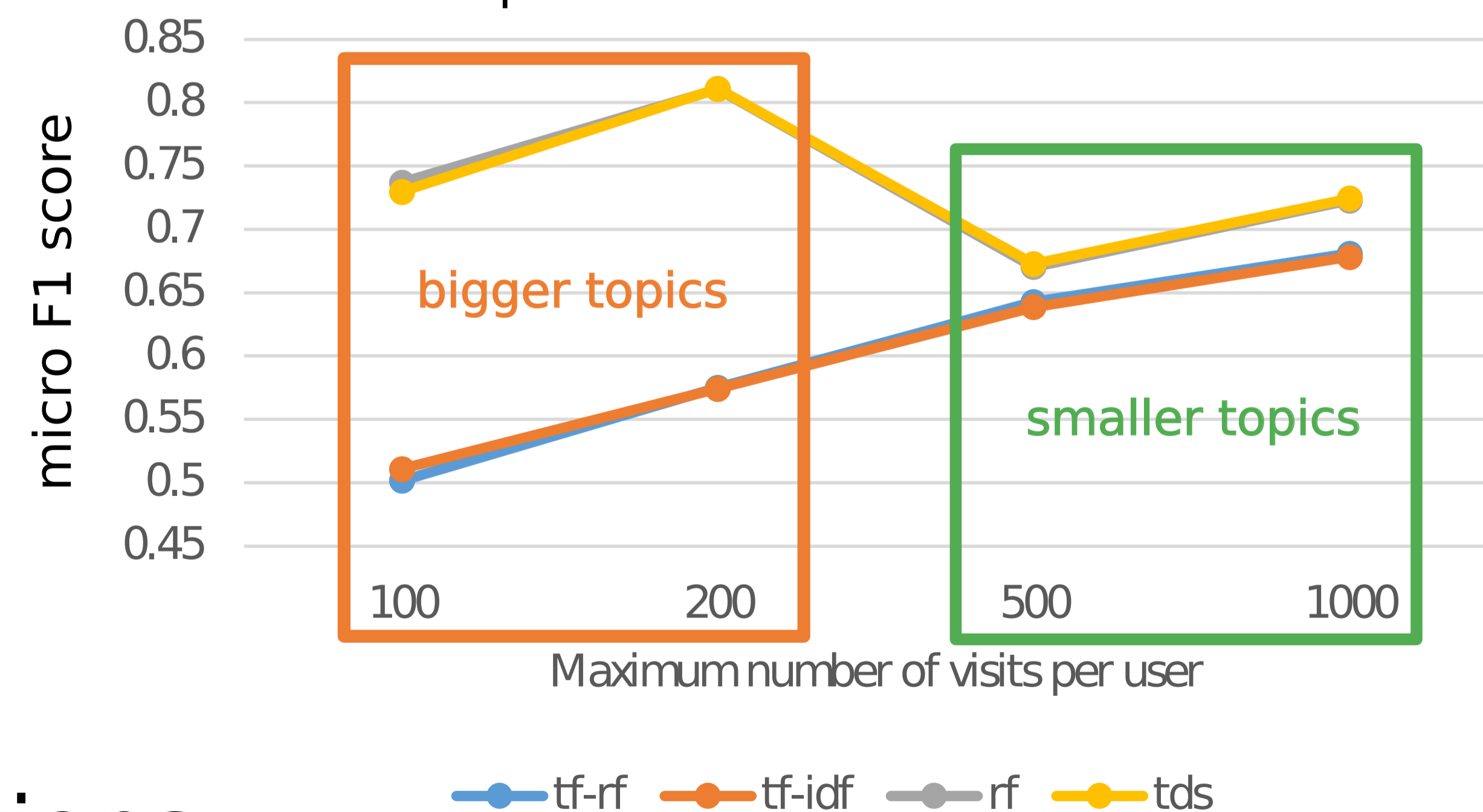
- For each user, we can aggregate personalised keywords (local interests) extracted from visited web pages to compute user's (global) interests

Results

Comparison of different metrics and keyword count



Comparison of different train data size



Contributions

Quantitative evaluation enables faster progress in research
We can analyse also very noisy data