

# Personalized Celebrity Video Search Based on Cross-space Mining

Zhengyu Deng, Jitao Sang and Changsheng Xu  
Institute of Automation, Chinese Academy of Science, Beijing, China  
{zydeng, jtsang, csxu}@nlpr.ia.ac.cn

## Introduction

We propose an Interest-Popularity Cross-space Mining based method to address the multi-domain engaged celebrity video search challenge. For the celebrity side, celebrity popularity is explored by leveraging expert information, e.g., the corresponding Wikipedia homepages. Standard topic modeling method of Latent Dirichlet Allocation (LDA) is adopted to extract the celebrity popularity distribution in abstract topic level. For the user side, we exploit user interests based on his/her online activities, e.g., video sharing, social tagging. LDA is again utilized for user interest topic extraction. Given the derived heterogeneous popularity and interest spaces, we introduce a cross-space correlation method. Semantic and context intra-word relations are refined by random walk to bridge the interest and popularity spaces. The framework of our proposed approach is shown in Fig. 1. The inputs include the celebrities' Wikipedia profile and the users' uploaded and favorite videos with associated tags. The output is the generated video re-ranking list. The framework contains three components, namely interest and popularity space construction, cross-space correlation and video re-ranking. Video re-ranking is based on joint probability distribution of user, celebrity and videos over the interest space.

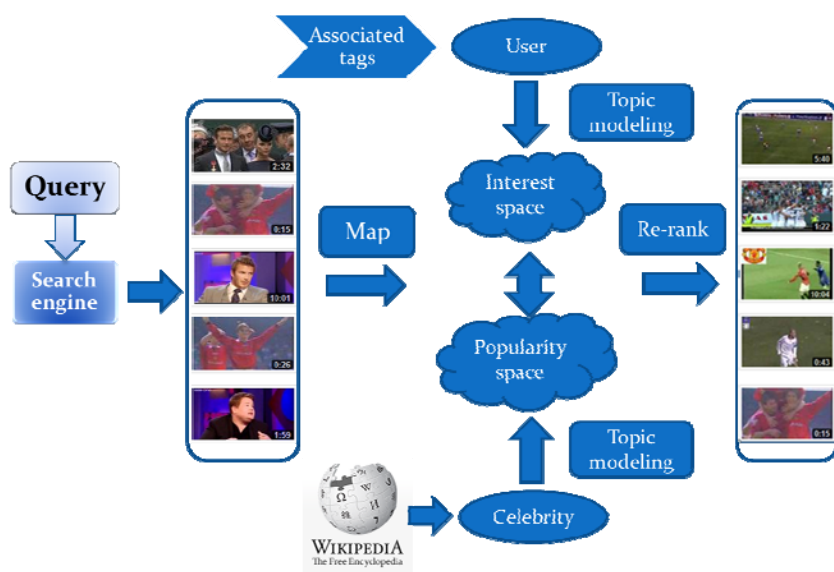


Fig. 1 Celebrity search system architecture.

## Contributions

- We introduce the novel problem of personalized celebrity video search, by exploiting the user interest and celebrity popularity in topic level.
- We propose a cross-space correlation method to connect heterogeneous spaces, which serves as a feasible solution to other cross-domain problems.
- With celebrity as a special case of distributed query, we provide one of the first attempts to address the query understanding challenge in personalized search problem.

Table 1. Parts of the Latent Topics

No	Interest Space	No	Popularity Space
2	movie disney film story	1	season game team sports
3	education book central	4	interview perform romance
5	music cover rock	7	family life dating ceremony
6	interview season party	9	film role series character
8	comedy funny humour	11	album released song awards

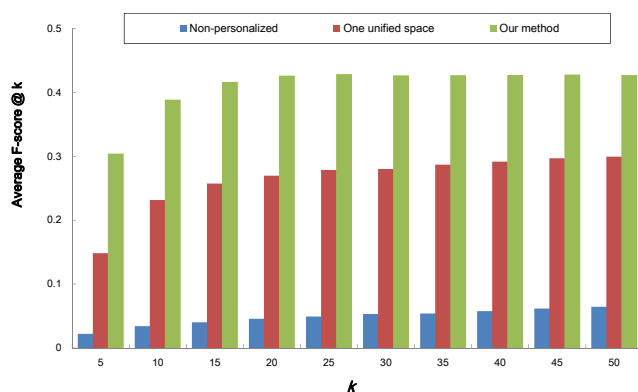


Fig. 2 Different approaches comparing with F-score.

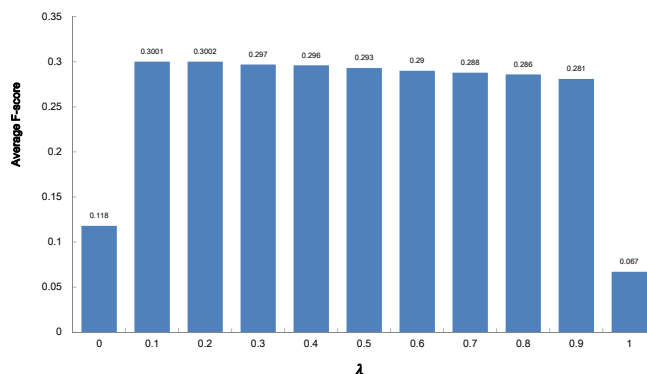


Fig. 3 The influence of random walk.