The Researchers Profile with Topic Modeling

Smail Boussaadi
Asistant professor, PhD student
Laboratory LIMPAF university of Bouira, Algeria
sboussaadi@gmail.com

Dr Hassina Aliane
Director of Information Sciences R&D Laboratory
Head of Natural Language Processing and Digital Content Team.
Cerist,Algiers, Algeria.
haliane@hotmail.com

Pr Ouahabi Abdeldjalil
1.Polytech Tours, Imaging and Brain, INSERM U930, University of Tours, Tours, France
2.University of Bouira, Computer Science Department, LIMPAF, Bouira, Algeria.
ouahabi@univ-tours.fr

2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS) | 978-1-7281-6921-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICECOCS50124.2020.9314588

***Abstract*— Modeling the interests of researchers in academic social networks is a crucial step in a process of recommending scientific articles, linked to their areas of competence and expertise. In this context, a researcher profile constructed from non-observable variables on the basis of articles which interests him by the LDA (Latent Dirichlet Allocation) topic modeling technique allows the system to capture knowledge about his area of competence and skills, in order to predict these needs in terms of relevant research articles. In this article we are interested in the results produced by two different implementations of LDA Gensim and Mallet on the basis of information provided by the researchers (explicit information), in order to compare their interpretability and checked if they are reliable sources for model the areas of competence and expertise of scientists.**

Keywords*— *topic modeling, LDA, Mallet, Genism, researcher profile

I. INTRODUCTION

Modeling of the academic interests of the researcher in scientific social networks allow the personalization process to generate information correlated with his area of expertise. Indeed, to mitigate the problem of cognitive overload and the phenomenon of disorientation faced by any researcher in sharing platforms like Academia or Researchgate, a process of filtering process is based on a fundamental element that is the profile of the researcher. The thematic modeling process consists in extracting and analyzing data relating to the researcher's activity by semantic exploration and extraction of themes on the basis of articles that wrote, downloaded or shared it... to obtain knowledge about his area of expertise and skills [1]. A researcher's interests may represent current topics related to their professional activity or articles related to their field of expertise [2]. Our main contribution in this work is a comparison of the performance of the LDA[3] algorithm for topics profile learning in an academic social network. A research axis between machine learning and automatic language processing (NLP). The objective being to model a researcher, member of a social scientific platform (Researchgate, Academia, ..) by research topics which interests him. To answer the question: are latent themes inferred by topics modeling algorithms such as LDA reliable sources for modeling the area of expertise of a researcher for papers recommendation.?

The document is structured as follows section 2 describes related work on subject inference based on the LDA method recommendation systems. Section 3 describes, the approach and the two implementations of LDA used in this study, section 4 shows the experimental results, as well as an evaluation of the subjects generated. The conclusion and future work are described in section 5.

II. RELATED WORK

This work is related to the most recent work in the field of modelling of academic interests through the use of techniques such as LDA, in the context of scientific recommendation systems. LDA has generated a great deal of research for recommendation systems in general. In [4] the authors grouped together a taxonomy of recommendations or LDA is used to infer the user profile or to represent the different recommendable items

A. LDA and the researcher's profile

Topic modeling based on the LDA method is a very popular technique for semantic exploration and inference of subjects.

In the work that follows; the authors used LDA-inferred latent subjects as reliable sources for modeling researchers' interests or for the representation of articles. In [5], LDA is used to derive the researcher's profile from the articles he has evaluated, the profile thus constructed is a vector where each component is associated with a topic. In [6], to extract the topics of interest from young researchers from their tweets, the authors used the tweeter-LDA algorithm, a variant of LDA, proposed by [7]. For the recommendation of citation and bibliographic references to researchers [8] proposed a combining algorithm, the extraction of latent themes by LDA with matrix factorization. And for its an interactive recommendation system which adapts to the researcher's temporal interests [9] integrated the LDA technique to infer the latent subjects from the researcher's

B. The LDA technique

The LDA technique considers each document as a multinomial distribution on k topics and each topic as a multinomial distribution on terms. And thanks to its flexibility, the method has given rise to numerous works such as the underlying estimation problems in document analysis [10] [11], the categorization and grouping of