Workshop on Monitoring Ph.D. Students Progress, Ph.D. Programme on Information and Communications Technology (Doc_TIC)
03 Jul 2014, University of Vigo, *Spain*

APPLICATIONS OF SOCIAL DATA MINING TO LEARNING ANALYTICS

Kais Dai, Supervised by Dr. Ana Fernández Vilas and Dr. Rebeca P. Díaz Redondo
Affiliation: I&C Lab. of AtlanTIC Research Center, Department of Telematics Engineering (University of Vigo)

Motivation:

- The conjunction of the world's economic crisis problem (which generated a high unemployment rate) with the lake of effective assessment tools to pilot educational policies (by anticipating the job market's needs) raised the expectations' bar from the Learning Analytics research field.
- Furthermore, the emergence of Massive Open Online Courses (MOOCs) as a new learning alternative for long-life learners, their diversity and abundance, highlighted the need of a more personalized learning experience.
- By taking into account the socialization of the web, which is generating big amounts of data (unstructured, heterogeneous, etc.), the informal data sources such as social networks (LinkedIn, Twitter, etc.) must be considered in the Learning Analytics context in order to get more featured view and address its challenges.
- Providing the learning stakeholders the right information in the right moment must be our main goal to optimize the learning environment and to furnish decision support at different steps of the learning process.

Objective:

Our main objective is to take advantage of informal data sources in order to provide a personalized recommendation's system for life-long learners and enhance their career path according to an objective driven approach guided by the job market needs. We expect to apply social mining techniques to merge and process a big amount of data (regarding previous learners' records and curriculums but also their social media fingerprints) and be able to predict non-desirable situations and warn both learners and online courses providers (E-learning platforms, MOOCs) to react properly.

Research Plan

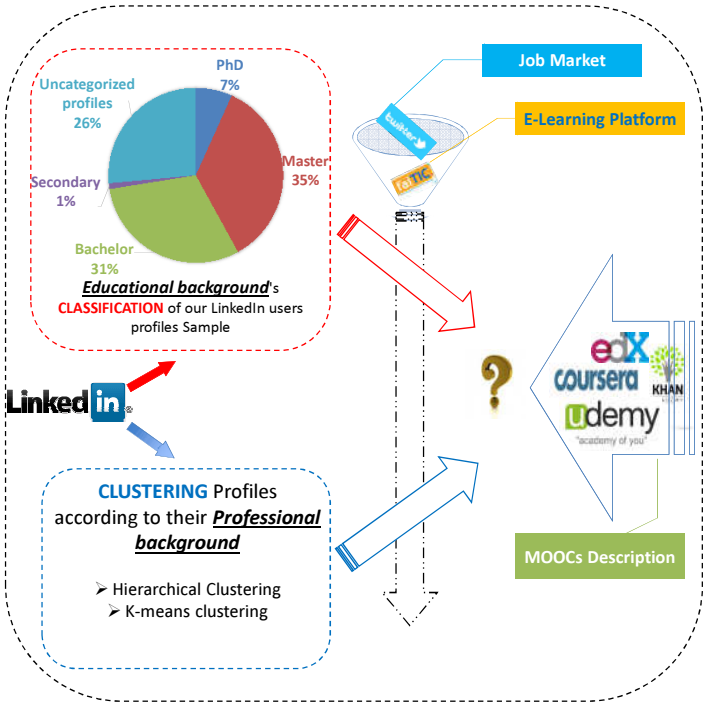
Done:

- ✓ Literature review and knowledge about the environment (Learning Analytics, Educational Data Mining, Big Data, Social Data Mining, Natural Language Processing, Machine Learning, etc.).
- ✓ Data set collection: Design of algorithms to obtain datasets from social media and MOOCs: Twitter APIs (Java, R), LinkedIn API (JavaScript, JSON, PHP, MySQL), LinkedIn public profiles scraping (Python), Coursera, Analysis of "FAITIC" (Learning Management System).
- ✓ Design and Implementation of algorithms based on Natural Language processing techniques to analyse the scraped LinkedIn Data.
- ✓ Text Mining (Application of parallel computing techniques):
  - Classification of the **Educational Background** (according to the UNESCO's education's degree levels)
  - Clustering of LinkedIn profiles According to their **Professional Background**
- ✓ Attendance to the "Big Data and Hadoop" technical course and to the "LASI-Local Madrid" International Workshop, In conjunction with 2nd Learning Analytics Summer Institutes (LASI 2014)

In progress:

- ➢ We expect to submit our results by the end of July and submit a paper in an international conference (Scopus index) and a paper in an international journal indexed in JCR
- ➢ Answer the questions: To which level/How much the educational degrees influence the professional success? Which educational degree to hold best positions? Which courses are the most appropriate to improve a specific professional career?
- ➢ Learning Management Systems analysis & Massive Open Online Courses analysis
- ➢ Discovering job market needs from Twitter data (job market analysis)
- ➢ Data fusion from different sources (LinkedIn, Twitter, MOOCs, LMs) to assist users in career long learning

Preliminary Results



Results Evaluation:

- We are dealing with data from real-world applications, and we use statistical methods to assess our results.
- For validation purposes, we are working on the implementation of a pilot via web applications (using APIs) or e-forms to get users feedbacks.

Next Year Planning

- ❑ Design the merging model for integrating data from all sources (MOOCs, Social Media, Formal E-Learning Platforms) which support the machine learning process to solve the research questions.
- ❑ Attendance to two conferences or workshops.
- ❑ Submitting in at least two international journals (in the JCR index)
- ❑ Short internship in an internationally recognized research group/Company.

References

[1] Rebecca Ferguson: "*Learning analytics: drivers, developments and challenges*", International Journal of Technology Enhanced Learning, Volume 4, Issue 5/6, Pages 304-317, January 2012.
[2] Camilo Palazuelos, Diego García-Saiz, Marta Zorrilla: "*Social Network Analysis and Data Mining : An Application to the E-Learning Context*", Pages 651–660, 2013.