

## **Authorship recognition and disambiguation of scientific papers using a neural networks approach**

*Wednesday, 21 March 2018 16:30 (30 minutes)*

One of the main issues affecting the quality and reliability of bibliographic records retrieved from digital libraries – such as Web of Science, Scopus, Google Scholar and many others – is the authorship recognition and author names disambiguation. So far these problems have been faced using methods mainly based on text-pattern-recognition for specific datasets, with high-level degree of errors.

In this paper, we propose an approach using neural networks to learn features automatically for solving authorship recognition and disambiguation of author names. The network learns for each author the set of co-writers, and from this information recovers authorship of papers. In addition, the network can be trained taking into account other features such as author affiliations, keywords, projects and research areas.

The network has been developed using the TensorFlow framework, and run on recent Nvidia GPUs and multi-core Intel CPUs. Test datasets have been selected from records exported in RIS format from the Scopus digital library, for several groups of authors working in the fields of computer science, environmental science and physics. The proposed methods achieves accuracies above 99% in authorship recognition and is able to effectively disambiguate homonyms.

We have taken into account several network parameters, such as training-set size and batch size, number of levels and hidden units, threshold and weight initialization, back-propagation algorithms, and analyzed the impact on accuracy of results.

This approach can be easily extended to any dataset and any bibliographic records provider.

**Primary authors:** Dr TOMASSETTI, Luca (University of Ferrara and INFN); Dr SCHIFANO, Sebastiano Fabio (University of Ferrara and INFN); Mr SGARBANTI, Tommaso (University of Ferrara)

**Presenters:** Dr TOMASSETTI, Luca (University of Ferrara and INFN); Dr SCHIFANO, Sebastiano Fabio (University of Ferrara and INFN)

**Session Classification:** Humanities, Arts & Social Sciences Session

**Track Classification:** Humanities, Arts, and Social Sciences (HASS) Application