Curriculum Vitae

Name:	**Roberto Pasolini**
Born:	4/3/1986 in Rimini (Italy)
Phone:	+39 333 5265824
e-mail:	roberto.pasolini@unibo.it
Web site:	https://www.unibo.it/sitoweb/roberto.pasolini/
Affiliation:	University of Bologna (Italy) Department of Computer Science and Engineering (DISI)
Work address:	Via Venezia 52, Cesena (FC), 47521 Italy
Work phone:	+39 0541 339210

Research Activity

- Recipient of **Research Grant** at the Interdepartmental Centre for Industrial Research in Health Sciences and Technologies (CIRI SdV) of Alma Mater Studiorum University of Bologna (Italy) from December 2016 to April 2018 working on a project titled “Design di una infrastruttura con elevate capacità computazionali per l’estrazione delle informazioni e l’implementazione di modelli di cura personalizzati” (*Design of a high performance computing infrastructure for data mining and for implementing personalised treatment and service models in healthcare*) under supervision of Prof. Lorenzo Chiari.
- Recipient of **Research Grant** at the Department of Computer Science and Engineering (DISI) of Alma Mater Studiorum University of Bologna (Italy) from March 2015 to September 2016, working on a project titled “Metodi e Algoritmi di Mining per Microdocumenti” (*Mining Methods and Algorithms for Microdocuments*) under supervision of Prof. Gianluca Moro.

Academic Degrees

- **Ph.D. Degree** in Electronic, Computer and Telecommunications Engineering, obtained on April 14th, 2015 at Department of Electrical, Electronic and Information Engineering “Guglielmo Marconi” (DEI) of Alma Mater Studiorum University of Bologna (Italy) with a thesis titled “Learning Methods and Algorithms for Semantic Text Classification across Multiple Domains”.
- **Master’s Degree** in Computer Engineering, obtained on March 23rd, 2011 at the Second Faculy of Engineering (located in Cesena) of Alma Mater Studiorum University of Bologna (Italy) with degree 110/110 with honors, presenting a thesis titled “Sviluppo di Algoritmi di Data Clustering Distribuiti e Applicazioni a Sistemi Decentralizzati per la Sicurezza in Rete” (*Development of Distributed Data Clustering Algorithms and Applications to Decentralized Systems for Network Security*).
- **Bachelor’s Degree** in Computer Engineering, obtained on December 17th, 2008 at the Second Faculy of Engineering (located in Cesena) of Alma Mater Studiorum University of Bologna (Italy) with degree 105/110 presenting a composition titled “Verifica e Simulazione di Sistemi Software col Model Checker Probabilistico PRISM” (*Verification and Simulation of Software*

Systems with the PRISM Probabilistic Model Checker).

Participations to Research Projects

- European H2020 Project TOREADOR (2016-2018, grant agreement No. 688797)
- Italian PRIN Project “Data-Driven Genomic Computing (GENDATA 2020)” [4, 10, 11]
- Italian PRIN Project “Autonomic Security” (2008-2010) [1]

Teaching Activity

- Teacher at the Department of Computer Science and Engineering (DISI) of Alma Mater Studiorum University of Bologna (Italy) of a learning module for the course “Data Intensive Applications”, held by prof. G. Moro from March to June 2018.
- Teaching support activity at the School of Economics, Management and Statistics in Rimini of Alma Mater Studiorum University of Bologna (Italy) for the course “Tecniche e Processi di Data Mining con Appl. Finanziarie” (*Data Mining Processes and Techniques with Financial Applications*) held by prof. G. Moro from October to December 2013, 2014 and 2015.
- Teaching support activity at the Faculty of Statistics in Rimini of Alma Mater Studiorum University of Bologna (Italy) for the course “Data Mining e Supporto alle Decisioni” (*Data Mining and Decision Support*), held by prof. G. Moro from October to December 2012.

Research Areas

During the doctorate course, efforts were focused on the study and application of data mining and machine learning, especially on unstructured textual data, requiring specific steps to extract structured information from it. We first tackled the problem of text categorization by topic, then we addressed sentiment analysis, where user opinions expressed in reviews and social media must be labeled as positive or negative. On one side, we performed an analysis of known techniques for text preprocessing, resulting in the development of new weighting schemes for words in documents [3, 10]. Such weighting schemes have been also employed in genomic computing, for the prediction of unknown associations between genes and functionalities.

On the other end, we dealt with cross-domain classification settings, where knowledge is extracted from a domain and applied to a different one: we developed a couple of novel transfer learning methods for text categorization and sentiment analysis with straightforward and efficient implementations, yielding state-of-the-art accuracy [2, 4, 9, 11, 13].

Other than classification of documents, text mining techniques have been employed in other problems, such as the development of a personalized search system for disambiguation of users’ queries to search engines based on personal interests [5] and the study of correlation between stock market trends and users’ content on social media [6]. Using general machine learning techniques we also obtained interesting results in the detection of network intrusions based on distributed machine learning on traffic statistics [1], the segmentation of job positions from LinkedIn and their recommendation based on users’ skills [8] and the detection of non-line-of-sight propagation in ultrawide-band signals [12].

During the participation to the H2020 TOREADOR project, aimed to developing Big Data analysis services for small and large enterprises, we studied the applicability of cited data mining techniques to

the analysis or very large volumes of data, leveraging specific paradigms, architectures and services for parallel and distributed computing.

Current research is focused on the study and application of deep learning methods, with particular interest in recurrent neural network architectures based on external memory. We tested different kind of network architectures on in-domain and cross-domain sentiment classification with different amounts of data, achieving accuracy improvements over traditional learning methods [7]. Taking inspiration from previous work, we also performed preliminary tests on the application of neural networks to the prediction of unknown gene functionalities, obtaining promising preliminary results with respect to the state of the art.

Technical Skills

- Experience with Linux-based, Android and Windows operating systems and with most common software applications. Basic knowledge of shell scripting in bash.
- Proficiency with Java and Python programming languages and commonly used libraries.
- Basic knowledge of C, C#, Prolog and Scala programming languages.
- Experience with languages and technologies for client-side and server-side Web development: HTML, CSS, JavaScript, PHP, JSP/Servlet, Spring Framework.
- Basic experience in development of mobile applications for the Android and J2ME platforms.
- Proficiency in use of SQL-based relational database engines (PostgreSQL, MySQL).
- Proficiency in statistical analysis and machine learning with R and WEKA.
- Experience with TensorFlow, Theano and Keras software libraries for deep learning.
- Basic experience in administration of Linux-based server machines and of VMWare ESXi virtualization services.

Publications

Journals

1. W. Cerroni, G. Moro, R. Pasolini, M. Ramilli. “Decentralized Detection of Network Attacks Through P2P Data Clustering of SNMP Data”. *Computers & Security* (Elsevier), Volume 52, pp. 1-16, July 2015.

Books and Collections

2. G. Domeniconi, G. Moro, A. Pagliarani, R. Pasolini. Cross-Domain Sentiment Classification via Polarity-Driven State Transitions in a Markov Model. *Knowledge Discovery, Knowledge Engineering and Knowledge Management: 7th International Joint Conference, IC3K 2015, Lisbon, Portugal, November 12-14, 2015, Revised Selected Papers*, 2017.
3. G. Domeniconi, G. Moro, R. Pasolini, C. Sartori. A Comparison of Term Weighting Schemes for Text Classification and Sentiment Analysis with a Supervised Variant of tf.idf. *Data Management Technologies and Applications: 4th International Conference, DATA 2015, Colmar, France, July 20-22, 2015, Revised Selected Papers*, 2016.

4. G. Domeniconi, G. Moro, R. Pasolini, C. Sartori. Iterative refining of category profiles for nearest centroid cross-domain text classification. *Knowledge Discovery, Knowledge Engineering and Knowledge Management: 6th International Joint Conference, IC3K 2014, Rome, Italy, October 21-24, 2014, Revised Selected Papers*, 2015.

Conference Proceedings

5. G. Moro, R. Pasolini, C. Sartori. Personalized Web Search via Query Expansion based on User's Local Hierarchically-Organized Files. *9th International Conference on Knowledge Discovery and Information Retrieval (KDIR)*, 2017.
6. G. Domeniconi, G. Moro, A. Pagliarani, R. Pasolini. Learning to Predict the Stock Market Dow Jones Index Detecting and Mining Relevant Tweets. *9th International Conference on Knowledge Discovery and Information Retrieval (KDIR)*, 2017.
7. G. Domeniconi, G. Moro, A. Pagliarani, R. Pasolini. On Deep Learning in Cross-Domain Sentiment Classification. 9th International Conference on Knowledge Discovery and Information Retrieval (KDIR), 2017. *(candidate to Best Student Paper Award)*
8. G. Domeniconi, G. Moro, A. Pagliarani, K. Pasini, R. Pasolini. Job recommendation from semantic similarity of LinkedIn users' skills. *5th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, 2016.
9. G. Domeniconi, G. Moro, A. Pagliarani, R. Pasolini. Markov Chain based method for in-domain and cross-domain sentiment classification. *7th International Conference on Knowledge Discovery and Information Retrieval (KDIR)*, 2015. *(candidate to Best Paper Award)*
10. G. Domeniconi, G. Moro, R. Pasolini, C. Sartori. A study on term weighting for text categorization: a novel supervised variant of tf.idf. *4th International Conference on Data Management Technologies and Applications (DATA)*, 2015. *(candidate to Best Student Paper Award)*
11. G. Domeniconi, G. Moro, R. Pasolini, C. Sartori. Cross-domain Text Classification through Iterative Refining of Target Categories Representations. *6th International Conference on Knowledge Discovery and Information Retrieval (KDIR)*, 2014. ***(recipient of Best Student Paper Award)***