Kevin Bello

Place and Date of Birth: Lima | 19 April 1993
Address: 610 Curazao Street, La Molina, Lima, Peru
email and webpage: kbellom@uni.pe | kevinbm.github.io

Research Interests

- Big Data
- Domain-Specific Machine Learning:
  - Natural Language Processing: topic modeling, information extraction, information retrieval, sentiment analysis, machine translation.
  - Computer Vision: convolutional neural networks for image recognition, image captioning, dimensionality reduction.
- General Purpose Machine Learning: graphical models, deep learning, kernel methods, neural networks.

Career Goal

To become a well-rounded computer scientist, capable of working with researchers from different disciplines and adequately integrate their specialties to contribute in large scale for betterment of humanity.

Education

Class of 2014-2 — Bachelor of Science in Mechatronics Engineering
National University of Engineering (UNI), Lima, Peru
Summa Cum Laude | Advisor: Dr. Alberto Coronado

Research Experience

Publications

- Cardenas, R., **Bello, K.**, Coronado, A., *et al*, "Panorama of the Market Demand for Mechanical Engineers in South American Countries", *Proceedings of the ASME International Mechanical Engineering Congress and Exposition (IMECE, November 2015, Houston, USA): Volume 5, Engineering Education and Professional Development*.
- **Bello, K.**, Cardenas, R., Coronado, A., *et al*, "Peruvian Labor Market Demand Analysis for Mechanical and Electrical Engineering", *CONIMERA 21st Peruvian Congress 2015, Lima, Peru*.
- Valle A., **Bello, K.**, *et al*, "Analysis of the Peruvian Labor Market Demand in the Area of Mining Maintenance", *Proceedings of the 2nd International Seminar on Mining Plant and Equipment Maintenance (MAPLEMIN), July 2015, Lima, Peru*.

Oral Presentations

- "Labor Market Demand Analysis for Engineering Majors in Peru Using Shallow Parsing and Topic Modeling", *Machine Learning Summer School (MLSS, September 2015), Poster Session, Kyoto University, Japan*.

Projects

Mar. - May 2015 — **Analysis of engineering job advertisements using Topic Models and Shallow Parsing**
Research Assistant | Advisor: Prof. Alberto Coronado | Detailed Information

Sept. - Dec. 2014 — **Recognition and classification of text chunks from job advertisements using Shallow Parsing**
Undergraduate Research Assistant | Advisor: Prof. Alberto Coronado | Detailed Information

May - July 2014 — **Analysis of job advertisements published on South American webpages using visualization techniques**
Undergraduate Research Assistant | Advisor: Prof. Alberto Coronado | Detailed Information

Oct. - Dec. 2013 — **An information retrieval system using Spherical K-means**
Independent Project | Advisor: Prof. Ricardo Rodriguez

Mar. - Sept. 2013 — **Artificial Intelligence and Control Systems Laboratory**
Independent Projects | Advisors: Prof. Guillermo Kemper & Prof. Ivan Calle
Project 1: Localization and path-planning of a synchro and differential-drive robot
Project 2: Fingertip detection using edge detection and connected-component labeling
Project 3: Design, implementation, and control of a 2-DOF robot for real time face tracking

Professional Experience

Empleatron (Start-up company), Lima, Peru
Co-founder, Software Engineer — January 2014 - May 2015
I was in charge of the back-end development of Empleatron–a web-based recommendation system that uses Peruvian job advertisements to suggest users a set of skills to acquire in order to thrive in their desired career path.

Honors and Awards

- Kyoto University's Grant to attend the *Machine Learning Summer School* in Kyoto, Japan — 2015
- Grant from the Research Institute of the Mechanical Engineering Dept. (INIFIM) to conduct research at UNI — 2015
- Top 10 in the Machine Learning Summer School Predictive Modeling Challenge in both tasks the classification and the regression challenge — 2015
- Highest accumulated GPA among all the students in my class(∼60) and top 1 percent in the last 10 years among all students at UNI (∼20,000)
- Dean's list — 2010, 2011, 2012, 2013, 2014
- Honorable Mention (top 5), ACM International Collegiate Programming Contest (ICPC), South American Regional Contest by IBM — 2012, 2014
- Peruvian Council of Science and Technology Grant to attend a summer course for computer science olympiads in Sao Paulo, Brasil — 2013
- Ranked 35 out of 7500 participants in the Worldwide IEEExtreme Programming Competition 7.0 — 2013
- *"Presidente Manuel Pardo y Lavalle Prize"*. It is the highest honor given to undergraduates at UNI — 2012
- University of Chile's Fellowship to participate in the *1st Latin American Theoretical Informatics School* — 2012
- 2nd Place in the national programming competition IEEExtreme - INTERCON, Peru — 2012
- Certificate of recognition for achieving the highest GPA among all the students of the Mechanical Engineering Department (∼800 students) at UNI — 2010

Computer Skills

Programming Languages/Frameworks	C/C++, Python, MATLAB, JAVA, R, Apache Spark, HTML, LaTeX
Operating Systems	LINUX [ArchLinux, Ubuntu], Windows
Databases	MySQL, MongoDB

Other Academic Activities

Researcher at Artificial Intelligence and Control Systems Laboratory (GISCIA) — 2013 - Present
Current president and member of the Artificial Intelligence and Control Systems Laboratory at National University of Engineering.

Course Instructor for Freshman Students at National University of Engineering
As part of the activities of GISCIA, I give lectures about introductory topics to engage students to these fields. Introduction to Probabilistic Robotics (May - July 2015), Introduction to Machine Learning (August - December 2014), Introduction to Algorithms (August 2013 - July 2014).

Active Contestant in Programming Competitions
Solving math and algorithmic problems is one of my hobbies, therefore I continuously participate in online programming competitions such as Google Code Jam, TopCoder, Facebook Hacker Cup, and Codeforces.

Summer Course for Computer Science Olympiads, Universidade Estadual de Campinas (UNICAMP) — January 2013
Attended a three-week course about algorithms at UNICAMP, Brasil. High-quality professors from Europe and South America gave lectures to the best university teams from South America.

1st Latin American Theoretical Informatics School (LATIN) — April 2012
Received a fellowship from the University of Chile to attend the 1st LATIN school that consisted of lectures by faculty of the Massachusetts Institute of Technology (USA), Universitat Politecnica de Catalunya (Spain), and Universidad Nacional Autonoma de Mexico (Mexico). I also attended some Plenary Lectures and Special Sessions of the LATIN Symposium.

Languages

- English: Advanced (TOEFL iBT: 102)
- Spanish: Native

Relevant Courses

National University of Engineering

Artificial Intelligence, Statistics and Probability, Linear Algebra, Multivariable Calculus, Numerical Methods, Digital and Electronic Circuits, Multi-body Dynamics, Data Communication and Networking, Physics, Biology for Engineers, Differential Equations, Complex Variable.

MOOCs Web based courses (e.g. Coursera, EdX), scored within 90% and 100%

Convex Optimization (S. Boyd), Statistical Learning (Hastie & Tibishirani), Neural Networks for Machine Learning (G. Hinton), Machine Learning (A. Ng), Probabilistic Graphical Models (D. Koller), Learning from Data (Y. Abu-Mostafa), Natural Language Processing (C. Manning & D. Jurafsky, M. Collins), Algorithms I & II (T. Roughgarden), Programming a Robotic Car (S. Thrun)

Hobbies

- Soccer: silver medal, Interschool Championship, Junin, Peru, 2008.
- Karate: bronze medal, Peruvian National Championship, Regional Phase, 2005.
- Amateur pianist

Personal References

Professor Alberto Coronado — macoronado@uni.edu.pe
Department of Mechanical Engineering
National University of Engineering

Professor Julio Urbina — jurbina@engr.psu.edu
Department of Electrical Engineering
The Pennsylvania State University

Professor Ricardo Rodriguez — robust@uni.edu.pe
Department of Mechanical Engineering
National University of Engineering

Project Summary

Analysis of engineering job advertisements using Topic Models and Shallow Parsing

Goal: Reconceptualize the classification of specialization programs (e.g. diploma) using unstructured text from engineering job advertisements.
Significance: Reveal latent connections among the majors with highest demand in the Peruvian labor market. Aiming to propose a new integration of specialization programs offered by Peruvian universities so that future professionals can be more competitive.
Methods: Six mixed-membership models were compared in order to assess the effect over model dimensionality, all of the models used Latent Dirichlet Allocation (LDA). The data source and the learning algorithm varied for each model. For the case of data source, two variations, one using the whole text, and the other using text only from functions and requirements of the job advertisement (a classifier of a previous work was used. See the description of the project below). And for the case of the learning algorithms, three variations. Two using Variational Expectation-Maximization, one with the hyperparameter being estimated and the other fixed, and the remaining with Gibbs Sampling. The 10-fold cross-validation method was used for model selection in order to find the optimal number of latent categories. The six mixed-membership models were implemented using R and C++.
Results: The analysis revealed several major clusters, which means that majors in the same cluster are required almost indistinctly, and that specializations converge to a mixture of common pools of knowledge and required skills. This suggests that some specialization programs should be a mixture of more than two different fields. Currently, specialization programs in Peru combine at most two different fields.

Recognition and classification of text chunks from job advertisements using Shallow Parsing

Goal: Recognize and classify text chunks such as functions, requirements, majors, and so on from job advertisements published in Peruvian websites.

Significance: After developing the classifier and running over hundreds of thousands job advertisements, the results will shed light the requirements and functions with highest demand for each major. Thus knowing these statistics will make an impact in the study programs taught by educational institutions in Peru.
Methods: The pre-processing step consisted of Natural Language Processing techniques such as regular expressions, word tokenization, and sentence segmentation. Then the problem was reduced to a sequence labeling problem (Shallow Parsing) in which the base phrases defined were: *Major* (major required for the position, Noun Phrase), *Requirements* (list of skills the applicant needs to have, Noun Phrase and Verbal Phrase), and *Functions* (list of tasks the applicant will have to accomplish, Verbal Phrase). A generative model using Hidden Markov Models and a discriminative model using Averaged Structured Perceptron were compared.
Results: The results were close to the state-of-the-art performances in the Spanish language, with a 87.1% F1 score for the *Major* predictions.

Analysis of job advertisements published on South American webpages using visualization techniques

Goal: Perform a descriptive analysis of online South America job advertisements using visualization techniques.
Significance: Providing statistics about the majors with highest demand can help people to get a notion of the current state of industry. Furthermore, people could observe and decide what career paths are more convenient for them to pursue.
Methods: A web crawler using the Scrapy Python framework was implemented. In addition, the job advertisements were stored in a SQL-model database. Also, a detector (to determine the majors required in the advertisement) was developed using regular expressions. The eighty most popular majors were considered in the detection phase. The statistics were shown through an interactive visualization by using the Javascript D3 library.
Results: The detector system achieved a precision of 96% in average. Analysis of the results are explained in the website www.empleatron.com (Spanish). A Treemap showing the distribution of the labor demand, and a Circleplot showing the relationship among the majors are presented in the webpage.