

Jorge Guevara

Ph.D.

100, Rua João Sbarai, apt 7
São Paulo, Brazil, CEP 05587-060

☎ (+55) 11 93393685

✉ jorgegd@br.ibm.com

🌐 www.vision.ime.usp.br/~jorjasso



Education

- 2010–2015 **PhD in Computer Science**, *Institute of Mathematics and Statistics, University of Sao Paulo*, Brazil.
- 2012–2013 **PhD internship at Laboratory for Computer Science, Information Technology and Systems**, *INSA de Rouen, University of Normandy*, France.
Under advise of professor Dr. Stephane Canu.
- 2007–2009 **Master in Computer Science**, *National University of Trujillo*, Peru.
- 2005 **Informatics engineer**, *Faculty of Physical and Mathematical Sciences, National University of Trujillo*, Peru.
- 2000–2004 **Bachelor in Computer Science**, *Faculty of Physical and Mathematical Sciences, National University of Trujillo*, Peru.

Doctoral Thesis

- Title *Supervised Machine Learning using kernel methods, probability measures and fuzzy set theory*
- Supervisor Professor Roberto Hirata Junior, University of Sao Paulo, Brazil.
- Internship Professor Stephane Canu, INSA-ROUEN, France.
- Supervisor
- Description This thesis explored the design and evaluation of machine learning algorithms on training sets of point-sets, i.e. each observation is itself a set of multidimensional points. We modeled each point-set either as a realization of a fuzzy random variable or a realization of a random probability measure. We introduced the novel concept of kernel on fuzzy sets and developed new algorithms to deal with such data.

Masters Thesis

- Title Speech Recognition Framework for Information Retrieval
- Supervisor Professor Ronald Leon
- Description We explored speech recognition techniques for text retrieval from speech audios in spanish language. This work was focused in the application of state-of-art speech recognition techniques on spanish sentences with the aim of perform speech information retrieval.

Informatics engineering thesis

Title Feature Extraction using Wavelets for Speech Recognition
Supervisor Professor Ronald Leon
Description We developed an algorithm for feature extraction from speech audio based on wavelet theory. This new feature extractor was experimentally evaluated on isolated speech recognition.

Bachelor graduate project

Title A Speech Recognition system for vocals recognition using a feedforward neural net.
Supervisor Professor Ronald Leon
Description We performed experiments with a feed-forward neural net with back-propagation algorithm to classify vocal sounds.

Professional Experience

- 2016–present **Researcher scientist**, IBM-RESEARCH, Sao Paulo -Brazil.
I am a researcher scientist within the Natural Resources Analytics group at IBM Brazil.
- 2014–2016 **Postdoc**, *Fapesp grants nro 2014/06549-7 and nro 2015/11693-2*, e-Science Models and methods for life and agricultural sciences. Software Manager, eSCIENCE LABORATORY -IME USP, Sao Paulo -Brazil.
- Teaching Experience as Full Professor**
- 2008–2009 **Speech Recognition**, *5rd year BSc*, National University of Trujillo, Peru.
2008–2009 **Computer Graphics**, *3rd year BSc*, National University of Trujillo, Peru.
2008–2009 **Image Processing**, *4rd year BSc*, National University of Trujillo, Peru.
2007–2008 **Numerical Computing**, *3rd year BSc*, National University of Trujillo, Peru.
2007–2008 **Artificial Intelligence**, *4rd year BSc*, National University of Trujillo, Peru.
- Teaching Assistant**
- 2014–2014 **Introduction to Computer Vision and Image Processing**, *3rd year BSc*, University Sao Paulo, Brazil.
2011–2011 **Computer Vision and Image Processing I**, *3rd year BSc*, University Sao Paulo, Brazil.
2010–2010 **Programming Laboratory II**, *2rd year BSc*, University Sao Paulo, Brazil.

Awards

- 2015 **FAPESP grant nro 2015/11693-2**, *Software Manager*, eScience labs, IME -USP.
2014 **BPDM Scholarship**, *Purdue University*, ACM SIGKDD 2014, New York.
2014 **IEEE-eScience**, *Scholarship*, Guaruja, Brazil.
2014 **FAPESP grant nro 2014/06549-7**, *Software Manager*, eScience labs, IME -USP.
2013 **Latam IEEE-eScience**, *Scholarship*, Sao Paulo, Brazil.
2013 **Summer School MLSS 2014**, *Tubingen*, Germany, Selected for attending.

- 2012 **CNPq grant**, 2012-2013, Sao Paulo, Brazil.
A year of doctoral internship at INSA-Rouen, France
- 2010 **CAPES grant**, 2010-2014, Sao Paulo, Brazil.
Grant for PhD studies
- 2009 **Teaching award**, *Laureate International Universities*, Trujillo, Peru.
Award for academic efficiency and good performance

Publications

- 2017 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Support Measure Data Description: A One-Class Classifier for Group Anomaly Detection, 2017 SIGACM KDD 2017 (Submitted).
- 2017 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Cross product kernels on fuzzy sets for fuzzy set similarity, 2017 IEEE International Conference on Fuzzy Systems - FUZZ-IEEE 2017 (Submitted).
- 2017 **Guevara Jorge**, and *Matthias Kormaksson, Jorge Guevara, Bianca Zadrozny*, A data-driven workflow for predicting horizontal well production using vertical well logs, DMOG 2017 - Workshop on Data Mining for Oil and Gas (Submitted).
- 2016 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Supervised machine learning with kernel embeddings of fuzzy sets and probability measures, Technical Report IME-USP.
- 2016 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Fuzzy Set Similarity using a Distance-Based Kernel on Fuzzy Sets , Book Chapter, Handbook of Fuzzy Sets Comparison - Theory, Algorithms and Applications, pages 103-120.
- 2015 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Support Measure Data Description for Group Anomaly Detection , SIGACM KDD, ODDx3'15, August 10-13, 2015, Sydney, NSW, Australia.
- 2014 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Detecting anomalous clusters of galaxies on the Sloan Digital Sky Survey Data (Poster), The 10th IEEE International Conference on e-Science, - IEEE e-Science 2014, Guarujá, Brazil.
- 2014 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Positive Definite Kernel Functions on Fuzzy Sets, 2014 IEEE International Conference on Fuzzy Systems - FUZZ-IEEE 2014, Beijing, China.
- 2013 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Support Vector Data Description for Uncertainty Data Sets (Poster), 2013 The Machine Learning Summer School, - MLSS 2014, Max Planck Institute for Intelligent Systems, Tübingen, Germany.
- 2013 **Guevara Jorge**, and *Hirata Jr, Roberto and Canu, Stephane*, Kernel Functions in Takagi-Sugeno-Kang Fuzzy System with Nonsingleton Fuzzy Input, 2013 IEEE International Conference on Fuzzy Systems - FUZZ-IEEE 2013, Hyderabad, India.
- 2009 **Guevara Jorge**, *Campos, O ; Navarro, R, R.* Extracción de características en el procesamiento digital de una señal para el reconocimiento automático del habla usando wavelets, In: VIII Jornadas Peruanas de Computación, JPC-2009.

- 2005 **Guevara Jorge**, *Campos, O ; Navarro, R*, Un Modelo para el Reconocimiento Automatico del Habla utilizando Wavelets B-Spline, In: 1st Euro-Latin American Workshop on Engineering Systems Trujillo. SISTING, 2005. p. 254-265.
- 2005 **Guevara Jorge**, *Campos, O ; Navarro, R*, Mejora del Reconocimiento Automatico del Habla utilizando una clase de Wavelets y Segmentacion no Uniforme, III Congreso internacional de cientificos peruanos RMCP-2005.

Patents

- 2016 **US 15/279632**, *Inventors: Vasconcelos, Marisa Affonso; Cardonha, Carlos Henrique; Nunes Goncalves, Bernardo; Guevara Diaz, Jorge Luis*, SYSTEM, METHOD AND COMPUTER PROGRAM PRODUCT FOR DISPATCHING OF MEDICAL EMERGENCIES AND WAIT TIME CONTROL.
- 2016 **To be filed**, *Inventors: Vasconcelos, Marisa Affonso; Cardonha, Carlos Henrique; Nunes Goncalves, Bernardo; Guevara Diaz, Jorge Luis*, DATA-DRIVEN URBAN INTERVENTIONS BASED ON CROWDSOURCING.

Softwares and Toolboxes

- MATLAB, R **Kernel on fuzzy sets**, *Cross product kernels on fuzzy sets on attribute noisy data.*
<https://github.com/jorjasso/crossProductKernel-noisyData>
- MATLAB **SMDD**, *Support Measure Description Models applied to the Group Anomaly Detection task.*
<https://github.com/jorjasso/SMDD-group-anomaly-detection>
- JAVA **Lorito**, *An isolated Spanish Speech Recognition System using Wavelets features.*
<https://github.com/jorjasso/Lorito-isolated-speech-recognizer>
- JAVA **SOM-TSP**, *A Self Organized Map neural network to solve the Traveling Salesman Problem .*
<https://github.com/jorjasso/kohonen-neural-network>
- MATLAB **Fuzzy Kernel hypothesis testing** , *Two kernel sample test with fuzzy kernels on a breast cancer dataset.*
<https://github.com/jorjasso/Two-sample-kernel-test-with-fuzzy-kernels>
- HTK, C++ **Speech Miner**, *Speech recognizer from Spanish Audios using Mel Frequency Cepstral Coefficients and Hidden Markov Models.*
<https://github.com/jorjasso/Speech-recognition-from-spanish-audios>
- MATLAB **Active set SVM**, *A support vector machine implementation using active sets and coarse search.*
<https://github.com/jorjasso/SVMActiveSet>
- JAVA **HMM**, *An implementation of a Hidden Markov Model using K-means for parameter initialization .*
<https://github.com/jorjasso/HMM>
- MATLAB **TSK-Kernels**, *TSK-Kernels on fuzzy sets applied to a support vector machine classifier on low quality datasets .*
<https://github.com/jorjasso/TSK-kernels-Low-quality>

JAVA **FNN**, *Feedforward neural network with a backpropagation algorithm for digit classification.*

<https://github.com/jorjasso/FNN-Digit-Classification>

PYTHON **VLAD**, *VLAD image descriptors.*

<https://github.com/jorjasso/VLAD>

Technical and Computer Skills

Machine learning	I have background on the analysis, design, implementation and evaluation of ML algorithms. I can formulate new ML algorithms , descriptive or generative, if its required.
Applied Machine learning	I am trained to apply several machine learning libraries from several languages on several ML tasks: classification, regression, clustering, feature engineering, anomaly/outlier detection, etc
Machine Learning methods	Support Vector Machines, Support Vector Data Description, Kernel Engineering, Hidden Markov Models for Speech recognition, Conditional Random fields.
Data science	I can apply ML algorithms for data science purpose: from raw data to insightful reports.
Fuzzy Technology	Fuzzy Learning Systems, Fuzzy Mathematics.
Signal processing	Fourier and Wavelet analysis
Technical Languages	JAVA, C++, Matlab, Python, R.
Toolkits	HTK Speech Recognition Toolkit. scikit-learn (python), dplyr (R), caret (R), git, svn
Version Control Systems	git, svn

Languages

Spanish	Mothertongue
English	Intermediate
French	Intermediate
Portuguese	Advanced

References

Roberto Hirata Jr
Stéphane Canu

hirata@ime.usp.br
stephane.canu@insa-rouen.fr

Jorge Guevara
100, Rua João Sbarai, apt 7
São Paulo, Brazil, CEP 05587-060
☎ (+55) 11 93393685
✉ jorgegd@br.ibm.com
🌐 www.vision.ime.usp.br/~jorjasso

March 1, 2017

Dear Sir or Madam,

I am a researcher scientist within the Natural Resources Analytics group at IBM Brazil. My work there basically consist on doing machine learning research using unconventional data like data from wells of oil and gas. I have interest on machine learning from theoretical to practical issues. My PhD thesis was about the design, implementation and evaluation of classifiers on data given by probability distribution and fuzzy sets. I proposed some kernel methods to deal with that. I did several theoretical contributions to the fuzzy mathematics area, the most important one is a new geometrical view of similarity measures between fuzzy sets by means of kernels. This open a new are of research where it is possible to embed the point-wise uncertainty on data or problems requiring fuzzy modelling, into the classifiers via a kernel function. I proposed state-of-the-art kernel algorithms to deal with low quality data, fuzzy data analytics, and distributional data, as well. In the past I spend almost 3 years teaching in Peruvian universities, working on numerical computing, artificial intelligence and speech recognition. Also I was teacher assistant at Sao Paulo University.