



XVI COLOQUIO INTERNACIONAL DE
GESTIÓN UNIVERSITARIA – CIGU

Gestión de la Investigación y Compromiso Social de la Universidad

Arequipa – Perú
23, 24 y 25 de noviembre de 2016

ISBN: 978-85-68618-02-8

**GENERACIÓN Y FORTALECIMIENTO DE VÍNCULOS
INTERUNIVERSITARIOS DESTINADOS AL DESARROLLO DE
APLICACIONES INFORMÁTICAS Y ESTADÍSTICAS EN BIOCENCIAS**

PRATO, LAURA BEATRIZ
Universidad Nacional de Villa María
lprato@unvm.edu.ar

FERNANDEZ, ELMER ANDRES
Universidad Nacional de Villa María
elmerfer@gmail.com

Resumen

El dominio de técnicas de análisis de Bio-datos, como la capacidad para diseñar y analizar experimentos cuyos resultados puedan ser luego transformados en nuevos dispositivos o técnicas de diagnóstico, es fundamental para hacer un uso más eficiente de los recursos destinados a la investigación y al desarrollo de productos biotecnológicos. Los pasos involucrados en la adquisición, análisis e interpretación de Bio-datos son numerosos y su correcto abordaje es crucial para el éxito de la aplicación [Buckingham2003]. Dominar este tipo de técnicas permitirá evaluar rápidamente su aplicabilidad en distintos ámbitos. En este contexto y en otros como la agrobiotecnología, la generación masiva de datos experimentales está requiriendo de la adaptación y/o creación de herramientas específicas para satisfacer los requerimientos de estas tecnologías y poder extraer información biológica relevante. En este sentido han trabajado la Universidad Nacional de Villa María y la Universidad Católica de Córdoba, ambas de Argentina, desde hace 5 años y se han establecido vínculos para hacerlo en adelante en conjunto con la Universidad de las Ciencias Informáticas de Cuba. Las tres Universidades en conjunto pretenden desarrollar metodologías estadístico/computacionales para proveer de técnicas de minería de datos a sectores de investigación y productivos de la región a través de un enfoque multidisciplinar, integrando conceptos y algoritmos provenientes de la Bioinformática, la Estadística, la Bioingeniería, la Biología y la Inteligencia Artificial, disciplinas que en general se tratan independientemente. La idea es internacionalizar este vínculo desarrollando actividades sobre la temática, perdurables en el tiempo, de manera de complementar e intercambiar conocimiento, proyectos y desarrollos.

Palabras clave:

Gestión – Vinculación - Bioinformática - Minería de Datos – Internacionalización

Abstract

The domain of techniques Bio-data analysis such as the ability to design and analyze experiments whose results can then be transformed into new devices or diagnostic techniques, it is essential to make more efficient use of resources for research and development of biotechnology products. The steps involved in the acquisition, analysis and interpretation of Bio-data are numerous and their correct approach is crucial to the success of the [Buckingham2003] application. Mastering these techniques allow you to quickly assess their applicability in various fields. In this context and in others as agricultural biotechnology, the massive generation of experimental data is requiring adaptation and / or development of specific tools to meet the requirements of these technologies and to extract relevant biological information. In this regard they have been working at the National University of Villa Maria and the Catholic University of Cordoba, both of Argentina, for 5 years and establishing links come forward to do so in conjunction with the University of Computers Science of Cuba. The three universities jointly seek to develop statistical / computational methodologies to provide data mining techniques to research and production sectors of the region through a multidisciplinary approach, integrating concepts and algorithms from bioinformatics, statistics, bioengineering, biology and artificial intelligence disciplines in general are treated independently. The idea is to internationalize this link development activities on the subject, enduring over time, so as to complement and share knowledge, projects and developments.

Keywords:

Management - Bonding - Bioinformatics - Data Mining – Internationalization

Introducción

La Universidad Nacional de Villa María en su Planificación Estratégica Institucional, ha asignando a la Internacionalización un lugar de importancia definiendo una serie de objetivos tendientes a concretarlo, entre ellos “Promover el intercambio de investigadores, docentes y estudiantes, en particular con aquellos lugares de mayor interés para la complementación del desarrollo académico” e “Impulsar una universidad sin fronteras”. Para el logro de estos objetivos la UNVM ha formalizado una serie de acuerdos con universidades de Latinoamérica, el Caribe y Europa. También ha generado programas de movilidad docente y estudiantil bilateral.

Además, la UNVM, mediante su Resolución Rectoral 046/2000 establece su “Reglamento de Políticas sobre Investigación y Desarrollo Tecnológico”, donde define sus áreas de prioridad, y vincula a éstas con las disciplinas enmarcadas en el presente proyecto. En su proyecto institucional define sus Fines específicos de investigación, donde, entre otros enuncia: Impulsar proyectos de innovación y transferencia tecnológica; Dar respuesta eficaz a los requerimientos provenientes de las instituciones públicas y privadas del medio y de la región; Estimular proyectos de investigación interdisciplinaria; -Potenciar la cooperación internacional en distintos programas; Priorizar líneas de investigación que permitan generar conocimientos científicos y tecnológicos que se requieran en el medio.

Por otro lado, la Universidad Nacional de Villa María y la Universidad Católica de Córdoba están trabajando conjuntamente desde hace cinco años en un proyecto compartido sobre Minería de Datos en Biociencias; donde se integran experiencias y criterios para desarrollar aplicaciones innovadoras en el campo de las biociencias. La meta principal es elaborar protocolos y metodologías de análisis, desarrollar métodos analíticos para generar nuevas estrategias diagnósticas, predictivas a partir de los datos recogidos que permitan extraer conocimiento en problemas biotecnológicos que se basen en investigación sólida de los procedimientos estadístico/bioinformáticos relevante para el manejo de datos experimentales a través de un trabajo cooperativo logrando vínculos entre las dos Universidades.

Como plan de desarrollo la Universidad Nacional de Villa María y la Universidad Católica de Córdoba comparten como objetivo “Promover el establecimiento de relaciones académicas y científicas con centros de investigación de universidades y otros organismos reconocidos, tanto nacionales como internacionales”.

Para internacionalizar este vínculo desarrollando actividades sobre la temática, se conformó un Proyecto de Redes con la Universidad de Ciencias Informáticas de Cuba, de manera de complementar e intercambiar conocimiento, proyectos y desarrollos.

Debido a la pertinencia de la temática que involucra a las tres universidades miembros de este proyecto se desarrollan acciones que incentiven la innovación tecnológica en el área de las biociencias, que permiten establecer programas de investigación conjunta, formación científico-tecnológica y académica y, el contacto e intercambio de docentes e investigadores con el fin de establecer y afianzar los vínculos entre instituciones universitarias argentinas y extranjeras. De esta manera potenciar la calidad de la docencia, la investigación y la extensión a través del intercambio académico, científico y cultural.

Antecedentes

La Universidad Nacional de Villa María y la Universidad Católica de Córdoba han firmado Convenios Marco de colaboración y desde el año 2010, llevan adelante proyectos de investigación de manera conjunta relacionados al área de la

Bioinformática y Aplicaciones estadísticas orientadas a las BioCiencias donde participan docentes investigadores y alumnos de ambas instituciones. Además han firmado convenios de prácticas supervisadas que son realizadas por sus alumnos dentro de sus actividades curriculares.

Como resultado de los proyectos de investigación, se han capacitado estudiantes y docentes de ambas instituciones, se han realizado presentaciones en congresos nacionales e internacionales y publicaciones en revistas especializadas, entre otros, en el ISCB Latin America 2012 en Santiago de Chile; en el Congreso Argentino de Bioinformática y Biología Computacional, en 2012, 2013, 2014 y 2015; en el 2do Congreso Internacional Red Universidad– Empresa, Octubre 2014 en México; y entre las publicaciones: TheMulti-Reference ContrastMethod: facilitating set enrichmentanalysis, Computers in Biology and Medicine, Volume 42, Issue 2 , Pages 188-194, February 2012, Glycanstructure in recombinant human FSH affectsendocrineactivity and global gene expressionon human granulosa cells,Volume 366, Issue 1, 5 February 2013, Pages 68–80. Además alumnos de ambas Universidades han realizado sus Trabajos Finales de Grado en el marco de estos proyectos, para la obtención del título de Ingenieros de Sistemas de la Facultad de Ingeniería de la UCC y la obtención del título de Licenciado en Informática en al UNVM.

En el año 2014, la Universidad Nacional de Villa María, la Universidad Católica de Córdoba y la Universidad de las Ciencias Informáticas conforman un Proyecto de Fortalecimiento de Redes Universitarias cuyo objetivo es “Establecer, promover y fortalecer la constitución de redes entre instituciones universitarias argentinas y extranjeras, a partir de la generación de espacios de intercambio en investigación y academia, en particular, para el desarrollo de aplicaciones informáticas y estadísticas en el área de las bio-ciencias.” En la actualidad existe un trabajo conjunto entre las instituciones en pos de los objetivos planteados logrando un intercambio académico, científico y cultural estableciendo futuras líneas de investigación conjunta y fortaleciendo el intercambio de conocimiento entre docentes e investigadores.

La Universidad Nacional de Villa María y la Universidad Católica de Córdoba han firmado Convenios Marco de colaboración y desde el año 2010, llevan adelante proyectos de investigación de manera conjunta relacionados al área de la Bioinformática y Aplicaciones estadísticas orientadas a las BioCiencias donde participan docentes investigadores y alumnos de ambas instituciones. Además han firmado convenios de prácticas supervisada que son realizadas por sus alumnos dentro de sus actividades curriculares.

Es importante destacar, que la Universidad Nacional de Villa María, tiene ya convenio marco firmado con la Universidad Católica de Córdoba y con la Universidad de Ciencias Informáticas de Cuba de manera separada y ya está en proceso de firma Convenios Marco y Específico tripartitos entre las tres Universidades mencionadas.

Objetivos

Objetivo General

Establecer, promover, fortalecer y consolidar la constitución de redes entre instituciones universitarias argentinas y extranjeras, a partir de la generación de espacios de intercambio en investigación y academia, en particular, para el desarrollo de aplicaciones informáticas y estadísticas en el área de las bio-ciencias, en el ámbito geográfico argentino-cubano.

Objetivos Específicos

- Afianzar los vínculos ya promovidos entre instituciones de educación superior argentinas y cubanas favoreciendo la posibilidad de trabajos académicos y de investigación en conjunto.
- Promover acciones de cooperación académica entre las universidades participantes en el área específica de Bioinformática y Estadística.
- Analizar y definir temáticas específicas de trabajo conjunto en las áreas de interés, con el fin de desarrollar capacitaciones para docentes e investigadores de las universidades argentinas y cubanas participantes.
- Realizar jornadas de formación, capacitación y actualización científica dirigida a los docentes e investigadores de las universidades argentinas y cubanas con el fin de promover trabajos de investigación de manera conjunta.
- Favorecer el contacto e intercambio de docentes e investigadores en las áreas relacionadas a la bioinformática y la Estadística para futuras acciones conjuntas en el campo académico y científico.
- Profundizar los lazos académicos entre las universidades generando a futuro especializaciones o desarrollo de carreras binacionales, mejorando las capacidades argentinas en el ámbito de la educación superior.

Etapas de Desarrollo

- Análisis y Planificación. Reuniones entre la UNVM y la UCC analizando las actividades ya desarrolladas y resultados obtenidos, y a partir de ello trazar los lineamientos generales sobre el intercambio con los integrantes de la Universidad de Cuba.
- Coordinación inicial. Reuniones por videoconferencia con los participantes de la UCI, a fin de establecer alguna línea de trabajo inicial y coordinar detalles de la visita.
- Preparación de material para el intercambio. Informes de lo desarrollado entre la UNVM y la UCC, a fin de divulgar internacionalmente los resultados obtenidos.
- Visita a la UCI en Cuba. Ejecución de la movilidad definida, y concreción de las reuniones planificadas entre todas las instituciones intervinientes. Definición concreta de líneas de trabajo común y plan de trabajo conjunto.
- Protocolos. Firma de los convenios y protocolos específicos necesarios, para la implementación de los programas elaborados.
- Evaluación de resultados. Reuniones conjuntas con los participantes para evaluar el impacto y programar cronograma de actividades relacionadas a la implementación definitiva de los planes propuestos.

Avances y Conclusiones

Resultados

A continuación se presenta un resumen de los resultados analizados por los avances realizados a la fecha por las Universidades Argentinas involucradas, en los proyectos que vienen realizando conjuntamente desde el año 2010.

A ello, debemos agregar que ya ha habido reuniones virtuales y presenciales (en Cuba) por representantes de las Universidades, definiendo los pasos a seguir en materia de actividades específicas de investigación y académicas, y para lo cual se está bosquejando el Convenio Marco y Específico de característica tripartito.

Además, ya está prevista la visita de la representante de la UNVM entre febrero y marzo de 2016 a la UCI, con el fin de concretar la firma de estos Convenios y participar de capacitaciones de posgrado en esta Universidad de Cuba.

A la actualidad, se han obtenido importantes logros durante la ejecución del presente proyecto, a saber:

A la estructura inicial del equipo, formada exclusivamente por docentes investigadores de ambas universidades argentinas, se han incorporado asesores para el área técnico-informática y alumnos de grado de carreras relacionadas a las ciencias informáticas de ambas universidades, con el propósito de formarse en el campo profesional y de investigación. Como consecuencia de trabajos desarrollados en el marco del proyecto, alumnos que intervienen han obtenido Becas de Estímulo a las Vocaciones Científicas, otorgadas por el Consejo Interuniversitario Nacional; y dos resultaron favorecidos con el Premio INTEL a la excelencia académica 2012 y 2013; con trabajos relacionados a evaluar las diferentes metodologías de programación distribuida orientadas a la explotación de arquitecturas multiprocesador, para la implementación de aplicaciones bioinformáticas computacionalmente intensivas basadas en lenguaje R.

Los alumnos Sergio Taleisnik y Juan Masashi Mishima, presentaron su Trabajo Final de la carrera Ingeniería de Sistemas de la Facultad de Ingeniería de la UCC, titulado “Sistema de análisis estadístico de experimentos proteómicos de geles de electroforesis bidimensional”, director Cristóbal Fresno, codirectora Laura Prato, presentado en Febrero de 2012.

El alumno Nicolás Ferreyra, presentó su Trabajo Final de la carrera Licenciatura en Informática del Instituto de Ciencias Básicas y Aplicadas de la UNVM, titulado “Paralelización del algoritmo SVMTOPC: determinación de estrategias para mejorar sus prestaciones”, directora Laura Prato, presentado en Junio de 2015.

Se han presentado múltiples trabajos en Congresos de índole nacional e internacional, tal el caso de:

- ISCB Latin America 2012 - Santiago de Chile, Marzo 2012 - Congress annals
- Goboot: testing set enrichment analysis robustness to background selection. Fresno C , Andrea S Llera , María R Girotti , María P Valacco , Juan A López , Osvaldo Podhajcer, Mónica G Balzarini, Federico Prada and Elmer A Fernández
- 3er Congreso Argentino de Bioinformática y Biología Computacional, Oro Verde, Septiembre 2012 (Poster):
 - DIGESuite: a Cytoscape plug-in for 2D-DIGE analysis. Sergio Taleisnik, Juan Mishima, Cristobal Fresno, Matías Semrik, Gabriela Ribero, Gabriela Merino, Laura Prato, Andrea Llera y Elmer Fernández.
 - GOboot: towards a robust SEA analysis. Cristobal Fresno, Andrea Llera, María Romina Girotti, María Pía Valacco, Juan A Lopez, Laura Zingaretti, Laura Prato, Osvaldo Podhajcer, Mónica G Balzarini, Federico Prada y Elmer Fernández.
 - Agi4x44.2c: a two-colour Agilent 4x44 Quality Control R library for large microarray projects. Germán Gonzalez, Cristobal Fresno, Gabriela Merino, Andrea Llera, Osvaldo Podhajcer y Elmer Fernández.
 - One vs One Artificial Neural Network strategy for gene expression multiclass classification. Remón L, Juárez L, Arab Cohen D, Fresno C, Prato L, Villoria L y Fernández E.
 - SVM Tree with Optimal Multiclass Partition applied to Gene expression signature classification. Pellarolo M, Arab Cohen D, Fresno C, Prato L y Fernández E.
- 4to Congreso Argentino de Bioinformática y Biología Computacional, Bariloche, Octubre 2013 (Poster):

- lmdme: Linear Modelson Designed Multivariate Experiments in R. Cristóbal Fresno, Mónica G Balzarini, Laura Prato, Elmer A Fernández.
- The course of parallelism in Bioinformatics: which R parallelization scheme to use?. Nicolás Ferreyra, Emiliano Marzioni, Cristóbal Fresno, Anibal Olivera, Gerardo Venier, Laura Prato, Elmer Fernández
- 5to Congreso Argentino de Bioinformatica y Biologia Computacional, Bariloche, Septiembre 2014 (Poster):
 - PSVMTOCP: árbol SVM paralelo para la partición óptima de clases. Nicolás Ferreyra, María Laura Zingaretti, Laura Prato, Elmer Fernández
- 2do Congreso Internacional Red Universidad-Empresa, Ciudad de México, México, Octubre de 2014:
 - Centro de Investigación, Desarrollo e Innovación en Aplicaciones Bioinformáticas. Laura Prato
- II Escuela de Computación de Alto Rendimiento (ECAR 2013), Mendoza, Julio 2013 (Sesión de Posters):
 - The course of parallelism in Bioinformatics: which R parallelization scheme to use?. Nicolás Ferreyra, Emiliano Marzioni, Cristóbal Fresno, Anibal Olivera, Gerardo Venier, Laura Prato, Elmer Fernández
- XX Congreso Argentino de Bioingeniería - SABI 2015.
 - The impact of quality control in RNA-seq experiments. Gabriela A Merino , Cristóbal Fresno, Federico Netto, Emmanuel Dias Netto, Laura Prato, Elmer Fernández
- VII Jornadas de Investigación, Villa María, Noviembre 2012, se presentaron:
 - Minería de datos en bio-ciencias: bioinformática, bioingeniería y biotecnología, Elmer A Fernández, Laura Prato, Cristóbal Fresno
 - Exploración de distintas librerías y funciones de r para análisis cluster María Laura Zingaretti, Laura Prato, Elmer Fernández
 - Procesamiento distribuido en bioinformática utilizando el lenguaje r, MaycoFraccarolli, Emiliano Marzioni, Nicolás Ferreyra, Cristóbal Fresno, Laura Prato
- VIII Jornadas de Investigación, Villa María, Noviembre 2013
 - Aplicación de cluster consenso para determinar el número de grupos en un problema de clasificación de cobertura digital. María Laura Zingaretti
 - Árbol de clasificación paralelo: una herramienta de minería de datos. Nicolás Ferreyra
- IX Jornadas de Investigación, Villa María, Noviembre 2014
 - Elector de marcadores moleculares usando biplot logístico extremo. María Laura Zingaretti, Paola Salvatierra, Laura Prato, Elmer Fernández
 - Aplicación de técnicas multivariadas para el análisis de la relación entre propiedades de suelos sódicos y comportamiento de raíces. Paola Salvatierra, Laura Prato, Elmer Fernandez
 - PSVMTOCP: árbol SVM paralelo para la partición óptima de clases. Nicolás Ferreyra, María Laura Zingaretti, Laura Prato, Elmer Fernández
- X Jornadas de Investigación, Villa María, Noviembre 2015

- Uso de técnicas estadísticas multivariadas para caracterizar patrones de lluvias de la República Argentina. Zingaretti María Laura, Prato Laura, Fernández Elmer
- Serie histórica de precipitaciones de la Provincia de Córdoba: una vista multivariada. Vargas Soria José, Prato Laura, Fernández Elmer.

Publicaciones en revistas de referato internacional:

- Fresno C, Llera AS, Girotti R, Valacco MP, López J; Podhajcer O, Balzarini M, Prada F, Fernández EA, The Multi-Reference Contrast Method: facilitating set enrichment analysis, *Computers in Biology and Medicine*, Volume 42, Issue 2 , Pages 188-194, February 2012
- Loreti N, Fresno C, Fernandez AE, Barrera D, Lira S, Larrea F, Campo S. Glycan structure in recombinant human FSH affects endocrine activity and global gene expression on human granulosa cells, *Volume 366, Issue 1, 5 February 2013, Pages 68–80*

Capacitaciones

- Estancia de Investigación en el laboratorio de Biometría y Estadística del Instituto de Estudios Avanzados (IDEA). Caracas, Venezuela. Enero- 2014. Participaron María Laura Zingaretti y Paola Salvatierra.
- Coordinación y dictado del seminario “Implementación de Metodología Estadística para el análisis de datos biológicos multivaria: stasis, stasis dual” - UNVM. Marzo-2014 – María Laura Zingaretti y Paola Salvatierra.
- Taller Gtec-Agencia – Fondo Argentino Sectorial – Ministerio de Ciencia, Tecnología e Innovación Productiva – Buenos Aires – 12 de Junio de 2014. Participó Laura Prato.
- Taller de Vigilancia Tecnológica para innovar en red – Universidad Autónoma Metropolitana – México D.F. – Octubre de 2014. Participó Laura Prato.
- Estancia de Investigación en la empresa Tecnalía en Bilbao, España – Octubre 2015. Participó Laura Prato
-

Referencias Bibliográficas

- Balzarini y cols (2004). Info-Gen: Software para análisis de datos genéticos. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Argentina. Copyright 362964 CESSI.
- Buckingham S. (2003) Programmed for success. *Nature*. V425 pp 209-215
- Churchill, Doerge RW. 1994. Empirical Threshold Values for Quantitative Trait Mapping. *Genetics* 138: 963-971.
- Fernández EA. (2003) Minería de Datos y redes Neuronales para el análisis de la cinética de diálisis. Tesis de Doctorado, Universidad de Santiago de Compostela, España
- Fernández EA, Balzarini M. (2005a) Improving Cluster Identification and Visualization in SOM like algorithms applied to Gene Expression Pattern Analysis. Enviado a *BMC Bioinformatics*.
- Fernández Elmer A, GirottiMaría R., López Juan A, Llera Andrea S., Podhajcer Osvaldo L, Cantet Rodolfo J. C. and BalzariniMónica Improving 2D-DIGE protein expression analysis by two-stage linear mixed models: Assessing experimental effects in a melanoma cell study *Bioinformatics*

- Fernandez EA, Souza Net EP, Abry P, Macchiavelli R, Balzarini M, Cuzin B, Baude C, Frutoso J, Gharib C Assessing erectile neurogenic dysfunction from heart rate variability through a Generalized Linear Mixed Model framework.. Computer Methods and Programs in Biomedicine [doi:10.1016/j.cmpb.2009.11.001](https://doi.org/10.1016/j.cmpb.2009.11.001)
- FernándezEA, Valtuille R, Willshaw P, Perazzo CA. (2001) Detection of Abnormality in the ECG Without Prior Knowledge by Using the Quantization Error of a Self-Organizing Map (SOM), tested on the European Ischemia Database, Medical & Biological Engineering & Computing. 39:330-337
- Fresno Cristóbal, Bioengineer; Andrea S Llera, Ph.D.; María R Girotti, Ph.D.; María P Valacco, Ph.D.; Juan A López; Osvaldo L Podhajcer, Ph.D.; Mónica G Balzarini, Ph.D.; Federico Prada, Ph.D.; Elmer A Fernández, The Multi-Reference Contrast Method: facilitating set enrichment analysis, Ph.D, Computers in Biology and Medicine.
- Gianola D, Fernando RL, Stella A. 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173: 1761–1776.
- Kang M, Balzarini M. Guerra J. 2004. Genotype-by-Environment interaction. In A. Saxton (ed.) SAS Genetic Book, SAS Institute, Cary NC (in press).
- Kohonen, T. (1997) Self-Organizing Maps (Springer, Berlin).
- Laciár E, Campos RJ, Brooks DH. Evaluación del daño miocárdico en enfermos chagásicos crónicos a partir del análisis del electrocardiograma de alta resolución... XV Congreso de Bioingeniería SABI 2005
- Podhajcer O. (2002). Estudios de genómica funcional mediante el uso de matrices de ADN (microarrays). Conferencia plenaria. Actas, XXXI Congreso Argentino de Genética, La Plata, 17-20 de setiembre de 2002.
- Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270, 467-470.
- U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, editors, [Advances in Knowledge Discovery and Data Mining](#), AAAI/MIT Press, 1996