

Líneas de Investigación del Laboratorio de Sistemas de Información Avanzados: Dinámica de Tecleo, Computación Afectiva, Extracción de Relaciones Semánticas, Blockchain & Smart Contracts.

Jorge Ierache, Hernán Merlino, Enrique Calot, Juan Rodriguez, German Concilio, Federico Rossi, Sabrina Campa, Ezequiel Aceto, Alejandra Ochoa, Nahuel González.

Laboratorio de Sistemas de Información Avanzados, Departamento Computación
Facultad de Ingeniería, Universidad de Buenos Aires.
Av. Paseo Colón 850 - C1063ACV - Buenos Aires -Argentina
Tel +54 (11) 4343-0893 / 4343-0092.

{jierache,hmerlino,ecalot}@lsia.fi.uba.ar

Resumen

El Laboratorio de Sistemas de Información Avanzados (LSIA) de la Facultad de Ingeniería de la Universidad de Buenos Aires (UBA) cuenta con distintas líneas de investigación: Dinámica de Tecleo, Extracción de relaciones semánticas en español, Computación afectiva, Blockchain y Smart Contracts. En varias de estas líneas se han obtenido aportes relevantes al área mientras que otras están comenzando y mantienen objetivos prometedores.

Palabras clave: *LSIA, Procesamiento del lenguaje Natural, Dinámica de tecleo, Computación afectiva, Blockchain, Smart Contracts.*

Contexto

El Laboratorio de Sistemas de Información Avanzados de la Facultad de Ingeniería UBA cuenta con distintas líneas de investigación [LSIA, 2019]. Fue creado en el 2011 y cuenta con diversas publicaciones; se radicó un proyecto UBACyT en el contexto de cadencia de tecleo, en colaboración con el ISIER UM, para el empleo de Interfaces Cerebro-Máquina (BMI). Recientemente se inició un Proyecto de desarrollo Estratégico (PDE) en el marco de patrones de dinámica de tecleo en el contexto *e-commerce*.

Introducción

En este artículo se enuncian las distintas líneas de investigación y su relación con las materias de grado, fomentando la investigación de alumnos de las carreras de Ingeniería y Licenciatura Informática. Se describen a continuación las distintas líneas de investigación del laboratorio.

Líneas de Investigación, Desarrollo e Innovación

El laboratorio cuenta con las siguientes líneas de investigación:

Keystroke Dynamics. Los patrones neurofisiológicos que vuelven única a una firma manuscrita se pueden observar también en el ritmo de tecleo de un usuario [Joyce & Gupta, 1990]. La técnica que analiza este tipo de patrones se llama Dinámica de Tecleo o Keystroke Dynamics [Calot et al. 2014], [González et al. 2015, 2016]. Desde el laboratorio se mantiene una línea de investigación que, además de articular keystroke dynamics con la extracción de información emocional del usuario, investiga mejores algoritmos de identificación de personas mediante su ritmo de tecleo.

Computación afectiva. Esta línea de investigación se centra en el análisis de factores emocionales de grupos de personas y en la toma de decisiones por software basada en las emociones inferidas. En particular, se realizan estudios sobre grupos de personas en distintos escenarios, detectando sus expresiones, midiendo ritmo cardíaco y monitoreando su actividad cerebral. Con ellos, se permite inferir su estado emocional (empleando métodos de aprendizaje automático supervisado) y se ejecutan acciones que refuerzan o retraen dicho estado.

Extracción de relaciones semánticas en español. Esta línea se orienta al desarrollo de un *parser* semántico, ya sea supervisado o no, que sea capaz de extraer relaciones semánticas de un texto no estructurado: lenguaje natural en español. La salida de este *parser* sería una red semántica, la cual tendría que servir para

generar, o bien para extender una ontología de forma semi-automática. Este *parser* estaría compuesto por una serie de pasos, en donde cada uno realizaría una tarea atómica utilizando algoritmos conocidos y probados dentro del área de PLN: Tokenización, POS *tagger*, *named entity recognizer*, *syntactic parser*, entre otros.

Blockchain & Smart Contracts. Esta línea se enfoca en el desarrollo de un framework para la automatización de Smart Contracts (SC) y en todo el proceso de implementación de los mismos en la Blockchain. Para la generación automática de SC se está utilizando programación genética, sistemas basados en conocimiento y sistemas de aprendizaje automático para la evaluación de los mismos.

Resultados y Objetivos

Keystroke Dynamics. Se mostró que el método de Mahalanobis no es siempre útil, especialmente si los datos son linealmente independientes, y que puede ser reemplazado por el método menos costoso de distancia euclídea normalizada [Calot et al., 2013]. A su vez, se creó un conjunto de datos de dinámicas de tecleo no estructurada que fue publicado en el sitio web del laboratorio [Calot, 2015]. Mediante este conjunto de datos y pruebas internas se comenzaron a evaluar estrategias más sofisticadas de extracción de características, filtrado del texto de entrada, selección de modelos y la alternancia de clasificadores que prometen mejorar el EER por debajo del 5% –frente a un 10% en usuarios estándar y 6% en usuarios de mayor pericia dactilográfica– y disminuir su variación entre usuarios. Se encuentra en desarrollo un proyecto de desarrollo estratégico vinculado a la transferencia a la industria reconocido por

la UBA (PDE-44-2019 Reconocimiento de patrones de teclado en ambientes Web). En este marco se realizó un trabajo colaborativo con una de las principales empresas de *e-commerce* de América Latina, empleando técnicas de modelado de cadencia de tecleo para autenticar al usuario, y luego identificar patrones de tipeo compartidos por uno más usuarios de la plataforma. Los resultados experimentales sobre una muestra de 2000 usuarios seleccionados al azar mostraron que es posible identificar a quien tipea con una precisión del 89%, [Concilio et al., 2018], autenticar sesiones con un EER por debajo del 8% y detectar usuarios controlados por una misma persona física, utilizando sólo las marcas de tiempo originadas por la presión de cada una de las teclas del teclado en campos de texto libres. En la muestra se lograron aislar 30 usuarios con comportamiento fraudulento confirmados.

Computación afectiva. Se está recolectando información de los estados de los individuos para nutrir la red de conocimientos de su comportamiento y permitir posteriormente inferir estados. El trabajo realizado consistirá en un framework capaz de detectar el estado predominante del grupo y permitir llevar el mismo a un estado emocional compartido entre todos los miembros del grupo. [Ierache et al., 2012], [Ierache et al., 2014], [Calot, E., Ierache J. et al, 2019]. El empleo de Interfaces Cerebro-Máquina BMI se realizó con el ISIER UM. En este sentido, se está incursionando en el reconocimiento de estados emocionales a partir del registro de bioseñales humanas. Particularmente, se explotará el uso de marcadores biométricos provenientes de señales electroencefalográficas (EEG) y de ritmo cardíaco (HR). En el LSIA se desarrolló una cámara de inmersión emocional para la realización de

experimentos [Rossi F. et al., 2017], [Calot E., Rossi F. et al, 2016] [Calot E., Ierache J., 2017]. Esta cámara propició la creación de un ambiente de estimulación con imágenes IAPS (International Affective Picture System), [Lang, P. J., Bradley, M. M. & Cuthbert, B, 1997], audios y videos, en el contexto de múltiples estados afectivos enmarcados en el modelo de Excitación-Valencia [James A. Russell, 1980, 1989]. Durante la estimulación a voluntarios se solicitó el tipeo de textos (en este orden se fusionaron las líneas de investigación de dinámica de tecleo y de computación afectiva). Además, se presentaron encuestas SAM (Self-Assessment Manikin) [Bradley, M. M. & Lang, P. J., 1994] para ser respondidas luego de cada estimulación. Si bien el EEG como base de inferencia ha sido tratado ampliamente, las líneas de investigación proponen establecer un marco neurocientífico sobre el cual se sustenten los criterios de diseño y ejecución experimentales aplicados. A su vez y, dado que el ritmo cardíaco no resulta ser una señal de uso extensivo en estudios de computación afectiva, se evaluará su robustez como característica de inferencia, analizando el desempeño de múltiples algoritmos de aprendizaje automático, brindando así un aporte al estado del arte del área.

Extracción de relaciones semánticas en español. Esto corresponde a una investigación en proceso en la cual ya se han obtenido resultados en un experimento realizado con prescripciones médicas, en donde se probaron dos métodos de clasificación de textos, con textos médicos en español: *Naïve Bayes Multinomial* (NBM) y *Support Vector Machines* (SVMs). La clasificación de textos, en este caso, si bien fue acotada al campo de la medicina, logró una precisión del 86% [Rodríguez et al., 2014]. Otros

avances en esta línea se han publicado en [Rodríguez et al., 2015], [Rodríguez et al., 2016a], [Rodríguez et al., 2016b], [Rodríguez et al., 2017a] y [Rodríguez et al., 2017b]

Blockchain & Smart Contracts. Los resultados hasta ahora obtenidos permiten esperar que el desarrollo de los mismos pueda ser llevado adelante por personas sin experticia en el tema, solo contestando un conjunto de preguntas. Un sistema experto generará el contrato y permitirá su puesta en etapa de pruebas o producción según lo defina el usuario.

Formación de Recursos Humanos

El laboratorio actualmente se conforma de dos investigadores formados, tres investigadores en formación, tres alumnos investigadores, un egresado investigador en formación. Se han radicado dos tesis de doctorado y diversas tesis de grado de la Facultad de Ingeniería de la UBA. Los docentes participan de materias de la carrera de Ingeniería en Informática y Licenciatura en Sistemas: (75.39) “Aplicaciones Informáticas”, (75.41) “Algoritmos y Programación II”, (75.50) “Introducción a los Sistemas Inteligentes”, (75.65) “Manufactura Integrada Por Computador”, (75.66) “Manufactura Integrada Por Computador II”, (75.67) “Sistemas Automáticos de Diagnóstico y Detección de Fallas I”, (75.68) “Sistemas de Soporte para Celdas de Producción Flexible”, (75.69) “Sistemas Automáticos de Diagnóstico y Detección de Fallas II” y (75.70) “Sistemas de Programación no Convencional de Robots”.

Referencias

BRADLEY, M. M. & LANG, P. J. Measuring Emotion: The Self-Assessment Manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59, 1994

CALOT, E. 2015. “Keystroke Dynamics keypress latency dataset”. Base de datos para investigación.
<http://lsia.fi.uba.ar/pub/papers/kd-dataset/>

CALOT, E.; PIRRA, F.; RODRIGUEZ, J.M.; PEREIRA, G.; IRIBARREN, J.; IERACHE, J. 2014. “Métodos Adaptativos de Educación de Dinámica de Tecleo Centrado en el Contexto Emocional de un Individuo aplicando Interfaz Cerebro Computadora”. XVI Workshop de Investigadores en Ciencias de la Computación, ISBN 978-950-34-1084-4.

CALOT, E.; RODRIGUEZ, J.M.; IERACHE, J. Improving versatility in keystroke dynamic systems. En *Proceedings del XIX Congreso Argentino de Ciencias de la Computación*, number 5606, 2013. ISBN 978-987-23963-1-2. URL <http://hdl.handle.net/10915/32428>

CALOT, E.; RODRIGUEZ, J.M.; IERACHE, J., 2014. Improving versatility in keystroke dynamic systems. En Jorge Raúl Finochietto y Patricia Mabel Pesado, editors, *Computer Science & Technology Series. XIX Argentine Congress of Computer Science, Selected papers*, páginas 289–298. Editorial de la Universidad Nacional de La Plata (EDULP), 2014b. ISBN 978-987-1985-49-4. URL <http://lsia.fi.uba.ar/papers/calot14b.pdf>

CALOT, E., ROSSI F., GONZÁLEZ N., HASPERUÉ, W., IERACHE, J.. Avances en educación de dinámica de tecleo y el contexto emocional de un individuo aplicando interfaz cerebro computadora. En *WICC 2016, Entre Ríos, Argentina*, páginas 872–876, jun. 2016. ISBN 978-950-698-377-2. URL <http://hdl.handle.net/10915/53247>

CALOT, E., IERACHE, J.. Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing. En *Proceedings del XXIII Congreso Argentino de Ciencias de la Computación (La Plata, 2017)*, number 10529, páginas 1030–1039, 2017. ISBN 978-950-34-1539-9. URL <http://hdl.handle.net/10915/63866>

CALOT, E., IERACHE, J., HASPERUÉ, W.. 2019. Robustness of keystroke dynamics identification algorithms against brain-wave variations associated with emotional variations.

En *Advances in Intelligent Systems and Computing*. Springer, c. En prensa

CONCILIO, G., IERACHE, J., MERLINO, H., CALOT, E.. Application of Keystroke Dynamics Modelling Techniques to Strengthen the User Identification in the Context of E-commerce. En XXIV CACIC 2018. En prensa. URL <http://lsia.fi.uba.ar/papers/concilio18.pdf>

ROSSI, F., CALOT, E., IERACHE J. Educación emocional de un individuo en contextos multimodales en computación afectiva. En XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, Buenos Aires, Argentina), jul. 2017. ISBN 978-987-42-5143-5

IERACHE, J., PEREIRA, G., IRIBARREN, J., SATTOLO, I. 2012. "Robot Control On The Basis Of Bio-Electrical Signals": "International Conference On Robot Intelligence Technology and Applications" (rita 2012). Series advances in intelligent and soft computing of springer.

IERACHE, J, PEREIRA., G, NERVO., F, IRIBARREN, J.. 2013. Estado emocional centrado en estímulos, aplicando Interfase cerebro-máquina. XX CACIC, ISBN 978-987-3806-05-6.

JOYCE, R.; GUPTA, G. 1990. Identity authentication based on keystroke latencies. *Commun. ACM* 33, 2 (February 1990), 168-176. <http://doi.acm.org/10.1145/75577.75582>

GONZÁLEZ, N., CALOT, E. Of Keystroke Dynamics In Free Text. En *Biometrics Special Interest Group (Biosig), 2015 International Conference Of The*, Páginas 1–5, Sep. 2015. ISBN 978-3-88579-639-8. DOI: 10.1109/Biosig.2015.7314606

GONZÁLEZ, N., CALOT, E. Y IERACHE, J... A Replication Of Two Free Text Keystroke Dynamics Experiments Under Harsher Conditions. En *2016 International Conference Of The Biometrics Special Interest Group (Biosig)*, Páginas 1–6, Sep. 2016. DOI: 10.1109/Biosig.2016.7736905

LANG, P. J., BRADLEY, M. M. & CUTHBERT, B. N. INTERNATIONAL AFFECTIVE PICTURE SYSTEM (IAPS): TECHNICAL MANUAL AND AFFECTIVE RATINGS. NIMH CENT. STUDY EMOT. ATTEN. 39–58, [HTTPS://DOI.ORG/10.1027/0269-8803/A000147](https://doi.org/10.1027/0269-8803/A000147) (1997)

LSIA: <http://lsia.fi.uba.ar> vigente marzo 2019

RODRÍGUEZ, J.M.; CALOT, E.P.; MERLINO H. 2014. "Clasificación de Prescripciones Médicas en Español". En "XX Congreso Argentino de Ciencias de la Computación". N° 1258. ISBN 978-987-3806-05-6.

RODRÍGUEZ, J., MERLINO, H., GARCÍA-MARTÍNEZ, R.. 2015. Revisión sistemática comparativa de evolución de métodos de extracción de conocimiento para la web. Argentina. La Plata. Libro. Artículo Completo. CACIC 2015.

RODRÍGUEZ, J., MERLINO, H., PESADO, P., GARCÍA -MARTÍNEZ, R. 2016a. Performance Evaluation of Knowledge Extraction Methods. Suiza. Basel. 2016. Libro. Artículo Breve. Congreso. International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA-AIE 2016).

RODRÍGUEZ, J., MERLINO, H., PESADO, P., GARCÍA -MARTÍNEZ, R. 2016b. Clasificación de distintos conjuntos de datos utilizados en evaluación de métodos de extracción de conocimiento creados para la web. Argentina. San Luis. Congreso. CACIC XXII

RODRÍGUEZ, J., MERLINO, H., PESADO, P., GARCÍA -MARTÍNEZ, R. 2017a. Evaluation of Open Information Extraction Methods Using Reuters-21578 Database. Estados Unidos de América. New York. 2017. Libro. Artículo Completo. Congreso. 2nd International Conference on Machine Learning and Soft Computing (ICMLSC '18).

RODRÍGUEZ, J., MERLINO, H., GARCÍA -MARTÍNEZ, R.. 2017b. Automatic Characteristics Extraction for Sentiment Analysis Tasks. Suiza. Basel. 2017. Libro. Artículo Completo. Congreso. XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017). Universidad Nacional de La Plata.

RUSSELL, J. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980. ISSN 0022-3514. DOI: 10.1037/h0077714

RUSSELL, J. Measures of emotion, páginas 83–111. Academic Press, 1989. ISBN 978-0-12-558704-4.