

LÍNEAS DE INVESTIGACIÓN DEL LABORATORIO DE SISTEMAS DE INFORMACIÓN AVANZADOS.

Enrique P. Calot, Ezequiel L. Aceto, Juan Manuel Rodriguez, Ariel M. Liguori, M. Alejandra Ochoa, Hernán Merlino, Enrique Fernandez Nahuel González, Francisco I. Pirra, Jorge S. Ierache

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.

{ecalot, eaceto, jmrodriguez, aliguori, hmerlino, ngonzalez, fpirra, jierache}@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: Internet de las Cosas, Extracción de relaciones semánticas en español, Explotación de Información, Dinámica de Tecleo, Redes definidas por software y Computación afectiva colectiva. En varias de estas líneas se han obtenido aportes relevantes al área mientras que las líneas que están comenzando mantienen objetivos prometedores.

Palabras clave: *LSIA, Internet de la cosas, Procesamiento del lenguaje Natural, Dinámica de tecleo, Computación afectiva colectiva, Software defined networking, Contextos finitos, Explotación de información*

Contexto

El Laboratorio de Sistemas de Información Avanzados de la Facultad de Ingeniería UBA cuenta con distintas líneas de investigación [LSIA, 2015]. Fue creado en 2011 y cuenta con más de 30 publicaciones en las líneas abajo mencionadas. Además radica un proyecto UBACyT, desarrolla en colaboración con el ISIER UM investigación aplicada en el marco del empleo de interfases Cerebro Maquina (BMI). Recientemente ha recibido apoyo en hardware por parte de Intel.

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. Finalmente se describen los resultados de las distintas líneas de

investigación del laboratorio, en caso de existir.

Líneas de Investigación, Desarrollo e Innovación

El laboratorio cuenta con las siguientes líneas de investigación:

Internet de las Cosas. El programa aquí presentado tiene como propósito instruir y fomentar el desarrollo de sistemas de estrecha relación con la *Internet de las Cosas* (IoT) en el ámbito universitario. Es de mucha importancia que los futuros profesionales del área de Sistemas se integren rápidamente a las líneas de trabajo relacionadas con IoT, dado que su participación es necesaria en un segmento de la industria que crece muy rápidamente de la mano de la electrónica de consumo.

El proyecto está siendo impulsado por el laboratorio y aplicado en la materia 75.39 “Aplicaciones Informáticas” dictada también en la Facultad de Ingeniería, para la carrera Licenciatura en Análisis de Sistemas.

El material requerido para llevar adelante la capacitación ha sido donado por Intel a través de su *Intel University Program*. El hardware utilizado es la placa "Intel Galileo" de 2da generación y el kit "*Grove - Starter Kit Plus (Intel Internet of Things)*" como soporte para facilitar la creación de prototipos reduciendo la necesidad de que los alumnos tengan conocimientos muy específicos de electrónica.

Extracción de relaciones semánticas en español. El 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 paso realizaría una tarea atómica utilizando algoritmos conocidos y probados dentro del área de PLN: Tokenización, POS *tagger*, *named entity recognizer*, *syntactic parser*, entre otros.

Explotación de Información. Se desarrolla en el marco de la materia (75.50) “Introducción a los Sistemas Inteligentes”. En esta línea se trabaja en aplicaciones para: la optimización de procesos en el ámbito de la industria, el e-commerce, en la detección de fraudes comerciales y financieros, para la predicción en el ámbito de las finanzas y acciones de bolsa y detección de patrones de comportamiento en la Web.

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]. 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.

Redes definidas por software. O por sus siglas en inglés *software defined networking* (SDN). El incremento de la utilización de redes IP en diversos ámbitos impulsado por las tecnologías móviles, IoT y la diversidad de contenido desprende una necesidad emergente de poder tomar el control de cada acción en la red, influyendo directamente en su funcionamiento a bajo nivel (interactuando en su plano de control)

como en las aplicaciones que sobre ella se manifiestan. Para lograr este nivel de control se realizan aplicaciones capaces de interactuar con cada dispositivo que conforma la red en cuestión mediante el protocolo OpenFlow, centralizando información de flujos y estados de la red, modelando el comportamiento de la misma y actuando en consecuencia. Esencialmente los estudios sobre esta temática se han centrado en la elaboración de sistemas de detección de patrones de tráfico y de sistemas de ruteo basado en las prioridades del usuario.

Computación afectiva colectiva. 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 la misma. En particular se realizan estudios sobre grupos de personas en distintos escenarios, detectando sus expresiones, midiendo pulsaciones y monitoreando su actividad cerebral y en base a ello se permite inferir su estado emocional, empleando métodos de aprendizaje supervisado, y ejecutar acciones que permitan mantener o corregir el mismo.

Resultados y Objetivos

Los resultados para cada proyecto son:

Internet de las Cosas. Durante el 2do cuatrimestre de 2014 hemos recibido la primera iniciativa por parte de un grupo de alumnos para crear una solución relacionada con IoT con potencial para ser aplicado en la Facultad una vez terminado. El desarrollo del mismo ha comenzado el 1er cuatrimestre de 2015 y estará finalizado en Julio del corriente. Esperamos que durante el transcurso del año se sumen al menos dos proyectos más a esta línea de trabajo de IoT.

Extracción de relaciones semánticas en español. Al encontrarse en una etapa inicial, los únicos resultados obtenidos

son los de un experimento realizado con las prescripciones medicas, en donde se probaron dos métodos de clasificación de texto: *Naïve Bayes Multinomial* (NBM) y *Support Vector Machines* (SVMs) con textos médicos en español. La clasificación de textos en este caso, si bien acotada al campo de la medicina, logró una precisión del 86% [Rodriguez et al., 2014].

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 set de datos de dinámicas de tecleo no estructurada que fue publicado bajo la página web del laboratorio [Calot, 2015]. Mediante este set de datos y pruebas internas se comenzó a evaluar estrategias más sofisticadas de extracción de características, filtrado del texto de entrada, selección de modelos y alternación de clasificadores que prometen empujar 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.

Redes definidas por software. La exploración de esta línea de conocimiento es desarrollada en la asignatura (75.39) “Aplicaciones Informáticas” para la Licenciatura en Análisis de sistemas Informáticos de la Facultad de Ingeniería de la UBA, como trabajo final de la misma se ha desarrollado un sistema de detección de patrones de tráfico de ataques de denegación de servicios distribuido (*DDoS, distributed denial of service*) sobre una red operada con un controlador central mediante *OpenFlow v1.3*. El aprendizaje empleado en la red era supervisado empleando redes neuronales y *backpropagation*, el filtrado

de entradas en los dispositivos de red se realizo mediante el agregado de filtros de entrada en la Tabla CAM (*Content Addressable memory*).

Computación afectiva colectiva. Se esta 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 mismo. [Ierache et al., 2012], [Ierache et al., 2014]. El empleo de interfases cerebro- maquina (BMI) se realiza con el ISIER UM.

Formación de Recursos Humanos

Desde el laboratorio se han formado investigadores, radicado dos tesis de doctorado y más de 10 tesis de grado en 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”. Además docentes del laboratorios han dirigido tesis de grado en la carrera Ingeniería Industrial y Postgrado en el Maestría de Ingeniería de Software ITBA.

Referencias

- 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. 2013. Improving versatility in keystroke dynamic systems. XIX Congreso Argentino de Ciencias de la Computación., ISBN 978-897-23963-1-2.
- 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, J, IRIBARREN. 2013. Estado emocional centrado en estímulos, aplicando Interfase cerebro-maquina. 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>
- LSIA, 2015. <http://lsia.fi.uba.ar/> , vigente marzo 2015
- RODRIGUEZ, 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.