

Proceso de Elicitación de Requerimientos para la Construcción de Modelos Predictivos Basados en Sistemas Inteligentes

Cinthia Vegega, Pablo Pytel, Hugo Ramón, Luciano Straccia, Anahi Bazet, Ariel Deroche, Sebastián Plawner & María Florencia Pollo-Cattaneo
 Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS)
 Programa de Maestría en Ingeniería en Sistemas de Información.
 Facultad Regional Buenos Aires. Universidad Tecnológica Nacional.
 cinthiavg@yahoo.com.ar; {flo.pollo, ppytel}@gmail.com

Resumen

¿Cuál opción es la más favorable? ¿Qué nos deparará a futuro? Estas son preguntas que nos hacemos cada vez que tenemos que tomar alguna decisión. Por esta razón, contar con mecanismos que permitan realizar predicciones acertadas es algo que la Humanidad siempre ha querido. Dicha necesidad no es propia únicamente de los individuos y también impacta en las organizaciones. Teniendo en cuenta estas cuestiones, el presente proyecto propone el desarrollo de un proceso que a partir de los deseos y necesidades de una organización, asista a la elicitación de los requerimientos para implementar exitosamente un Modelo Predictivo basado en Sistemas Inteligentes.

Palabras clave: Ingeniería de Software. Elicitación de Requerimientos. Modelos Predictivos. Sistemas Inteligentes. Aprendizaje Automático.

En el marco de las actividades del Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) se comienza una nueva línea de trabajo en el campo de la automatización de procesos orientados a la Toma de Decisiones en organizaciones, mediante la aplicación de

Modelos Predictivos, basados en Sistemas Inteligentes. De esta manera, se articula dentro de los objetivos de GEMIS en el campo de la Informática, la generación de nuevos conocimientos en el área de la Ingeniería de Software con la aplicación de tecnologías no convencionales provenientes del Aprendizaje Automático. Además, se desarrolla dentro del ámbito del PID con incentivos UTN 'UTI3799TC', y una Tesis de Maestría.

Introducción

La Toma de Decisiones es un proceso esencial en la vida de las organizaciones [1]. Como se indica en [2], las organizaciones son sistemas no lineales altamente sensibles a las innovaciones, a los eventos o al azar propio del ambiente empresarial. Esto lleva a que diariamente deben decidir cómo asignar sus valiosos recursos basados en predicciones [3] sobre el tiempo, el esfuerzo y/o los riesgos que actividades. Esta situación se encontrará acenuada debido al entorno altamente complejo y de difícil predicción del siglo XXI [4]. Además genera que las organizaciones deban actuar más allá de los sistemas tradicionales de gestión e incorporar nuevos mecanismos para la "creación y potenciación del conocimiento organizativo" tales como

los provistos por la Inteligencia Artificial [5; 6]. Un caso de estos nuevos mecanismos se puede encontrar en el ámbito de los Modelos Predictivos.

A pesar de que tradicionalmente se han aplicado Técnicas Estadísticas y Modelos Paramétricos para generar predicciones [7], en las dos últimas décadas se han incorporado diversos métodos asociados al Aprendizaje Automático [3; 8]. Estos métodos son considerados normalmente como 'cajas negras' o modelos basados en datos [9] y permiten construir modelos que permiten encontrar la relación entre situaciones pasadas y futuras, utilizando los datos históricos disponibles.

No obstante, desafortunadamente los Modelos Predictivos suelen ser imprecisos o, en algunos casos, se equivocan, generando a menudo respuestas que son incomprensibles [3]. En este sentido, la calidad de la información requerida es sumamente importante para poder tomar decisiones bajo certeza [10]. Al reducir la ignorancia sobre el problema y su contexto es posible generar mejores predicciones. Como indica el físico Richard Feynman, *"lo que siempre deberíamos tener en cuenta para evitar grandes fracasos, catástrofes y pánicos no es lo que sabemos, sino lo que no sabemos que no sabemos"* [11]. Por consiguiente, además de recolectar los datos históricos que se aplicarán para construir el Modelo Predictivo, también es imprescindible identificar las características generales del dominio en donde se lleva a cabo la predicción, y así poder detectar situaciones o eventos de los que no se poseen datos, pero que el modelo deberá considerar.

En este contexto, la presente línea de trabajo propone un nuevo proceso ingenieril que asista al Ingeniero de Software en la difícil tarea de recolectar, entender, identificar y documentar la

información necesaria para implementar un Modelo Predictivo basado en Sistemas Inteligentes (también denominado como Modelo Predictivo Inteligente o MPI). Para ello, se consideran las particularidades del problema predictivo y de las arquitecturas más utilizadas para resolverlo, así como, las necesidades de las organizaciones argentinas.

Líneas de Investigación, Desarrollo e Innovación

De acuerdo a la meta establecida en la sección anterior, a continuación se exponen los motivos que justifican el proceso propuesto:

o **Modelos Predictivos Inteligentes**

Se considera al Aprendizaje Automático (o Machine Learning) como la disciplina que estudia y modela los procesos de aprendizaje con sus múltiples manifestaciones para poder ser trasladados a las computadoras [9]. De esta manera, se busca reproducir aspectos del comportamiento de los seres inteligentes para implementar sistemas software en ámbitos complejos [12]. Los sistemas software así generados se denominan Sistemas Inteligentes [13] y pueden ser aplicados en diferentes dominios para resolver diferentes tipos de problemas. Uno de estos problemas tiene que ver con la construcción de modelos capaces de inferir a partir de datos históricos las dependencias entre valores pasados y valores futuros a corto plazo [8], es decir, los Modelos Predictivos.

Debido al incremento de la cantidad de datos disponibles en los repositorios de las organizaciones, y mejoras de los rendimientos de procesamiento del hardware, es posible aplicar métodos computacionales intensivos para construir este tipo de modelo [14]. En este sentido se destacan las Redes Neuronales

Artificiales, o RNA, [15; 16] y las Redes Bayesianas, o RB, [17; 18] como las principales arquitecturas a ser utilizadas para este tipo de problema [14; 19; 20].

○ **Dificultades de la construcción de Modelos Predictivos Inteligentes**

Según indica [8], predecir el futuro es una de las tareas más importantes y difíciles de las ciencias aplicadas. En este sentido, resulta obvio que contar con datos representativos del problema le permitirá al modelo generar mejores predicciones. Sin embargo, obtener los datos no es una tarea trivial. Por ejemplo, tómesese la cuestión de determinar la cantidad de información histórica que se necesita para producir los mejores resultados [21]. Según [22], la respuesta habitual a la pregunta “¿cuántos datos se necesitan?” es “lo más posible”. Cuantos más datos se tenga, mejor se podrá identificar la estructura del modelo y los patrones que se utilizan para la predicción pero, en la práctica, es imprescindible ponerle algún límite. Entonces, para definir la cantidad de datos a utilizar es necesario primero identificar las fuentes de datos disponibles y comprender sus características. Lo mismo sucede al momento de determinar los atributos apropiados que deberán ser seleccionados como variables de entrada y salida del modelo [21]. Además de analizar las fuentes de información histórica, también se deberán utilizar otras fuentes, como son las opiniones de expertos [22].

Estas necesidades no escapan a los Sistemas Inteligentes. Como dice [12], “*el Aprendizaje Automático no es magia, no se puede conseguir algo de la nada*”. Una de las cuestiones principales de la implementación de estos sistemas tiene que ver con los pasos asociados a recolectar los datos, integrarlos, limpiarlos y pre-procesarlos. Para lo cual, es imprescindible identificarlos y

conocerlos primero [23]. Otro tipo de dificultad asociada a la construcción de los MPI tiene que ver con determinar el tipo de tecnología a aplicar y las parametrizaciones a ser consideradas (como por ejemplo, su topología). De acuerdo al Teorema de ‘No Free Lunch’ [24], no existe ningún algoritmo de Aprendizaje Automático que pueda ser aplicado para cualquier problema.

Del análisis anterior se desprende que, las principales dificultades asociadas a la construcción de un MPI dependen de los aspectos relativos al problema que se quiere resolver. En tal sentido, es imprescindible conocer las expectativas que el cliente tiene sobre el modelo que se traducirán en los objetivos que deberá cumplir. Dado que los mismos normalmente estarán muy relacionados con las metas estratégicas y tácticas de la organización, también se deben entender características propias de la organización, su contexto y los procesos de negocio realizados. Una vez que estos elementos son identificados, es necesario realizar un reconocimiento inicial de las fuentes de información disponibles en la organización identificando cuáles fuentes se encuentran informatizadas y cuáles no. En el caso de los repositorios, es muy importante también, determinar la estructura, naturaleza y calidad de los datos disponibles. Por consiguiente, se considera indispensable elicitación todos estos requerimientos en forma completa y detallada para poder construir así el modelo que satisfaga las necesidades predictivas del cliente.

○ **Elicitación de Requerimientos para Modelos Predictivos Inteligentes**

Dentro de la Ingeniería de Software existe la Ingeniería de Requisitos que es la encargada de recopilar, analizar y validar los requerimientos de los distintos interesados (o stakeholders) sobre el

sistema software a implementar [25]. Entre sus actividades se aplican diferentes técnicas de elicitación, cuyo principal objetivo es entender los deseos y las necesidades de dichos interesados [26].

Sin embargo, a diferencia de los proyectos de desarrollo de software tradicional, la problemática abordada en los proyectos de construcción de Sistemas Inteligentes es diferente. Por lo tanto, existe otra disciplina denominada Ingeniería del Conocimiento o INCO [27] la cual se encuentra orientada a la construcción de Sistemas Basados en Conocimientos en general y los Sistemas Expertos en particular. Estos últimos son Sistemas Inteligentes que tienen la capacidad de emular parte del comportamiento de un experto humano para realizar una tarea en un dominio muy complejo y poco estructurado. Aunque existen metodologías que guían este proceso de Adquisición de Conocimientos [28-30], éstas se concentran en definir los hechos y reglas que serán almacenados en una Base de Conocimiento.

Esto significa que las metodologías de la INCO no son aplicables para construir otros Sistemas Inteligentes que no tengan Base de Conocimiento, como es el caso de las RNA y las RB. En función de lo que se ha identificado anteriormente, los requerimientos de dichas arquitecturas se encuentran asociados a otros elementos del problema y su dominio. A partir de un relevamiento efectuado en el campo metodológico, se han encontrado diversas publicaciones vinculadas a la definición de estos elementos tales como [23; 31- 37]. No obstante, la mayoría de estas publicaciones se encuentran más orientadas a la identificación de aspectos relativos a la construcción de RB, dejando de lado las particularidades de las RNA. Además, dentro de las propuestas orientadas a las RB, cada una se

concentra en algún aspecto dejando de lado otros por lo que no existe una guía integral para el desarrollo de esta arquitectura. Por consiguiente, se considera de interés unificar las propuestas anteriores en un único proceso que asista a las primeras etapas de la construcción de una RNA y/o una RB como MPI. Para ello, además se reutilizará la experiencia existente en el uso de las técnicas de elicitación de requerimientos de la Ingeniería de Software y la INCO, lo cual dotará al proceso de desarrollo con las características del abordaje ingenieril contribuyendo así al avance del conocimiento científico mediante el uso de técnicas consistentes.

Resultados y Objetivos

La presente línea de trabajo propone un modelo de proceso definido y estandarizado para la elicitación de los requisitos necesarios para la implementación de Modelos Predictivos mediante RNA o RB. De esta manera, se procura aportar a los Ingenieros involucrados en la implementación de estos tipos de Sistemas Inteligentes, una guía que les brinde soporte durante las etapas iniciales del proyecto teniendo presente las particularidades de los mismos.

Dentro de los objetivos específicos considerados se encuentran:

- La identificación de los conceptos necesarios a ser elicitados para la construcción de MPI.
- La especificación de los procesos necesarios para la elicitación de los conceptos identificados.
- La selección de técnicas de extracción y educación de conocimientos a ser aplicadas en dichos procesos.
- La definición del tipo de arquitectura y topología a ser aplicado para la

construcción del MPI, teniendo en cuenta los requerimientos elicitados.

Formación de Recursos Humanos

El equipo de trabajo se encuentra conformado por tres investigadores formados, un tesista de maestría, dos graduados y un alumno en la carrera de grado. Asimismo, se prevé incorporar alumnos avanzados en la carrera de Ingeniería en Sistemas de Información con posibilidades de articular sus Trabajos Finales de Carrera de Grado. De esta manera se espera generar un verdadero espacio integrado de investigación en carreras de grado y posgrado.

Referencias

- [1]. Robbins, S. & Coulter, M. (2010). *Administración*. Décima Edición. Prentice Hall.
- [2]. Arévalo, L. E. B. (2013). *La organización empresarial como sistema adaptativo complejo*. Estudios Gerenciales, 29(127), 258-265.
- [3]. Mair, C., Kadoda, G., Lefley, M., Phalp, K., Schofield, C., Shepperd, M. & Webster, S. (2000). *An investigation of machine learning based prediction systems*. Journal of Systems and Software, 53(1), 23-29.
- [4]. García, F. J. M., Martínez, M. A. P. & García, J. S. (2003). *Gestión Estratégica del Conocimiento*. Asociación Universitaria Iberoamericana de Postgrado.
- [5]. Nilsson, N. J. (2014). *Principles of Artificial Intelligence*. Morgan Kaufmann.
- [6]. Russell, S. J., Norvig, P., Davis, E., Russell, S. J. & Russell, S. J. (2010). *Artificial Intelligence: a modern approach* (Vol. 2). Englewood Cliffs: Prentice Hall.
- [7]. Shepperd, M. & Kadoda, G. (2001). *Comparing software prediction techniques using simulation*. IEEE Transactions on Software Engineering, 27(11), 1014-1022.
- [8]. Bontempi, G., Taieb, S. B. & Le Borgne, Y. A. (2013). *Machine learning strategies for time series forecasting*. In Business Intelligence (pp. 62-77). Springer Berlin Heidelberg.
- [9]. Alpaydin, E. (2014) *Introduction to machine learning*. MIT press.
- [10]. Rodríguez, M. & Márquez Alegría, M. (2015). *Manejo de problemas y toma de decisiones* (Vol. 8). Editorial El Manual Moderno.
- [11]. Haldane, A. G. & Madouros, V. (2012). *The dog and the frisbee*. Revista de Economía Institucional, 14(27), 13-56.
- [12]. Domingos, P. (2012). *A few useful things to know about machine learning*. Communications of the ACM, 55(10), 78-87.
- [13]. Cohen, P. R. & Feigenbaum, E. A. (2014). *The handbook of Artificial Intelligence*. Vol. 3. Butterworth-Heinemann.
- [14]. De Gooijer, J. G. & Hyndman, R. J. (2006). *25 years of time series forecasting*. International journal of forecasting, 22(3), 443-473.
- [15]. Wang, S. C. (2003). *Artificial Neural Network*. In Interdisciplinary Computing in Java Programming (pp. 81-100). Springer US.
- [16]. Wang, L. & Fu, K. (2009). *Artificial Neural Networks*. Wiley Encyclopedia of Computer Science and Engineering. 181-188.
- [17]. Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. The MIT Press.
- [18]. Premchaiswadi, W. (2012) *Bayesian Networks*. Ed. In-Tech.
- [19]. Zhang, G., Patuwo, B. E. & Hu, M. Y. (1998). *Forecasting with artificial neural networks: The state of the art*. International journal of forecasting, 14(1), 35-62.
- [20]. Chatfield, C. (2016). *The analysis of time series: an introduction*. CRC press.
- [21]. Walczak, S. (2001). *An empirical analysis of data requirements for financial forecasting with neural networks*. Journal of management information systems, 17(4), 203-222.
- [22]. Hyndman, R. J. & Kostenko, A. V. (2007). *Minimum sample size requirements for seasonal forecasting models*. Foresight, 6(Spring), 12-15.
- [23]. Trujillano, J., March, J. & Sorribas, A. (2004). *Aproximación metodológica al uso de redes neuronales artificiales para la predicción de resultados en medicina*. Med Clin (Barc), 122(s1).
- [24]. Wolpert, D. H. & Macready, W. G. (1997). *No free lunch theorems for optimization*. IEEE transactions on evolutionary computation, 1(1), 67-82.
- [25]. Nuseibeh, B. & Easterbrook, S. (2000). *Requirements engineering: a roadmap*. In Proceedings of the Conference on the Future of Software Engineering (pp. 35-46). ACM.
- [26]. Robertson, S. & Robertson, J. (2012). *Mastering the requirements process: Getting requirements right*. Addison-wesley.
- [27]. García Martínez, R. & Britos, P. (2004). *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería.
- [28]. Buchanan, B. G., Barstow, D., Bechtal, R., Bennett, J., Clancey, W., Kulikowski, C., Mitchell, T. & Waterman, D. A. (1983). *Constructing an expert system*. Building expert systems, 50, 127-167. Addison Wesley.
- [29]. Gómez, A., Juristo, N., Montes, C. y Pazos, J. (1997). *Ingeniería del Conocimiento*. Editorial R. Areces. Madrid.
- [30]. Schreiber, G., Akkermans, H., Anjewierden, A., Hoog, R. D., Shadbolt, N. R. & Wielinga, B. (2000). *Knowledge engineering and management: the CommonKADS methodology*. MIT press.
- [31]. Chen, S. H. & Pollino, C. A. (2012). *Good practice in Bayesian network modelling*. Environmental Modelling & Software, 37, 134-145.
- [32]. Constantinou, A. C., Fenton, N., Marsh, W. & Radlinski, L. (2016). *From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support*. Artificial intelligence in medicine, 67, 75-93.
- [33]. Henrion, M. (2013). *Practical issues in constructing a Bayes' belief network*. arXiv preprint arXiv:1304.2725.
- [34]. Masegosa, A. R. & Moral, S. (2013). *An interactive approach for Bayesian network learning using domain/expert knowledge*. International Journal of Approximate Reasoning, 54(8), 1168-1181.
- [35]. Mendes, E. (2011). *Knowledge representation using Bayesian networks - A case study in Web effort estimation*. In Information and Communication Technologies (WICT), 2011 World Congress on (pp. 612-617). IEEE.
- [36]. Pitchforth, J. & Mengersen, K. (2013). *A proposed validation framework for expert elicited Bayesian Networks*. Expert Systems with Applications, 40(1), 162-167.
- [37]. Zhou, Y., Fenton, N. & Neil, M. (2014). *Bayesian network approach to multinomial parameter learning using data and expert judgments*. International Journal of Approximate Reasoning, 55(5), 1252-1268.