

ESTIMACIÓN EN GESTIÓN DE PROYECTOS DE SOFTWARE BASADA EN EXPLOTACIÓN DE INFORMACIÓN

Bogado, Verónica¹, Dapozo, Gladys¹; García Martínez, Ramón²

1. Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura
Universidad Nacional del Nordeste

2. Grupo Investigación en Sistemas de Información. Departamento Desarrollo Productivo y Tecnológico
Universidad Nacional de Lanús

vro.s.bg@gmail.com, gndapozo@exa.unne.edu.ar, rgarcia@unla.edu.ar

RESUMEN

El desarrollo del software requiere de la estimación para controlar y administrar los recursos que se necesitan utilizar antes y durante el proyecto. Son numerosas las variables relacionadas con los recursos humanos, el contexto y las políticas que intervienen en su proceso que pueden afectar los resultados finales. Sin embargo, la sistematización de la estimación puede lograr muy buenos resultados, principalmente cuando la organización cuenta con datos históricos de estimaciones y métricas de proyectos anteriores. En este trabajo se presentan las líneas de investigación y desarrollo orientadas hacia el análisis, estudio y discusión de métodos de estimación para el desarrollo de software, aplicando técnicas de explotación de información para la construcción de modelos que aporten a la toma de decisiones en la Ingeniería de Software. Se espera que los resultados sean transferidos al medio como forma de contribuir al desarrollo de la industria del software en la región NEA.

Palabras clave: Explotación de información, Gestión de proyectos de software, Técnicas de estimación.

CONTEXTO

Las líneas de I/D presentadas en este trabajo se encuadran en las actividades de colaboración y asistencia técnica y académica especialmente orientada a la formación de recursos humanos en carreras de Informática, realizado entre la Universidad Nacional de Misiones (UNAM), Universidad Nacional del Nordeste (UNNE) y la Universidad Nacional de Lanús (UNLa).

INTRODUCCION

La construcción de métodos de estimación de proyectos de software que logren resultados predictivos sobre los recursos a emplear que se ajusten de la mejor manera posible a la realidad obtenible, es un problema abierto en el campo de los sistemas de información [1].

El desarrollo de software requiere de la estimación para controlar y administrar los recursos que se necesitan utilizar antes y durante el proyecto. No se puede considerar la estimación como una ciencia exacta, ya que existen numerosas variables humanas, técnicas, del entorno y políticas, que intervienen en su proceso y que pueden afectar los resultados finales. Sin embargo, cuando es llevada a cabo en forma sistemática, se pueden lograr resultados con un grado aceptable y convertirla en un instrumento útil para la toma de decisiones [2].

Por otra parte, el desarrollo tecnológico alcanzado en nuestros días, unido al consecuente abaratamiento de los recursos, han propiciado que cualquier organización sea capaz de almacenar todos los datos generados por su actividad. Esta facilidad para generar y almacenar información ha fomentado en los últimos años el desarrollo y perfeccionamiento de técnicas de Explotación de Información y Aprendizaje Automático para la extracción de conocimiento a partir de grandes conjuntos de datos (KDD, *Knowledge Discovery Databases*) y el desarrollo de Sistemas de Ayuda a la Toma de Decisiones (DSS, *Decisions Support Systems*).

En general, los DSS son aplicables a cualquier campo o actividad social o industrial. Las áreas de aplicación de los DSS son muy diversas,

entre ellas: la gestión y planificación de negocios, atención sanitaria, militar, y cualquier área de gestión en el que se encuentren situaciones de toma de decisión complejas. Entre estas áreas se encuentra la Ingeniería de Software y, en concreto, la gestión de proyectos de desarrollo de software donde se usan las citadas técnicas principalmente para decisiones estratégicas y tácticas en la gestión de nivel superior.

En este contexto, teniendo almacenadas las variables que miden el proceso de desarrollo de un proyecto software, se puede establecer un modelo que ayude a encontrar las condiciones más idóneas para abaratar el costo y reducir el tiempo de realización del proyecto.

En la gestión y desarrollo de sistemas de Ingeniería de Software se deben tomar múltiples decisiones en innumerables situaciones del desarrollo y mantenimiento de dichos sistemas. Las fases en las que es preciso tomar decisiones acertadas comienzan en las evaluaciones iniciales sobre la corrección de los requisitos, evaluación de las arquitecturas, diseños, prototipos y otros elementos de las estructuras de las aplicaciones. En las fases de diseño, codificación y pruebas los técnicos también deben tomar decisiones acerca de los productos intermedios. Además, en todas las actividades de gestión el responsable del proyecto debe resolver adecuadamente en las distintas situaciones que se plantean. El aporte de información, cuantitativa o cualitativa, contribuirá a mejorar el proceso de toma de decisiones [3].

Las ventajas que ofrecen las técnicas de Explotación de Información para el análisis de información, junto a la necesidad de desarrollar DSS en el campo de la Ingeniería de Software, generan interesantes posibilidades de estudio y desarrollo.

Explotación de Información

La explotación de información engloba un conjunto de técnicas encaminadas a la extracción de conocimiento procesable pudiendo abordarse la solución a problemas de predicción, clasificación y segmentación en la extracción de conocimiento no trivial que reside

de manera implícita en los datos disponibles en distintas fuentes de información [4].

La Explotación de Información puede definirse como un elemento fundamental de un proceso más amplio, el cual tiene una primera etapa de preparación de datos, luego el proceso de explotación de información, la obtención de patrones de comportamiento, y la evaluación e interpretación de los patrones descubiertos [5].

Este conjunto de técnicas pueden aplicarse a la solución de problemas de predicción, clasificación y segmentación. Los fundamentos de estas técnicas se encuentran en el análisis estadístico y en los sistemas inteligentes. Del análisis estadístico se toma las siguientes técnicas: análisis de varianza, análisis de regresión, prueba Chi-cuadrado, análisis de agrupamiento, análisis de componentes, análisis discriminante, series de tiempo. De los sistemas inteligentes se usa las siguientes tecnologías: algoritmos genéticos, algoritmos que generan árboles de decisión (TDIDT - *Top Down Induction Decision Trees*), redes neuronales, redes bayesianas [6]. La regresión logística es un método lineal que intenta modelar la probabilidad de ocurrencia de un evento de interés, donde la variable dependiente es categórica dicotómica o policotómica, a los efectos de facilitar la interpretación [7]. Las redes neuronales son modelos computacionales inspirados en las características neurofisiológicas del cerebro humano y están formadas por un gran número de neuronas dispuestas en varias capas e interconectadas entre sí mediante conexiones con pesos. Las redes neuronales constituyen herramientas analíticas que permiten examinar los datos con el objeto de descubrir y modelar las relaciones funcionales existentes entre las variables. Pueden comportarse como técnicas de aproximación o de clasificación universales [8]. Los árboles de decisión son una serie de reglas o condiciones organizadas de forma jerárquica, a modo de árbol. Son muy útiles para encontrar estructuras en espacios de alta dimensionalidad y en problemas que mezclan datos categóricos y numéricos [9].

Métodos de Estimación

Fenton y Pfleeger establecieron cuatro categorías de métodos de estimación que tienen como objetivo predecir el costo y esfuerzo del proyecto [9].

1. Opinión de expertos. Un desarrollador o gestor describe los parámetros del proyecto y los expertos realizan estimaciones basadas en su experiencia.
2. Analogía. Enfoque más formal que el anterior, los expertos comparan el proyecto propuesto con uno o más proyectos anteriores intentando encontrar similitudes y diferencias particulares.
3. Descomposición. Análisis minucioso de las características que afectan al costo del proyecto mediante la descomposición del producto o del proceso software. En el primer caso se realizan estimaciones individuales sobre los componentes en que se descompone el producto y en el segundo sobre las tareas de bajo nivel que forman parte de otras superiores. Las estimaciones de bajo nivel se pueden combinar siguiendo diferentes procedimientos para producir una estimación global sobre el proyecto completo.
4. Modelos. Técnicas que identifican los factores clave que contribuyen al esfuerzo y generan un modelo matemático que relaciona dichos factores con el esfuerzo. Los modelos se basan normalmente en información obtenida de experiencias pasadas.

Los métodos del último grupo representan el enfoque más formal y son los que proporcionan resultados más fiables. En un principio recibieron el nombre de métodos algorítmicos o paramétricos debido a que utilizaban procedimientos como la regresión para producir un modelo constituido por una o varias expresiones matemáticas que relacionan el esfuerzo con una variable primaria, generalmente el tamaño, y varios factores de ajuste secundarios. El modelo COCOMO es el más representativo de este grupo.

Actualmente, los métodos de estimación basados en modelos empíricos incorporan técnicas heurísticas consideradas no algorítmicas en el sentido de “no deterministas”, es decir, no sólo se considera la

solución, sino la probabilidad de que esa solución sea cierta. La inducción de árboles de decisión, redes neuronales, algoritmos genéticos y otras técnicas de aprendizaje automático se utilizan para construir modelos de predicción de diferentes atributos del software.

Estimación del esfuerzo

La estimación del esfuerzo de desarrollo de un proyecto software representa un papel importante dentro de la creación y mantenimiento de productos software. La misma permite evaluar la viabilidad del proyecto, así como analizar sus posibles alternativas y gestionar los recursos necesarios para el desarrollo del mismo. Por lo tanto, disponer de una buena estimación en las etapas iniciales del desarrollo del proyecto software resulta clave para las empresas de software [10].

Esta estimación del esfuerzo implica un proceso de predicción de tiempo y recursos necesarios para llevar a cabo el proyecto en tiempo y forma, teniendo como base información incierta e incompleta. Para el proceso de estimación del esfuerzo se puede usar como entradas datos tomados de planes de proyectos anteriores o de iteraciones (fases anteriores de un proyecto), presupuestos o estudios realizados (investigaciones).

Además de contar con cierta información inicial, es necesario contar con herramientas y técnicas que permitan reflejar con mayor precisión la realidad de los proyectos.

En la literatura existen varios enfoques de estimación de esfuerzo. Uno de estos enfoques es la estimación del experto, técnicas que se basan en el juicio del experto y su experiencia, reflejadas en trabajos como [11]. Existen otras propuestas que emplean modelos de regresión, los cuales usan parámetros independientes (tamaño del software entre otros), teniendo como ventaja la base matemática para medir la bondad de ajuste (determina cuanto se ajusta el modelo a los datos); sin embargo, tienen la desventaja que ser modelos muy susceptibles a valores *outlier*, además de necesitar un conjunto de datos grande lo cual es un problema en el campo de la estimación de software ([12], [13]). Otras propuestas se enfocan en la aplicación de

redes neuronales, que si bien éstas se diseñan para un conjunto de datos de entrada (puntos de función, lenguajes de implementación, etc.) como para sus salidas (esfuerzo de desarrollo), las mismas carecen de la capacidad de explicación y no proveen un entorno para una adaptación directa de los resultados ([12], [13]). Por otro lado, el razonamiento basado en casos ha demostrado ser otra opción para la estimación de esfuerzo en proyectos de software; esta técnica resuelve nuevos problemas a partir de soluciones de problemas anteriores, permitiendo de esta forma justificar los resultados, basándose en los casos previos, es muy similar a la estimación del experto [12]. La estimación del esfuerzo, a pesar de ser un tema muy discutido y existiendo gran número de propuestas, es un proceso que sigue siendo principalmente realizado por el experto.

El problema que presentan las estimaciones del esfuerzo hoy en día es la tendencia a ser sobre optimistas, llevando a que el desarrollo demore más de lo especificado en el contrato con el cliente, incrementando los costos presupuestados y no cumpliendo con los tiempos pautados. Este problema se agrava aún más en empresas medianas y pequeñas, dado que éstas tienen como objetivo prioritario el de incrementar el portfolio de proyectos y captar nuevos clientes, existiendo una tendencia a atemorizarse ante la pérdida de un cliente, prometiendo así tiempos imposibles de sobrellevar en el transcurso de la ejecución del plan.

Los modelos existentes para la estimación son variados ofreciendo valores predictivos de costos y esfuerzo para el desarrollo de software. Pero cualquiera sea el método escogido, es importante que la organización recolecte y mantenga un conjunto de datos históricos de estimaciones y métricas de proyectos anteriores, de forma que las estimaciones futuras sean más confiables.

En este contexto resulta interesante el estudio de métodos, técnicas y herramientas para estimación de esfuerzo que sean adaptables a las necesidades de las empresas de la zona, que por sus características se encuentran en una etapa inicial de madurez.

Empresas de Software y Servicios Informáticos (SSI) de la región NEA

Las empresas de SSI de las provincias de Corrientes y Chaco en forma conjunta con los respectivos gobiernos provinciales y las universidades con carreras de Informática radicadas en la región han concretado la conformación de los Polos IT Chaco (2005) e IT Corrientes (2007).

En un estudio acerca de la caracterización de estas empresas, se detectó que su perfil es similar al descrito en el informe sobre las empresas de software del país realizado por la CESSI [14]. La mayoría de las empresas tiene menos de 10 empleados, su plantel de recursos humanos está formado por profesionales informáticos de alta calificación, las actividades se orientan principalmente al desarrollo de software, las principales dificultades para el crecimiento están relacionadas con el financiamiento y el nivel económico interno, la falta de certificación de calidad y problemas de comercialización [15].

Conocer a las empresas de software, cómo funcionan y cuáles son sus dificultades, permitirá definir acciones concretas de vinculación entre las universidades de la región y las entidades productoras de software, a fin de promover e implementar mejores prácticas de desarrollo de software orientadas especialmente a las pequeñas empresas a fin de favorecer su desarrollo y contribuir a la promoción de la industria del software en la región del Nordeste Argentino (NEA).

LINEAS DE INVESTIGACION Y DESARROLLO

Las principales líneas del proyecto están orientadas hacia el análisis, estudio y discusión de métodos de estimación de desarrollo de software, aplicando técnicas de explotación de información y construcción de modelos que aporten a la toma de decisiones en la Ingeniería de Software.

Por otra parte, dentro de los objetivos del proyecto, se destaca como relevante que los resultados sean transferidos al medio como forma de contribuir al desarrollo de la industria del software en la región NEA.

RESULTADOS OBTENIDOS/ESPERADOS

En esta primera etapa del desarrollo del proyecto las actividades propuestas son:

- Relevamiento y análisis de las distintas técnicas utilizadas en la Ingeniería de Software para la estimación en proyectos de desarrollo de software.
- Investigación sobre las prácticas de estimación que utilizan las empresas de software de la región, determinación del grado de satisfacción con los resultados estimados e identificación de los distintos factores que afectan la calidad de la estimación.
- Generación de un repositorio de datos con valores de métricas de proyectos que permita la aplicación de los métodos para medir y estimar el esfuerzo necesario para el desarrollo de un nuevo proyecto.
- Especificación de un modelo de estimación/predicción de esfuerzos para el desarrollo de software utilizando alguna de las técnicas del campo de Explotación de Información.

FORMACION DE RECURSOS HUMANOS

En el marco del proyecto se desarrolla tesis de Maestría en Ingeniería de Software, tesis doctoral y planes de investigación de becarios de investigación de pregrado.

REFERENCIAS

- [1] Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. “Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información”. Anales del CACIC 2010 - XVI Congreso Argentino de Ciencias de la Computación, pag. 664-673, 2010.
- [2] Rodríguez Brito, D., Febles Estrada, A., Delgado Martínez, R., Machado Cento.A.G., Monagas Reyes, M.A. “Método de Estimación para Proyectos de Desarrollo de Software de da Universidad de las Ciencias Informáticas (UCI)”. Décima Edición de la Semana Tecnológica. Disponible en <http://semanatecnologica.fordes.co.cu/public/site/363.pdf>
- [3] Francisco Ferrer Troyano, Raúl Giráldez Rojo, Roberto Ruiz Sánchez. “Estimación y toma de decisiones mediante Minería de Datos”. Capítulo del libro “Técnicas Cuantitativas para la Gestión en Ingeniería de Software”, pag. 245-266. Editorial Netbiblo. 2007.
- [4] Britos, P. “Procesos de explotación de información basados en sistemas inteligentes”. Tesis Doctoral Universidad Nacional de la Plata. La Plata. Buenos Aires. Argentina. 2008.
- [5] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. , Uthurusamy, R. “Advances in Knowledge Discovery and Data Mining”. AAAI/MIT Press. 1996.
- [6] García Martínez, R., Servente, M., Pasquini, D. “Sistemas Inteligentes”. Editorial Nueva Librería. Buenos Aires. 2003.
- [7] Britos, P. “Minería de Datos”. Buenos Aires: Nueva Librería. 2005.
- [8] Castillo, E., Cobo, A., Gutiérrez, J. M., Pruneda, R. E. “Introducción a las Redes Funcionales con Aplicaciones. Un Nuevo Paradigma Neuronal”. Editorial Paraninfo S.A. Madrid. España. 1999.
- [9] Moreno Garcia M., Garcia Peñalvo F., “Modelos de Estimación del software basados en técnicas de aprendizaje automático”. Capítulo del libro “Técnicas Cuantitativas para la Gestión en Ingeniería de Software”. Pag. 109 a 127. Editorial Netbiblo. Año 2007.
- [10] López, J. E., Dolado, J.J., “Reducción de las probabilidades condicionadas en la estimación del esfuerzo software mediante Redes Bayesianas. Actas de los Talleres de las Jornadas de Ingeniería de Software y Bases de Datos, Vol. 4, No. 1, 2010.
- [11] Jorgensen, M. “A review of studies on expert estimation of software development effort”, The Journal of Systems and Software 70 (2004) 37–60, 2004.
- [12] Finnie, G. R.; Wittig, G. E.; Desharnais, J.M. “A Comparison of Software Effort Estimation Techniques: Using Function Points with Neural Networks, Case-Based Reasoning and Regression Models”. Journal of Systems and Software, 1997.
- [13] Bilge Başkeleş, Burak Turhan, Ayşe Bener. “Software Effort Estimation Using Machine Learning Methods”. IEEE. 2007.
- [14] Cámara de Software y Servicios Informáticos - CESSI. “Situación actual y desafíos futuros las PYMES de software y servicios informáticos”, <http://www.cessi.org.ar/index.htm>
- [15] Estayno, M., Dapozo, G.; Greiner, C.; Cuenca Pletsch, L.; Pelozo, S. “Caracterización de las pymes de software de la región NEA orientada hacia un marco de mejora de la calidad”. Anales del CACIC 2009. Pag. 901 a 910. ISBN 978-897-24068-4-1