

Estimación del esfuerzo en desarrollo de software con minería de datos

Estimating software development effort with Data Mining

Mauricio Vargas Sánchez*

Nelson Armando Vargas Sánchez**

Fecha de recepción: 23 de marzo de 2013

Fecha de aceptación: 30 de abril de 2013

Resumen

En este artículo se revisan conceptos sobre indicadores de gestión y se modela matemáticamente el cálculo del esfuerzo en desarrollo de software a través de técnicas de Minería de Datos mediante algoritmos de agrupamiento (clustering). También se da una solución general al problema y se referencia una solución particular a partir de un caso práctico tomado de una base de datos de proyectos del ISBSG (International Software Benchmarking Standards Group).

Palabras clave:

Estimación de esfuerzo, dimensionamiento de proyectos, indicadores de gestión, minería de datos, clustering.

* Ingeniero de Sistemas Universidad Nacional de Colombia, Licenciado en Matemáticas Universidad Distrital Francisco José de Caldas. Especialista en Gerencia Informática Universidad EAN. Magister en Gestión de Organizaciones Universidad EAN-Universidad du Quebec a Chicoutimi. Consultor Empresarial Estrategia Humana Ltda. Correo electrónico: mvargas.un@gmail.com

** Ingeniero de Sistemas Universidad Nacional de Colombia, Especialista en Telecomunicaciones Universidad Piloto de Colombia, Docente Universidad Distrital Francisco José de Caldas. Correo electrónico: navargass@yahoo.com

Abstract

In this paper we explore the concepts of key performance indicators or KPI and we calculate the effort of software development using a mathematical model through mining techniques with clustering algorithms. We propose a general solution to the problem and is referenced a particular solution, in a case study taken from a database project ISBSG (International Software Benchmarking Standards Group).

Keywords

Effort estimation, projects sizing, key performance indicators (KPI), data minning, clusterng.

1. Introducción

Una de las principales responsabilidades de las áreas de tecnologías de información en las organizaciones es la ejecución de proyectos de software. Éstos se desarrollan para apoyar el logro de los objetivos estratégicos, sin perder de vista la optimización adecuada de los recursos que se emplean en su ejecución. Dentro de los recursos relevantes está el tiempo que se utiliza para cumplir a cabalidad con cada una de las actividades que conforman el proyecto. El tiempo (horas/hombre) debe gestionarse para evitar atrasos, incremento en los costos, impacto financiero del proyecto, incumplimiento de compromisos contractuales y obtener recuperación de la inversión de forma temprana.

Existen varios métodos para calcular el esfuerzo en horas hombre en la ejecución de proyectos de software. Se puede utilizar el que se considere más apropiado dependiendo de la organización, del tipo de proyecto y de la experiencia. En este artículo se mostrará una solución generalizada para la cuantificación del esfuerzo en proyectos de desarrollo de software, a partir de información histórica de proyectos que se han desa-

rollado y de los cuales se tiene información que se toma como base para la realización de predicciones con el apoyo de herramientas de Minería de Datos.

2. Indicadores de gestión

Los Indicadores son la medida del estado y desempeño de un macroproceso, proceso a actividad, en un momento determinado e indican el grado en que se están logrando los objetivos. En consecuencia, se entiende por indicador el conjunto de variables cuantitativas o cualitativas que se va a medir y monitorear.

Un indicador de gestión se define como una relación entre variables que permite observar aspectos de una situación y compararlos con las metas y los objetivos propuestos. Dicha comparación permite observar la situación y las tendencias de evolución de la situación o fenómenos observados. [9]

La definición de indicadores de gestión en el ámbito tecnológico, de procesos y de servicio permite controlar oportunamente las desviaciones en el cumplimiento de los resultados esperados (metas) y por ende de

manera oportuna ejecutar los planes de acción adecuados para la corrección de dichas desviaciones. Es fundamental la definición de los Indicadores de Gestión más relevantes para el negocio, con el fin de orientar los esfuerzos en su medición, análisis de desviaciones y ajustes necesarios para apoyar convenientemente la toma de decisiones a nivel estratégico, teniendo en cuenta que el fin último de los indicadores es el control y la mejora continua de los servicios y los procesos.

La utilización de herramientas proporcionadas por disciplinas como la estadística, las probabilidades y el posicionamiento que ha tenido la minería de datos en la transformación de datos en conocimiento se muestran como alternativas fuertes para respaldar la definición de metas asertivas en el momento de definir formalmente los indicadores de gestión gerenciales en las áreas de tecnologías de información, y es por esto que se pretende formular la aplicabilidad de las mismas a partir de la información histórica de las organizaciones.

2.1 Definición del indicador

Para efectos del cálculo del indicador de la eficacia en la ejecución de proyectos de software, cada periodo de medición definido por la frecuencia del indicador (en este caso se medirá mensualmente), se debe identificar los proyectos que deben ser ejecutados y definir la meta que se tomará como referencia en dicho periodo de tiempo. La propuesta de definición de la meta se basa en el cálculo del “esfuerzo” en horas hombre previstas para la ejecución de los diferentes proyectos en el periodo de tiempo considerado.

Como elemento adicional para el cálculo del indicador se plantea hacer una ponderación para los proyectos planeados a desarrollar en un periodo de tiempo, calificando mayormente a los proyectos considerados más estratégicos dentro de la organización.

2.2 Datos para el cálculo del indicador

A continuación se ilustra de manera generalizada la fuente de los datos para el cálculo

Tabla 1. Definición del indicador

Nombre	Eficacia en la ejecución de proyectos de software
Descripción	Indica el porcentaje de cumplimiento del tiempo empleado
	(Tiempo Real) en la ejecución de los proyectos de
	desarrollo de software, respecto a los tiempos planeados
	de ejecución de los mismos
Tipo	Indicador de eficacia
Frecuencia	Mensual
Responsable	Gerente de proyectos
Formula	$100\% * \left[1 - \sum_{i=1}^N W_i * \frac{ER_i - EPI}{EPI} \right]$
Meta	$\sum_{i=1}^n W_i * EPI$

Fuente: elaboración propia.

Tabla 2. Datos para el cálculo del indicador de ejecución de proyectos

Proyecto	Peso	Esfuerzo Planeado	Esfuerzo Real
Proyecto 1	W1	EP1	ER1
Proyecto 2	W2	EP 2	ER 2
...
Proyecto K	Wk	EP k	ER k
...
Proyecto N	WN	EP N	ER N

Fuente: elaboración propia.

culo del Indicador: proyectos planeados a desarrollar en un periodo de tiempo (proyecto), ponderaciones individuales (peso), metas definidas como el esfuerzo en horas hombre (esfuerzo planeado) y la medición del esfuerzo en horas hombre para cada proyecto (esfuerzo real).

2.3 Cálculo del indicador

Con los datos contenidos en la tabla anterior se procede al cálculo del indicador de acuerdo a la formulación definida en la Hoja de Vida del Indicador:

$$100\% * \left[1 - \sum_{i=1}^N W_i * \frac{ER_i - EP_i}{EP_i} \right]$$

Donde:

W_i: Corresponde al peso ponderado de cada proyecto.

ER_i: es el esfuerzo real (en horas/hombre) del i-ésimo proyecto.

EP_i: es el esfuerzo planeado (en horas/hombre) del i-ésimo proyecto relevantes y el esfuerzo en horas hombre utilizado para la ejecución de los mismos.

2.4 Definición de la meta

Para este caso la meta a calcular es el "Esfuerzo" en horas hombre necesario para la ejecución de cada uno de los proyectos planeados en el periodo de tiempo que se está considerando.

La solución a este problema se propone conseguir utilizando herramientas de Minería de Datos, para lo cual se requiere tener una fuente de datos histórica de proyectos en la que se indique las características (atributos) más relevantes y el esfuerzo en horas hombre utilizado para la ejecución de los mismos.

2.5 Atributos relevantes de los proyectos

El esfuerzo requerido para la ejecución de un proyecto de software depende de muchos atributos que tienen un impacto directo en el mismo. Dentro de los más relevantes se pueden mencionar: Tipo de aplicación, tamaño del equipo de proyecto, tipo de lenguaje de programación, usuarios de las áreas de negocio, puntos de función, utilización de herramientas CASE, metodología de desarrollo, arquitectura, etc.

3. Solución general del problema

A partir de la información histórica de los proyectos se debe definir una función matemática que pronostique el esfuerzo necesario para el desarrollo de nuevos proyectos. Esto lo podemos expresar de la siguiente manera:

En la tabla anterior podemos ver representado cada uno de los proyectos históricos (P₁, P₂, ..., P_k) con sus valores para cada uno de sus atributos (X₁, X₂, ..., X_n) y el esfuerzo en horas hombre (E₁, E₂, ..., E_k) utilizado. Lo que pretendemos pronosticar es el esfuerzo (E_{k+1}) necesario para la ejecución de un nuevo proyecto de software (P_{k+1}).

Tabla 3. Atributos para el Cálculo del Esfuerzo

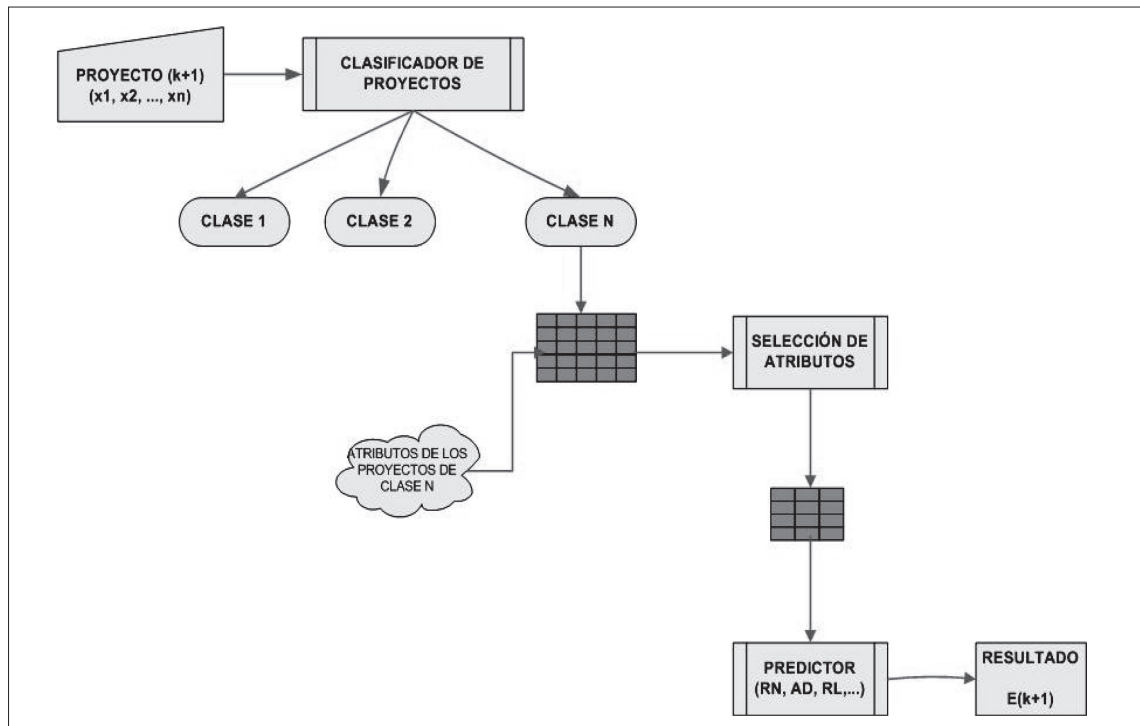
Esfuerzo = F (X₁, X₂, ..., X_n) = F(Xⁿ)

		ATRIBUTOS					ESFUERZO
		A ₁	A ₂	A _n	
PROYECTOS	P ₁	₁ X ₁	₁ X ₂	₁ X _n	E ₁
	P ₂	₂ X ₁	₂ X ₂	₂ X _n	E ₂

	P _k	_k X ₁	_k X ₂	_k X _n	E _k
P _{k+1}	_(k+1) X ₁	_(k+1) X ₂	_(k+1) X _n	E _{k+1}	

Fuente: elaboración propia.

Figura 1. Generalización de la Solución



Fuente: elaboración propia.

Existen varias alternativas para encontrar la solución, dentro de las cuales están las siguientes: Métodos Analíticos (Función Lineal, Función Polinómica, Función de Distribución de Probabilidad), Métodos de Aproximación Numérica (Simulación Numérica, Minería de Datos).

La propuesta es utilizar técnicas de Minería de Datos mediante la ejecución de algoritmos de Agrupamiento (Clustering) con el fin de "segmentar" en grupos de proyectos similares para luego implementar en cada uno de estos grupos un algoritmo de regresión que permita pronosticar el esfuerzo en horas hombre de los nuevos proyectos considerados similares.

3.1 Cálculo del esfuerzo de proyectos

El nuevo proyecto P_{k+1} se representa por un vector n-dimensional donde cada una de sus componentes son los atributos o características relevantes, $P_{k+1} = \langle X_{k+1,1}, X_{k+1,2}, \dots, X_{k+1,n} \rangle$. Este proyecto debe ser objeto de una clasificación realizada por la utilización de un algoritmo de agrupamiento (Clustering) el cual generará N diferentes grupos de proyectos (Clase 1, Clase 2, ..., Clase N).

Posteriormente para cada clase de proyectos se utilizará un algoritmo de Selección de Atributos con el fin de reducir el número de atributos identificando aquellos de mayor relevancia. De esta manera tendremos un modelo simplificado con una aceptable representación del caso de estudio.

Finalmente con la utilización de un algoritmo de predicción numérica (utilizando redes neuronales, árboles de decisión, regresión lineal o no lineal) se establece la fórmula matemática que determina el esfuerzo promedio para los proyectos que hacen parte de la Clase N.

4. Solución al caso particular

A continuación se hará referencia a una solución particular a partir de un caso prác-

tico realizado por el ISBSG (International Software Benchmarking Standards Group) en el que se toma como referencia una base de datos de 709 proyectos de desarrollo de software con características disímiles que sugiere una agrupación de proyectos similares para poder predecir con mayor precisión el esfuerzo necesario para futuros proyectos.

En este caso se propone trabajar el caso con un atributo que representa el "tamaño" de un proyecto. Este atributo se conoce como "Puntos de Función" de un proyecto, el cual es una combinación de algunos atributos representativos de los proyectos de desarrollo de software que se mencionaron con anterioridad (arquitectura, lenguaje de programación, tipo de aplicación, etc.) [2]

5. Referencias

- [1] J. Bendar y Robersto., *Estimating size and effort*. Recuperado de <http://www.inf.ed.ac.uk/teaching/courses/sapm/2006-2007/slides/estimation.pdf>; consultado el 20 Nov 2012.
- [2] J. Cuadrado, M. Sicilia y M. Garre, M. *Segmentación recursiva de proyectos software para la estimación del esfuerzo de desarrollo de software*. Recuperado de <http://www.sc.ehu.es/jiwdocoj/remis/docs/CuadradoAdis2004.pdf>; consultado el 10 Dec 2012.
- [3] D. Garmus, *The principles of sizing and estimating projects using international function points user group (IFPUG) function points*. Recuperado de <http://www.compaid.com/caiinternet/ezone/garmus-principlesofsizing.pdf>; consultado 25 Nov 2012.
- [4] Hernández, J., Ramírez, M. J., y Ferri, C. *Introducción a la Minería de Datos*. Pearson Prentice Hall, Madrid, 2004.
- [5] Jain, A., Murty, M., y Flynn, P. *Data clustering: A review*. Recuperado de <http://www.cs.rutgers.edu/~mlittman/courses/lightai03/jain99data.pdf>; consultado 15 Ene 2013.

- [6] J. M. Jaramillo, *Indicadores de Gestión: Herramientas para lograr la competitividad*. 3R Editores, Bogotá, 1998.
- [7] R. Kaplan y D. Norton, *The Balanced Scorecard*. HBS Press, Boston, 1996.
- [8] C. Pérez y D. Santin, *Minería de Datos: Técnicas y Herramientas*. Paraninfo S.A., Madrid, 2007.
- [9] H. Serna, *Indices de Gestión: Cómo diseñar un sistema integral de medición de gestión*. 3R Editores, Bogotá, 2005.
- [10] SPPS Andino. *Información Creativa*. Recuperado de <<http://www.catalogodesoftware.com/otrosproductos.aspx?aid=256>>; consultado el 13 Nov 2012.