

Segmentación Recursiva de Proyectos Software para la Estimación del Esfuerzo de Desarrollo Software

J. Cuadrado Gallego¹, Miguel Ángel Sicilia¹, Miguel Garre Rubio¹

¹ Dpto de Ciencias de la Computación, Universidad de Alcalá,
28871, Ctra. de Barcelona, 33.6, Alcalá de Henares, España

jjcg@infor.uva.es
msicilia@uah.es
miguel.garre@uah.es

Abstract. En este artículo se continúa con el trabajo realizado por J. Cuadrado y M.A. Sicilia sobre la idea de mejora en la estimación del esfuerzo en el desarrollo de software. Esta mejora se debe al hecho de usar varios modelos de regresión, uno por grupo de proyectos obtenido, en vez de uno solo que incluya a todos los proyectos. Estos proyectos son heterogéneos e inducen a pensar que la estimación sería más fiable si se agrupase por proyectos lo más homogéneos posible entre sí. Por lo tanto, hemos realizado nuevas agrupaciones recursivas, utilizando la base de datos ISBSG y el algoritmo EM, hasta llegar a obtener modelos de regresión segmentados con diferentes parámetros que nos proporcionan unas estimaciones bastante aceptables. Por último hemos comprobado que haciendo sucesivas divisiones recursivas llegamos a clusters indivisibles, cuyas características proporcionan las mejores estimaciones posibles.

1 Introducción

La correcta estimación del esfuerzo que va a requerir realizar un proyecto software es un aspecto crucial para una empresa de desarrollo software. Una estimación errónea puede hacerle perder mucho dinero, además de a futuros clientes. De ahí la importancia de que las estimaciones se realicen de la forma más eficaz posible.

Tal y como se indica en el artículo de J. Cuadrado², los métodos de estimación del esfuerzo basados en la utilización de técnicas estadísticas aplicadas a bases de datos de proyectos históricas, proporcionan ecuaciones matemáticas en las que la variable dependiente es el esfuerzo o el tiempo, y las variables independientes son diferentes aspectos del proyecto o del producto o de ambos. Estas ecuaciones son como por ejemplo las que utilizan la función potencia del tipo $e=a \cdot s^b$, donde e es el esfuerzo estimado y s alguna medida del tamaño del proyecto. El utilizar una única ecuación de este tipo para toda la base de datos de proyectos, donde los proyectos son heterogéneos, proporciona unos resultados muy pobres. Por ejemplo, utilizando la herramienta Reality de la base de datos de proyectos ISBSG versión 8 (Internacional Software Benchmarking Standard Group[†]), aplicada a 709 proyectos, se obtiene la siguiente ecuación:

$$e=47.73 \cdot s^{0.76}, \quad (1)$$

donde el esfuerzo se expresa en horas, y el tamaño en puntos de función. Un análisis de la bondad del ajuste nos da como resultado $MMRE=1.18$ y $PRED(.30)=25.6\%$. Ambas medidas son difícilmente aceptables, dado el alto grado de desviación sobre la inmensa mayoría de los datos.

Algunos autores^{1,4} han sugerido que la segmentación de los datos contenidos en las bases de datos históricas podría ser un camino adecuado para la obtención de ecuaciones matemáticas que proporcionen una mayor exactitud en las estimaciones. Una forma de obtener mejores ajustes consiste en utilizar algoritmos de agrupamiento (*clustering*) conocidos, para dividir el dominio de proyectos, y en caso de que el ajuste aún no sea “suficientemente” bueno, producir de forma recursiva nuevas divisiones.

[†] <http://www.isbsg.org/>

Los resultados obtenidos, aplicando el algoritmo EM sobre los 709 proyectos, son los siguientes:

Table 1. Calidad de los modelos de regresión obtenidos tras la segunda segmentación.

	Modelo	MMRE	PRED(.3)	MMRE (original)	PRED(.3) (original)	Nº Proyectos
Cluster 0-0	$e = 246,9 \cdot fp^{-0,0225}$	0,47	34,4%	3,2	6%	32
Cluster 0-1	$e = 5716 \cdot fp^{-0,4994}$	0,37	55%	1,9	21,5%	102
Cluster 0-2	$e = 54220 \cdot fp^{-0,8357}$	0,13	94%	1,71	30,5%	36
Cluster 1	$e = 9647 \cdot fp^{-0,3}$	0,39	46,2%	1,03	30%	255
Cluster 2	$e = 26670 \cdot fp^{-0,27}$	0,37	53%	0,78	20,5%	185
Cluster 3	$e = 29120 \cdot fp^{-0,083}$	0,53	43%	0,74	18%	88
Cluster 4	$e = 152600 \cdot fp^{-0,1286}$	0,28	72%	0,6	9%	11
Media		0,36	57%			

Se han obtenido siete clusters, con valores de MMRE y PRED(.30) mucho mejores que los que proporciona el algoritmo *Reality* para todos los proyectos. Sin embargo estos valores podrían mejorar si se siguiese el proceso de división recursiva. Y esto es lo que vamos a realizar en el presente trabajo: partiendo de los datos obtenidos en la tabla 1, realizar sucesivas segmentaciones hasta obtener unos valores medios de $MMRE \leq 0,25$ y $PRED(.3) \geq 75\%$. Más aún, trataremos de obtener estos valores para cada cluster obtenido. Y por último, establecer cuál debe ser el límite de división realizado, si es que existe.

El resto del artículo se organiza de la siguiente manera. En la segunda sección se describe la técnica utilizada para la obtención del modelo de estimación segmentado. En la tercera sección se describe como caso de estudio un análisis cuantitativo de la efectividad de la técnica sobre los datos de la tabla 1. Finalmente, las conclusiones de la investigación y algunas posibles líneas de continuación futura se describen en la cuarta sección.

2 Descripción de la técnica de obtención del modelo segmentado

La técnica de segmentación y regresión recursiva puede describirse esquemáticamente mediante el siguiente pseudocódigo:

```

Generar_modelo(CD:conjunto_datos, MMRE_deseado, PRED_deseado,
               PRED_PARAM: real; TIPO_MODELO: funcion): Modelo;
MODELO:=obtener_modelo(CD, TIPO_MODELO);
MMRE := calcular_mmre(CD, MODELO);
PRED := calcular_pred(CD, PRED_PARAM, MODELO);
Si (MMRE >= MMRE_deseado o PRED <=PRED_deseado
    o <<pocos datos>>) Entonces
    Clusters :=Dividir(CD);
    Desde i:=1 hasta Clusters.longitud hacer
        Modelos[i]:=Generar_modelo(CD{Clusters[i]},
                                   MMRE_Deseado, PRED_Deseado, PRED_PARAM,
                                   TIPO_MODELO);
    Fin_Desde
Return Modelos;
Fin_Si

```

```
Return MODELO;  
Fin
```

En el pseudocódigo anterior, se asume que la salida del algoritmo de *clustering* (invocación a *Dividir*) es una estructura de datos con tantas posiciones como grupos obtenidos, en la cual, `Clusters[i][centro]` indica el “centro” estimado del cluster, `Clusters[i][alcance]` el alcance estimado del radio del grupo respecto al centro, y `Clusters[i][tamaño]` el tamaño aproximado en número de elementos del grupo. Esos datos se utilizan para filtrar el conjunto de datos original, lo cual se expresa mediante $CD\{Clusters[i]\}$.

El algoritmo descrito podría aún hacerse más genérico en varias direcciones. Por ejemplo, podrían especificarse otros tipos o criterios de medidas de calidad del ajuste, utilizar varios algoritmos de agrupamiento simultáneamente, examinando qué alternativa es mejor, o podrían utilizarse combinaciones de modelos matemáticos para las funciones de estimación.

3 Caso de estudio

Nuestro objetivo es obtener unos modelos de regresión segmentados con un valor medio de $MMRE \leq 0,25$ y $PRED(.30) \geq 75\%$. Para ello partimos de los datos obtenidos en la tabla 1, y aplicamos de forma recursiva el algoritmo EM, sobre los clusters que creemos pueden ofrecer mejores resultados.

Como algoritmo de *clustering*, se ha utilizado la implementación del algoritmo EM proporcionada en las herramientas de fuente abierto WEKA[†]. El algoritmo de agrupamiento EM (Expectation-Maximization)³ es una variante de algoritmos de agrupamiento en árbol que calcula probabilidades de pertenencia a un cluster, basadas en una o más distribuciones de probabilidad, utilizando como objetivo la maximización de la probabilidad global. Este algoritmo es especialmente adecuado para casos en los que no se precisa pre-determinar el número de clusters que han de generarse como salida. Dado que en el contexto que aquí tratamos no se tiene ninguna información *a priori* sobre la forma o número de grupos en la base de datos de proyectos, el algoritmo EM es preferible a algoritmos que predeterminan el número de grupos de salida, aunque EM también se puede utilizar indicando el número de clusters de salida que queremos obtener. El algoritmo EM asigna una probabilidad de pertenencia a cada elemento de cada cluster, sin embargo en este estudio se utiliza una asignación de puntos a clusters determinados, determinada por el centro (media) de los clusters en el dominio del tamaño del software, que será el que se utilice al hacer estimaciones de proyectos concretos.

En lo que sigue se describen los datos obtenidos según el esquema algorítmico descrito en la sección anterior.

Lo primero que nos planteamos es la elección del primer cluster a subdividir, para ello observamos la tabla 1 y vemos que el cluster 0-0 tiene un valor muy bajo de $PRED(.30)$, sin embargo se descarta por estar formado por muy pocos elementos. Otro posible candidato es el cluster 3, ya que proporciona el peor valor de $MMRE$ ($MMRE=0,53$), por lo tanto parece muy factible su posibilidad de mejora. Éste será, por lo tanto, el primer cluster que subdividamos. Al aplicar el algoritmo EM sobre los 88 proyectos del cluster 3 obtenemos los siguientes clusters de segundo nivel:

Table 2. Resultados de la segmentación del cluster 3

	Mean (fp)	Std. Dev (fp)	Approx. Size
Cluster 3-0	1837,1565	624,8073	31
Cluster 3-1	1451,0797	783,1552	19
Cluster 3-2	518,519	267,1391	38

Teniendo en cuenta los tres nuevos sub-grupos obtenidos, el análisis de la calidad del ajuste queda como muestra la siguiente tabla.

[†] <http://www.cs.waikato.ac.nz/~ml/weka/>

Table 3. Calidad de los modelos de regresión obtenidos tras la segmentación

	Modelo	MMRE	PRED(.3)	MMRE (original)	PRED(.3) (original)
Cluster 0-0	$e = 246,9 \cdot fp^{-0,0225}$	0,47	34,4%	3,2	6%
Cluster 0-1	$e = 5716 \cdot fp^{-0,4994}$	0,37	55%	1,9	21,5%
Cluster 0-2	$e = 54220 \cdot fp^{-0,8357}$	0,13	94%	1,71	30,5%
Cluster 1	$e = 9647 \cdot fp^{-0,3}$	0,39	46,2%	1,03	30%
Cluster 2	$e = 26670 \cdot fp^{-0,27}$	0,37	53%	0,78	20,5%
Cluster 3-0	$e = 147500 \cdot fp^{-0,4193}$	0,24	71%	1,03	6,5%
Cluster 3-1	$e = 117200 \cdot fp^{-0,2968}$	0,13	84,2%	0,56	26,3%
Cluster 3-2	$e = 108400 \cdot fp^{-0,2022}$	0,28	63,15%	0,57	23,68%
Cluster 4	$e = 152600 \cdot fp^{-0,1286}$	0,28	72%	0,6	9%
Media		0,30	63,66%		

Los resultados continúan mejorando, nos hemos acercado al objetivo, valores medios de $MMRE \leq 0,25$ y $PRED(.3) \geq 75\%$, pero seguimos sin alcanzarlo. Vamos a intentar conseguirlo con una nueva división. En esta ocasión seleccionaremos el cluster 1, por la misma razón que seleccionamos anteriormente el cluster 3.

Si tomamos los 255 proyectos que forman el clúster 1 y volvemos a aplicar el algoritmo EM, obtenemos los siguientes grupos de segundo nivel.

Table 4. Resultados de la segmentación del cluster 1

	Mean (fp)	Std. Dev (fp)	Approx. Size
Cluster 1-0	146,7543	60,2494	70
Cluster 1-1	504,2978	145,4643	19
Cluster 1-2	270,6069	76,3552	95
Cluster 1-3	124,0791	39,6394	71

Teniendo en cuenta los cuatro nuevos sub-grupos obtenidos, el análisis de la calidad del ajuste queda como muestra la siguiente tabla.

Table 5. Calidad de los modelos de regresión obtenidos tras la segmentación

	Modelo	MMRE	PRED(.3)	MMRE (original)	PRED(.3) (original)
Cluster 0-0	$e = 246,9 \cdot fp^{-0,0225}$	0,47	34,4%	3,2	6%
Cluster 0-1	$e = 5716 \cdot fp^{-0,4994}$	0,37	55%	1,9	21,5%
Cluster 0-2	$e = 54220 \cdot fp^{-0,8357}$	0,13	94%	1,71	30,5%
Cluster 1-0	$e = 19360 \cdot fp^{-0,5394}$	0,18	90%	0,9	38,57%
Cluster 1-1	$e = 56610 \cdot fp^{-0,6869}$	0,08	94,73%	0,66	63,15%
Cluster 1-2	$e = 38360 \cdot fp^{-0,5573}$	0,20	84,21%	1,1	26,31%
Cluster 1-3	$e = 62320 \cdot fp^{-0,5721}$	0,23	85,9%	1,18	39,43%
Cluster 2	$e = 26670 \cdot fp^{-0,27}$	0,37	53%	0,78	20,5%

Cluster 3-0	$e = 147500 \cdot fp^{-0.4193}$	0,24	71%	1,03	6,5%
Cluster 3-1	$e = 117200 \cdot fp^{-0.2968}$	0,13	84,2%	0,56	26,3%
Cluster 3-2	$e = 108400 \cdot fp^{-0.2022}$	0,28	63,15%	0,57	23,68%
Cluster 4	$e = 152600 \cdot fp^{-0.1286}$	0,28	72%	0,6	9%
Media		0,25	73,47%		

A pesar de que los resultados obtenidos en esta ocasión están muy cercanos a los deseados, debemos intentar seguir mejorando, ya que aún quedan grupos que lo permitirán con toda seguridad. En concreto el cluster 2. Éste será el que utilicemos a continuación para la subdivisión.

Si tomamos los 185 proyectos del clúster 2 y volvemos a aplicar el algoritmo EM, obtenemos los siguientes grupos de segundo nivel.

Table 6. Resultados de la segmentación del cluster 2

	Mean (fp)	Std. Dev (fp)	Approx. Size
Cluster 2-0	301,886	146,5272	77
Cluster 2-1	237,3868	94,7184	55
Cluster 2-2	744,8528	322,8134	53

Teniendo en cuenta los tres nuevos sub-grupos obtenidos, el análisis de la calidad del ajuste queda como muestra la siguiente tabla.

Table 7. Calidad de los modelos de regresión obtenidos tras la segmentación

	Modelo	MMRE	PRED(.3)	MMRE (original)	PRED(.3) (original)
Cluster 0-0	$e = 246,9 \cdot fp^{-0.0225}$	0,47	34,4%	3,2	6%
Cluster 0-1	$e = 5716 \cdot fp^{-0.4994}$	0,37	55%	1,9	21,5%
Cluster 0-2	$e = 54220 \cdot fp^{-0.8357}$	0,13	94%	1,71	30,5%
Cluster 1-0	$e = 19360 \cdot fp^{-0.5394}$	0,18	90%	0,9	38,57%
Cluster 1-1	$e = 56610 \cdot fp^{-0.6869}$	0,08	94,73%	0,66	63,15%
Cluster 1-2	$e = 38360 \cdot fp^{-0.5573}$	0,20	84,21%	1,1	26,31%
Cluster 1-3	$e = 62320 \cdot fp^{-0.5721}$	0,23	85,9%	1,18	39,43%
Cluster 2-0	$e = 38520 \cdot fp^{-0.3942}$	0,20	93,5%	0,72	28,57%
Cluster 2-1	$e = 122300 \cdot fp^{-0.5247}$	0,19	89%	0,90	16,36%
Cluster 2-2	$e = 64340 \cdot fp^{-0.3453}$	0,18	94,3%	0,73	13,2%
Cluster 3-0	$e = 147500 \cdot fp^{-0.4193}$	0,24	71%	1,03	6,5%
Cluster 3-1	$e = 117200 \cdot fp^{-0.2968}$	0,13	84,2%	0,56	26,3%
Cluster 3-2	$e = 108400 \cdot fp^{-0.2022}$	0,28	63,15%	0,57	23,68%
Cluster 4	$e = 152600 \cdot fp^{-0.1286}$	0,28	72%	0,6	9%
Media		0,23	78,96%		

Estos valores son muy buenos y podríamos parar, ya que se cumplen los requisitos de un valor medio de $MMRE \leq 0,25$ y $PRED(.3) \geq 75\%$, pero creemos interesante fijarnos de forma individual en los clusters obtenidos e intentar mejorar alguno más de ellos. De los que tienen valores mejorables, el único que proporciona suficientes elementos para hacer una subdivisión aceptable es el cluster 0-1, el cuál será el último que abordaremos.

Si tomamos los 102 proyectos del clúster 0-1 y volvemos a aplicar el algoritmo EM, obtenemos los siguientes grupos de tercer nivel.

Table 8. Resultados de la segmentación del cluster 0-1

	Mean (fp)	Std. Dev (fp)	Approx. Size
Cluster 0-1-0	120,938	42,1837	44
Cluster 0-1-1	66,9898	21,6657	58

Teniendo en cuenta los dos nuevos sub-grupos obtenidos, el análisis de la calidad del ajuste queda como muestra la siguiente tabla.

Table 9. Calidad de los modelos de regresión obtenidos tras la segmentación

	Modelo	MMRE	PRED(.3)	MMRE (original)	PRED(.3) (original)
Cluster 0-0	$e = 246,9 \cdot fp^{-0,0225}$	0,47	34,4%	3,2	6%
Cluster 0-1-0	$e = 9829 \cdot fp^{-0,7115}$	0,23	68,18%	2,3	11,36%
Cluster 0-1-1	$e = 21650 \cdot fp^{-0,733}$	0,18	81%	1,61	29,31%
Cluster 0-2	$e = 54220 \cdot fp^{-0,8357}$	0,13	94%	1,71	30,5%
Cluster 1-0	$e = 19360 \cdot fp^{-0,5394}$	0,18	90%	0,9	38,57%
Cluster 1-1	$e = 56610 \cdot fp^{-0,6869}$	0,08	94,73%	0,66	63,15%
Cluster 1-2	$e = 38360 \cdot fp^{-0,5573}$	0,20	84,21%	1,1	26,31%
Cluster 1-3	$e = 62320 \cdot fp^{-0,5721}$	0,23	85,9%	1,18	39,43%
Cluster 2-0	$e = 38520 \cdot fp^{-0,3942}$	0,20	93,5%	0,72	28,57%
Cluster 2-1	$e = 122300 \cdot fp^{-0,5247}$	0,19	89%	0,90	16,36%
Cluster 2-2	$e = 64340 \cdot fp^{-0,3453}$	0,18	94,3%	0,73	13,2%
Cluster 3-0	$e = 147500 \cdot fp^{-0,4193}$	0,24	71%	1,03	6,5%
Cluster 3-1	$e = 117200 \cdot fp^{-0,2968}$	0,13	84,2%	0,56	26,3%
Cluster 3-2	$e = 108400 \cdot fp^{-0,2022}$	0,28	63,15%	0,57	23,68%
Cluster 4	$e = 152600 \cdot fp^{-0,1286}$	0,28	72%	0,6	9%
Media		0,21	79,97%		

La media de valores de MMRE y PRED(.3) es bastante buena, 0,21 y 79,97% respectivamente. Respecto a cada uno de los clusters obtenidos, el único que sigue teniendo valores “malos” es el cluster 0-0, MMRE=0,47 y PRED(.3)=34,4%, el cual podríamos subdividirlo, pero saldrían clusters demasiado pequeños y por lo tanto nada representativos. Los clusters 3-2 y 4 se consideran como buenos.

Como podemos observar de la tabla 9, hemos obtenido 15 clusters que dan lugar a 15 ecuaciones, las cuales proporcionan unas estimaciones bastante aceptables, mucho mejores que si se utiliza una única ecuación para estimar el esfuerzo.

De esta manera, cuando queramos estimar el esfuerzo de un nuevo proyecto lo haremos en dos pasos:

1º Asignar el proyecto a uno de los clusters obtenidos.

2º Aplicar la ecuación correspondiente a este cluster sobre los datos del proyecto, y así obtener el esfuerzo estimado.

El primer paso lo haremos teniendo en cuenta los valores medios de cada uno de los clusters obtenidos. Se calculan las diferencias entre el valor fp del proyecto y el valor medio de fp de cada grupo, la menor de ellas nos indicará el cluster más probable al que pertenece. Si hubiese probabilidad de pertenencia a más

de un grupo optamos por aplicar la ecuación de los grupos a los que podría pertenecer y posteriormente se realiza una media aritmética del resultado obtenido. De este modo se obtiene una solución de compromiso entre estimaciones de clusters a los cuales potencialmente podría pertenecer el punto dado.

Por último cabe preguntarse que sucedería si continuamos aplicando el proceso de subdivisión. Lo hemos intentado con los siguientes clusters y ha sucedido lo siguiente:

- Cluster 0-0: no da lugar a nuevas subdivisiones. Se queda finalmente un grupo de 32 proyectos, y con valores $MMRE=0,47$ y $PRED(.30)=34,4\%$.
- Cluster 0-2: obtenemos 4 nuevos clusters, los cuales tienen aproximadamente un valor de $PRED(.30)$ cercano o igual al 100%. Para cada uno de los cuatro clusters obtenidos intentamos una nueva subdivisión, no siendo posible en esta ocasión.

De estos datos podemos concluir que cuando el algoritmo EM obtiene unos agrupamientos “óptimos”, no proporciona nuevas subdivisiones sobre ellos. Lo cual parece lógico, si ya tenemos lo que buscamos, para qué seguir buscando.

La palabra “óptimo” en este contexto significa tener clusters con valores $PRED(.30)$ cercanos al 100%, o si no es así, llegar a clusters cuya mejora no es posible, tal y como sucede con el cluster 0-0

4 Conclusiones y trabajo futuro

La principal conclusión que podemos sacar del estudio realizado, es que no tiene sentido utilizar un único modelo de regresión, obtenido a partir de la totalidad de los proyectos pasados de una base de datos heterogénea, para estimar esfuerzos futuros. Es mucho más útil utilizar algún algoritmo de agrupamiento, que en base a parte de los atributos de los proyectos, nos proporcione grupos de proyectos con comportamientos similares. Más aún, esta división puede realizarse de forma recursiva, de manera que con cada subdivisión aumentemos la similitud de los miembros al grupo, y por lo tanto los valores perseguidos de $MMRE$ y $PRED(.30)$.

Hemos observado que esta subdivisión podemos realizarla hasta llegar a los niveles deseados, es decir, aquellos que nos proporcionan clusters con valores de $MMRE \leq 0,25$ y $PRED(.30) \geq 75\%$. Incluso si seguimos dividiendo llegamos a obtener clusters “terminales”, que no proporcionan nuevas subdivisiones, y alcanzan valores $PRED(.30) \approx 100\%$. De ello podemos derivar a futuros trabajos, ya que podríamos plantearnos remodelar el algoritmo de obtención de clusters, de manera que realizase subdivisiones hasta alcanzar ciertos valores de $MMRE$ y $PRED(.30)$, o que realizase subdivisiones hasta obtener clusters “terminales”, con la restricción de que estos clusters fuesen de un tamaño mínimo, por debajo del cual no serían útiles para realizar estimaciones.

Otros trabajos a los que podría dar lugar el estudio realizado, nos dirigirían a estudiar otros algoritmos de agrupamiento con el fin de utilizar los más apropiados para nuestra tarea, y poder compararlos entre sí. Además habría que intentar utilizar algunos atributos más, no solamente el esfuerzo y los puntos de función, sino también quizás el tipo de aplicación, tipo de herramienta utilizada en la implementación, etc. Esto puede que ayudase a obtener mejores agrupamientos, y si no es así porqué.

Por último habría que intentar mejorar la asignación a clusters, y esto teniendo en cuenta cada tipo de algoritmo de clusterización utilizado.

5 Agradecimientos

Este trabajo ha sido realizado con el soporte del proyecto CICYT: IN2GESOFT. TIN2004-06689-C03-00. INnovación e INtegración de métodos para el desarrollo y GESTión cuantitativa de proyectos SOFTWARE.

6 Referencias

1. Abran, A. (2003). Software Estimation: Black Box or White Box. Presentado en el Workshop ADIS 2003.

2. Cuadrado, J., Rodríguez, D., Sicilia, M.A. (2004). Modelos segmentados de estimación del esfuerzo de desarrollo del software: Un caso de estudio con la base de datos ISBSG. Presentado en
3. Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via de EM algorithm. *Journal of the Royal statistical Society, Series B*, 39(1): 1-38.
4. Reifer, D., Boehm, B., Chulani, S. (1999). The Rosetta Stone. Making COCOMO 81 Estimates Work with COCOMO II. *CrossTalk*, 12(2), 11-15.