

Tabla de contenidos

Tabla de contenidos	VII
Índice de tablas	XI
Índice de figuras	XII
Resumen	XV
Abstract	XVII
Agradecimientos	XIX
1. Introducción	1
1.1. Planteamiento del problema	1
1.2. Objetivos y aportaciones originales	4
1.3. Método de trabajo	6
1.3.1. Planificación general	6
1.3.2. Método de evaluación	8
1.4. Estructura del documento	8
2. Estado de la cuestión	11
2.1. Introducción	11
2.2. Clasificación de los Modelos de Estimación del Coste del Software	12
2.2.1. Clasificación de Barry Boehm (1981)	12
2.2.2. Clasificación de DeMarco(1982)	16
2.2.3. Clasificación de Conte, Dunsmore y Shen (1986)	16
2.2.4. Clasificación de Kitchenham 1991 y 1996	18
2.2.5. Clasificación de Fairley 1992	19
2.2.6. Clasificación de Walkerden y Jeffery 1997	20
2.2.7. Clasificación realizada por Barry Boehm, Chris Abts y Sunita Chulani en el año 2000	21
2.2.8. Clasificación de Wieczorek 2001	27
2.2.9. Estado actual	29

2.3.	Modelos de Estimación del Coste del Software	47
2.3.1.	SDC - System Development Corporation (1965-1967)	47
2.3.2.	Modelo de “Farr and Zagorski” 1965	48
2.3.3.	Aron (1969)	49
2.3.4.	Modelo NAVAIR - NAVal AIR Development Center Model (1971)	50
2.3.5.	PRICE H - RCA (1973)	50
2.3.6.	TRW Wolverton TRW (1974)	50
2.3.7.	Tecolote - Tecolote Research, Inc. (1974)	51
2.3.8.	GRC - General Research Corporation (1974-79)	51
2.3.9.	ESD - Electronic Systems Division 1975-78	51
2.3.10.	Doty (1977)	51
2.3.11.	Aerospace (1977)	52
2.3.12.	Boeing (1977)	52
2.3.13.	Walston y Felix / IBM-FSD (1977)	52
2.3.14.	SLICE - System Life Cycle Estimation (1977)	53
2.3.15.	Modelo de Putnam (1978)	53
2.3.16.	Modelo COCOMO - COConstructive COst MOdel(1979)	54
2.3.17.	Modelo PRICE S - RCA (1979)	57
2.3.18.	FPA - Function Point Analysis (1979)	58
2.3.19.	Putnam SLIM (1979)	60
2.3.20.	Modelo Jensen (1979)	62
2.3.21.	Sech-Square (Chi-Cuadrado) / Parr(1980)	63
2.3.22.	Bailey-Basili (1981)	63
2.3.23.	SOFTCOST - Deep Space Network (DSN)- (1981)	64
2.3.24.	COPMO (1983)	64
2.3.25.	ESTIMACS (1983)	64
2.3.26.	SEER-SEM (1989)	65
2.3.27.	COCOMO II (1995)	65
2.3.28.	SPR Checkpoint (1996)	69
2.3.29.	COBRA (1998)	70
2.3.30.	SELECT (1998)	70
2.4.	Clustering	71
2.4.1.	Clustering Jerárquico (<i>Hierarchical Clustering</i>)	73
2.4.2.	Métodos de Particionado y Recolocación (<i>Partitioning Relocation Clustering</i>)	75
2.4.3.	Algoritmos basados en Rejillas (<i>Grid-Based Methods</i>)	79
2.4.4.	Métodos basados en la Co-Ocurrencia de Datos Categóricos (<i>Co-Ocurrence of Categorical Data</i>)	80
2.4.5.	Clustering basado e Restricciones (<i>Constraint-Based Clustering</i>) .	82
2.4.6.	Algoritmos para Datos de Grandes Dimensiones	82
2.5.	Resumen	83

3. Planteamiento del problema	87
3.1. Introducción	87
3.2. Heterogeneidad de los proyectos de software	90
3.2.1. Heteroscedasticity	93
3.3. Influencia de los cost drivers sobre la estimación del coste del software . .	94
3.4. Resumen	95
4. Descripción del modelo paramétrico segmentado	97
4.1. Introducción	97
4.2. Descripción general del modelo	98
4.2.1. Segmentación de proyectos	99
4.2.2. Modelización paramétrica de los segmentos obtenidos	100
4.3. Definición formal del modelo	101
4.3.1. Modelo general	101
4.3.2. Modelo específico	104
4.4. Resumen	109
5. Evaluación	111
5.1. Metodología general llevada a cabo para realizar la evaluación del modelo segmentado	111
5.2. Base de datos de proyectos	112
5.3. Análisis de regresión	117
5.4. Estudio del caso base	120
5.4.1. Precisión de los modelos obtenidos	121
5.5. Selección de los algoritmos de clustering utilizados	122
5.5.1. K-medias	123
5.5.2. EM	124
5.5.3. Proceso de clusterización llevado a cabo por la herramienta RCT	128
5.6. Criterios para la evaluación del modelo paramétrico segmentado	133
5.6.1. Evaluación del proceso de clustering	134
5.6.2. Evaluación de las curvas de regresión	134
5.6.3. Evaluación de las estimaciones llevadas a cabo por el modelo sobre valores de test	136
5.6.4. Comparación de las evaluaciones realizadas	140
5.6.5. Predicción del esfuerzo que costará llevar a cabo un nuevo proyecto de software	141
5.7. Ejecución y análisis de resultados	143
5.7.1. Comparación de los algoritmos k-medias y EM, con el modelo básico	143
5.7.2. Experimentación con las diferentes formas de regresión de RCT .	150
5.7.3. Análisis recursivo mediante RCT	157
5.7.4. Particionado basado en el conocimiento de expertos y posterior clusterización automática de proyectos	165
5.8. Resumen	181

6. Conclusiones y trabajo futuro	183
6.1. Resumen de objetivos y de su evaluación	183
6.1.1. Resumen del objetivo general	184
6.1.2. Objetivos específicos y su evaluación	184
6.2. Conclusiones	185
6.3. Aportaciones originales	187
6.4. Líneas de trabajo futuras	189
6.4.1. Extensiones a los experimentos realizados	189
6.4.2. Introducción de nuevas características	191
6.4.3. Otras líneas de trabajo futuras	192
A. Recursive Clustering Tool - RCT	195
A.1. Descripción de la interface de usuario	195
A.2. Sesión de trabajo con RCT	199
A.3. Generalización del proceso	208
A.3.1. Parámetros de clusterización	208
A.3.2. Parámetros de la curva de regresión	209
A.3.3. Misceláneas	210
Bibliografía	213

Índice de tablas

2.1. Modos de proyecto COCOMO	55
2.2. Coeficientes de COCOMO	55
2.3. Cost drivers de las formas Intermedia y Detallada de COCOMO	56
5.1. Resultados de la aplicación de todos los modelos de regresión sobre 1246 proyectos de entrenamiento, y utilización de 300 de test	120
5.2. Resultados del algoritmo k-medias	145
5.3. Resultados del algoritmo EM	148
5.4. Análisis de regresión, mediante FuReA, para el modelo paramétrico global	148
5.5. Comparación k-medias, EM y modelo paramétrico global	149
5.6. Clusters obtenidos por RCT	151
5.7. Regresión lineal: <i>Linear (least-absolute-deviation)</i>	152
5.8. Regresión lineal: <i>Linear (least-squares)</i>	152
5.9. Regresión linealizada mediante doble transformación logarítmica: <i>Linear log-log Plot (least-absolute-deviation)</i>	153
5.10. Regresión linealizada mediante doble transformación logarítmica: <i>Linear log-log Plot (least-squares)</i>	153
5.11. Regresión exponencial: <i>Non-Linear Power (least-squares)</i>	154
5.12. Comparación de los valores obtenidos con las diferentes regresiones . . .	154
5.13. Clusters obtenidos a partir del cluster 6	155
5.14. Parámetros de las regresiones de los clusters recursivos	156
5.15. División recursiva del cluster 0	159
5.16. Características de los clusters procedentes del cluster 0	159

5.17. División recursiva del cluster 1	159
5.18. Características de los clusters procedentes del cluster 1	160
5.19. División recursiva del cluster 3	160
5.20. Características de los clusters procedentes del cluster 3	161
5.21. División recursiva del cluster 4	162
5.22. Características de los clusters procedentes del cluster 4	163
5.23. División recursiva del cluster 5	164
5.24. Características de los clusters procedentes del cluster 5	164
5.25. Comparación del proceso recursivo con el modelo no recursivo y con el total	165
5.26. Clusters obtenidos a partir del cluster#7	167
5.27. Características de los clusters procedentes de la partición cluster#7 . . .	169
5.28. Clusters obtenidos a partir del cluster#8	169
5.29. Características de los clusters procedentes de la partición cluster#8 . . .	170
5.30. Clusters obtenidos a partir del cluster#9	171
5.31. Características de los clusters procedentes de la partición cluster#9 . . .	171
5.32. Clusters obtenidos a partir del cluster#10	171
5.33. Características de los clusters procedentes de la partición cluster#10 . . .	172
5.34. Clusters obtenidos a partir del cluster#11	173
5.35. Características de los clusters procedentes de la partición cluster#11 . . .	173
5.36. Clusters obtenidos a partir del cluster#12	174
5.37. Características de los clusters procedentes de la partición cluster#12 . . .	175
5.38. Clusters obtenidos a partir del cluster#34	175
5.39. Características de los clusters procedentes de la partición cluster#34 . . .	176
5.40. Clusters obtenidos a partir del cluster#35	176
5.41. Características de los clusters procedentes de la partición cluster#35 . . .	177
5.42. Clusters obtenidos a partir del cluster#36	177
5.43. Características de los clusters procedentes de la partición cluster#36 . . .	178
5.44. Modelo particionado y posteriormente clusterizado	178

Índice de figuras

2.1. Clasificación de Modelos de Estimación según (Wieczorek & Briand 2001)	28
2.2. Modelo de Estimación mediante una Red Neuronal	34
2.3. a) Representación difusa del valor lingüístico joven. b) Representación clásica. c) Representación numérica	37
3.1. Curva de regresión que representa a todos los proyectos de la base de datos.	91
3.2. Conjunto de datos que presenta un patrón <i>heteroscedastic</i>	93
4.1. Descripción gráfica del modelo de estimación paramétrico segmentado . .	106
5.1. Ejemplo de regiones en el plano e,fp	138
5.2. Funciones de distribución de dos de los clusters obtenidos	139
5.3. Clusters obtenidos con el algoritmo k-medias	146
5.4. Clusters obtenidos con el algoritmo EM	147
5.5. Clusters obtenidos con la herramienta RCT	151
5.6. Segmentación recursiva del cluster 6	156
5.7. Clustering recursivo mediante RCT	158
5.8. Clustering recursivo del cluster 0	160
5.9. Clustering recursivo del cluster 1	161
5.10. Clustering recursivo del cluster 3	162
5.11. Clustering recursivo del cluster 4	163
5.12. Clustering recursivo del cluster 5	165
5.13. Particionado de la base de datos ISBSG según los atributos METHO y CASET	167

5.14. Jerarquía de clusters obtenidos tras el proceso de particionado y posterior clusterización	168
5.15. Segmentación del cluster#7, METHO yes, CASET yes	169
5.16. Segmentación del cluster#8, METHO yes, CASET no	170
5.17. Segmentación del cluster#9, METHO yes, CASET unknown	170
5.18. Segmentación del cluster#10, METHO no, CASET yes	172
5.19. Segmentación del cluster#11, METHO no, CASET no	173
5.20. Segmentación del cluster#12, METHO no, CASET unknown	174
5.21. Segmentación del cluster#34, METHO unknown, CASET yes	175
5.22. Segmentación del cluster#35, METHO unknown, CASET no	176
5.23. Segmentación del cluster#36, METHO unknown, CASET unknown	177
5.24. Distribución de los proyectos de test sobre los de entrenamiento	179
A.1. Pantalla inicial con la base de datos de proyectos cargada	200
A.2. Pantalla inicial tras la separación de proyectos de training y de test	201
A.3. Ventana de selección de los parámetros de la segmentación	202
A.4. Árbol de clusters obtenidos	203
A.5. Informe del proceso de clustering	204
A.6. Selección de los parámetros del análisis de regresión	205
A.7. Resultados del proceso de análisis	206
A.8. Resultados del proceso de test	207
A.9. Gráfico con los clusters obtenidos	208