

Comparación de modelos destinados a anticipar la conveniencia de trazar proyectos de desarrollo de software

Giró, Juan / Vazquez, Juan / Meloni, Brenda y Constable, Leticia
Universidad Tecnológica Nacional, Facultad Regional Córdoba,
Departamento de Ingeniería en Sistemas de Información

Abstract

La escasez de evidencias de que los progresos en el campo de la trazabilidad son efectivamente aprovechados por la industria del software ha estimulado el desarrollo de modelos conducentes a un mejor conocimiento del problema y poder anticipar los resultados esperables en proyectos. Para ello fue necesario identificar los factores de mayor impacto en el éxito de los procesos de trazabilidad y proponer modelos que permitan hacer predicciones a partir de esos factores. En este trabajo se hace un análisis comparativo de tres modelos de predicción, se utiliza un caso de estudio y se discuten los resultados obtenidos. Para la comparación se recurre al Análisis ROC con el fin de identificar al modelo más conveniente. A pesar de las ventajas y creciente divulgación del Análisis ROC como herramienta de evaluación de modelos dicotómicos, por el momento esta técnica ha tenido poca difusión en la ingeniería de software.

Palabras Clave

Ingeniería de Software, Trazabilidad de Requerimientos, Modelos de Predicción de Trazabilidad, Análisis ROC.

Introducción

Las evidencias de que los progresos en el campo de la trazabilidad de requerimientos de proyectos de desarrollo de software no llegan a ser efectivamente aplicadas en la industria [1] condujeron a la necesidad de entender mejor el problema y sus causas. Revisando las experiencias desfavorables con referencia a la trazabilidad se comprobó que pueden ser reunidas en tres grupos: *i)* los proyectos en que ésta no fue iniciada, fue prematuramente abandonada o no cubrió las expectativas desde un punto de vista técnico, *ii)* aquellos en que fue técnicamente exitosa con un costo de implementación mayor que el beneficio obtenido y *iii)* los que con un elevado costo condujeron a un resultado pobre o nulo, es decir una combinación de las dos primeros.

Lo expuesto resulta sorprendente ya que en la actualidad es unánime el reconocimiento de la trascendencia e importancia de la trazabilidad de requerimientos como soporte de los procesos de desarrollo de software [2][3], habiendo sido incorporada en todas las normas y modelos de desarrollo vigentes.

Cabe aquí acotar que se consideran proyectos exitosos o con resultados favorables aquellos en que la trazabilidad se mantuvo vigente a lo largo de todo el proceso de desarrollo y sus resultados cubrieron las expectativas de las organizaciones en lo que hace a los beneficios obtenidos y los recursos demandados. Está fuera del alcance de este trabajo definir indicadores que permitan valorar, calificar y establecer comparaciones con referencia al impacto de la trazabilidad en el propio proyecto o en la calidad del software producido, los que serán objetivos de próximas etapas.

Antes de continuar es necesario enfatizar que al hablarse de trazabilidad de requerimientos en proyectos de desarrollo de software se está haciendo referencia a la gestión que vincula las numerosas etapas de sus ciclos de vida, asegurando el éxito del proyecto, brindando la necesaria garantía de coherencia, completitud y corrección al software producido y posibilitando su eficaz mantenimiento correctivo y preventivo en el resto de su vida útil.

Al analizarse las líneas de estudio en el campo de la trazabilidad, se comprueba que en su mayor parte están orientadas a desarrollar nuevas metodologías y herramientas, habiendo un esfuerzo mucho menor destinado a estudiar el resultado de la aplicación de las mismas en la industria y

las causas de las dificultades ya señaladas. Además, los escasos documentos destinados a analizar el origen de las dificultades de la trazabilidad abordan el problema en forma cualitativa y en la mayoría de las veces el enfoque es demasiado general.

Surgió así la presunción de que no es fortuito que ciertos proyectos puedan ser exitosamente trazados y otros no, por lo que debe haber una combinación de condiciones objetivas que conducen a que se llegue a uno u otro resultado. Esta idea llevó a plantear la hipótesis de que *deben existir factores que condicionan el éxito de los procesos de trazabilidad y que es factible identificarlos*.

La comprobación de la hipótesis enunciada orienta la actividad que se viene cumpliendo en el proyecto “Aseguramiento de la Trazabilidad en Proyectos de Desarrollo de Sistemas de Software” [4], y en este marco se propusieron factores y modelos destinados a anticipar los resultados de procesos de trazabilidad, que vienen siendo progresivamente mejorados [5][6][7][8].

En este trabajo se comparan tres modelos de predicción de trazabilidad que se apoyan en los mismos factores y métricas. Uno de ellos ya fue presentado con anterioridad [7] y los otros dos tienen su origen en un modelo que originalmente solo estaba destinado a una interpretación gráfica del problema [5]. Por tratarse de modelos destinados a resolver un problema dicotómico, se recurre al Análisis ROC como herramienta de comparación a fin de identificar la opción más conveniente. La organización del documento es la siguiente:

en la sección 2 se resumen las características de los modelos estudiados, en la sección 3 se presentan los resultados obtenidos con un caso de estudio, en la sección 4 se hace un análisis comparativo de los mismos y finalmente en la sección 5 se presentan las conclusiones de este trabajo y actividades futuras.

Características de los modelos estudiados

Se reconocen tres entidades principales que están estrechamente relacionadas entre sí, que son: a) el propio *producto software*, b) el *proyecto*, que responde a cierto modelo de proceso y ampara la construcción del producto y c) la *organización*, que constituye el escenario en el que el proyecto es desarrollado. Por lo tanto, se anticipa que los factores buscados estarán asociados a dimensiones de estas tres entidades.

También se establecieron criterios con respecto a la selección de los factores, la forma en que éstos son evaluados y su interpretación con respecto al problema tratado, que son los siguientes: a) deben ser *cuantificables*, b) se les aplicará un multiplicador de escala para expresarlos en el *intervalo cero a cinco*, c) los valores *crecientes* contribuyen más favorablemente a la trazabilidad de un proyecto y d) deben ser *ortogonales entre sí*. Se definen así, agrupados por entidades, ocho factores propuestos [6][7][8] para predecir la conveniencia de realizar la trazabilidad de requerimientos en proyectos de desarrollo de software, que son resumidos a continuación en las Tabla 1 y Tabla 2:

Tabla 1: Definición de entidades, factores, variables, variables de referencia, métricas e intervalos

Entidad	Factor	Variable	Variable de referencia V_r y descripción	Métrica $f(V_r)$ (Variante B)	Intervalo de V_r
Producto	Tamaño	t	Tamaño del producto en Puntos de Función PF	$t = 5 * PF / 1000$	100 - 1000
	Vigencia	v	Expectativa de vida útil VU [años]	$v = 5 * VU / 10$	0,5 - 10
	Reutilización	r	Previsión de futura reutilización RE [%]	$r = 1 + RE/20$	0 - 80
	Confiabilidad	c	Indicador de la confiabilidad exigida CO (*)	$c = 1 + 0,8 * CO$	0 - 5
Proyecto	Plazo	p	Previsión de la duración del proyecto DP [años]	$p = DP$	0 - 5
	Equipo	e	Efectividad y capacidad del equipo EF (**)	$e = EF$	1 - 5
Organización	Madurez	m	Nivel de madurez $CMMI$ de la organización	$m = CMMI$	1 - 5
	Dependencia	d	Autonomía NA (***) en adopción de normas	$d = NA$	1 - 5

Tablas 2: Definición de los indicadores CO (*), EF (**), Nivel de Madurez CMMI y NA (***)

Confiabilidad CO		Efectividad EF		Madurez CMMI		Autonomía NA	
0	No importante	1	Pobre	1	Inicial	1	Independiente
1	Baja	2	Baja	2	Gestionado	2	Normas propias
2	Media	3	Media	3	Definido	3	Normas clientes
3	Alta	4	Alta	4	Cuant.Gestion.	4	Nor. casa matriz
4	Muy alta	5	Muy alta	5	Optimizado	5	Combina 3 y 4
5	Absoluta						

En las referencias [5][6][7][8] ya señaladas se presenta un detalle del alcance y justificación de los factores y métricas utilizados: *i*) en [5] se presentan los primeros modelos propuestos y se justifican las entidades y factores seleccionados, *ii*) en [6] y [7] se mejoran los modelos originales brindando detalles que justifican las variantes en los factores adoptados y *iii*) en [8] se analizan diferentes métricas para seleccionar a través de una evaluación exhaustiva aquellas que conducen a una clasificación más efectiva, que son las identificadas como “variante B” en la Tabla 1.

A partir de los ocho factores enunciados, se desarrollan a continuación los siguientes tres modelos: *i*) Vectorial “V”, *ii*) de Áreas “A” y *iii*) de Radio Medio “R”.

Modelo Vectorial “V”

Con el fin de reducir la dimensión del problema y facilitar la visualización de las poblaciones de datos se propuso la asignación de una variable a cada una de las tres entidades involucradas: el *producto* (η_1), el *proyecto* (η_2) y la *organización* (η_3). Para calcular el módulo resultante de cada una de las tres variables se tuvo en cuenta la ortogonalidad entre los factores:

$$\eta_1 = \sqrt{(t^2 + v^2 + r^2 + c^2)} \quad (1)$$

$$\eta_2 = \sqrt{(p^2 + e^2)} \quad (2)$$

$$\eta_3 = \sqrt{(m^2 + d^2)} \quad (3)$$

Luego, a partir de esta reducción de dimensiones surgió la idea de utilizar el módulo de la resultante de los ocho factores, en adelante denominado “ ρ ”, como parámetro representativo o “indicador” de cada caso considerado:

$$\rho = \sqrt{(\eta_1^2 + \eta_2^2 + \eta_3^2)} = \sqrt{(t^2 + v^2 + r^2 + c^2 + p^2 + e^2 + m^2 + d^2)} \quad (4)$$

Es importante reconocer el significado del indicador ρ , que representa el radio de una fracción de casquete esférico en el espacio de tres dimensiones, o en el hiperespacio de ocho, que separa los casos exitosamente trazables de los que no lo son. Es decir que es el radio de una hipersuperficie espacial que discrimina ambas poblaciones. Por lo tanto, el objetivo es aquí determinar el valor más apropiado de ρ , denominado ρ_c , que permita separar de la mejor forma las poblaciones de los proyectos no trazables de los trazables.

Modelo de Áreas “A”

La siguiente propuesta se apoya en un primer modelo [5][6], denominado básico, que equivale a reconocer que cada uno de los ocho factores incide de manera independiente sobre el diagnóstico buscado, permitiendo un rápido reconocimiento de las expectativas que ofrece la trazabilidad de un proyecto.

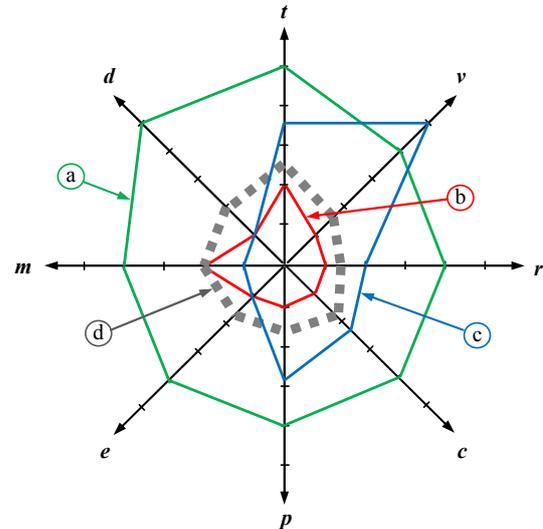


Figura 1. Representación de la trazabilidad de proyectos en un diagrama de radar

Para visualizar el modelo se recurre a un “diagrama radar”, que es un esquema plano con una cantidad de ejes radiales igual a la dimensión del modelo, en este caso ocho. En la Figura 1 se muestra uno de tales diagramas, donde los polígonos corresponden a: *a*) un proyecto trazable, *b*) un proyecto no trazable, *c*) un proyecto atípico y *d*) la “zona gris” que representa la frontera discriminante entre ambos casos.

Las posibilidades de clasificación que ofrece este modelo son muy limitadas, pero fueron la base del ya anticipado “Modelo de Áreas *A*”, en el que se adopta como indicador de cada proyecto el área encerrada por su polígono octogonal. Al haber ocho ejes y por lo tanto ocho triángulos con un ángulo interior que es siempre de 45°, se suman los productos de cada uno de los ocho catetos “*k*” (b_k) por las correspondientes alturas ($h_k = c_k \cdot \sin 45^\circ$) y se obtiene la Ecuación (5):

$$A = \frac{1}{2} \sum_{k=1}^8 b_k \cdot h_k = \frac{1}{2} \sum_{k=1}^8 b_k \cdot c_k \cdot \sin(45) \\ = \frac{\sqrt{2}}{4} \sum_{k=1}^8 b_k \cdot c_k = 0.25\sqrt{2}(tv + vr + rc + cp + pe + em + md + dt) \quad (5)$$

Este nuevo modelo ya no considera en forma independiente la incidencia de cada factor en el diagnóstico, porque ahora el indicador de “corte” o de clasificación pasa a ser el área encerrada por el polígono “*d*” de la Figura 1, que resulta de la contribución de todos los factores. Además, con este enfoque desaparecen los “proyectos atípicos”, que eran los que cruzaban la frontera del polígono discriminante. Esto significa que en ciertos proyectos las debilidades en algunos factores se ven compensadas por las fortalezas en otros, permitiendo que sean trazados exitosamente.

Modelo de Radio Medio “*R*”

El Modelo de Áreas representa un enorme progreso con respecto a la mera representación de los polígonos en el gráfico de radar, posibilitando la disponibilidad de un indicador cuantificable. No obstante, podría

objetarse el hecho de que la condición de un proyecto sea representada por un área, lo que no parece ser muy intuitivo. Para superar esta dificultad se buscó la equivalencia con un polígono que encierre la misma área, pero con un idéntico valor para todos sus radios, lo que conduce al indicador denominado “Radio Medio” *R* que se define con la Ecuación (6) a partir de la Ecuación (5):

$$A = \frac{\sqrt{2}}{4} \sum_{k=1}^8 b_k \cdot c_k = \frac{\sqrt{2}}{4} \sum_{k=1}^8 R^2 = 2\sqrt{2}R^2 \\ \text{de donde resulta } R = \sqrt{\frac{0.5A}{\sqrt{2}}} \quad (6)$$

Análisis ROC

Una vez definidos los tres modelos a ser considerados, el paso siguiente es definir la herramienta de comparación a efectos de identificar el más conveniente. Para este caso es recomendable el análisis ROC [9], que es una técnica destinada a evaluar clasificadores dicotómicos y que recientemente ha experimentado gran difusión en campos muy variados, tales como la bioingeniería, aprendizaje automático y minería de datos. A pesar de ello, todavía es muy poco usado en la ingeniería de software y esto justifica una breve presentación. El Análisis ROC persigue dos objetivos principales: *i*) poder elegir objetivamente el mejor entre varios modelos de clasificación y *ii*) optimizar el valor de corte más conveniente en el modelo elegido. En este caso se aplica el Análisis ROC con el primer objetivo, que es el de comparar clasificadores.

Tabla 3: Presentación de los resultados posibles en una tabla de contingencia

Modelo Clasificador		Condición real	
		Positivos	Negativos
Resultados con cierto valor de corte	Positivos	Verdaderos positivos (VP)	Falsos positivos (FP)
	Negativos	Falsos Negativos (FN)	Verdaderos negativos (VN)

Si se considera poblaciones de datos con resultados positivos y negativos, y que

tienen como referencia un determinado indicador, al definirse un cierto valor de corte quedan inmediatamente establecidos cuatro agrupamiento de los datos, que son los representados en la Tabla 3.

Estos agrupamientos dan lugar a la definición de ciertos parámetros que son la base del análisis ROC, según se definen a continuación.

$$\text{Sensibilidad} = FVP = \frac{VP}{VP+FN} \quad (7)$$

$$\text{Especificidad} = FVN = \frac{VN}{VN+FP} \quad (8)$$

$$\text{Especificidad} = FVN = 1 - FFP \quad (9)$$

$$\text{Exactitud} = \frac{VP+VN}{VP+FN+VN+FP} \quad (10)$$

Las curvas ROC representan la sensibilidad (Ec. 7) en función de FFP (Ec. 9) y el mejor modelo es el que encierra la mayor área bajo esta curva, denominada AUC [10][11].

A partir de estos parámetros la curva ROC representa globalmente el desempeño de un modelo clasificador, es necesariamente creciente y refleja el compromiso existente entre *sensibilidad* y *especificidad*. En efecto, si para obtener mayor *sensibilidad* se modifica el valor de corte entre las poblaciones de las muestras negativas y positivas, se lo hace a expensas de disminuir al mismo tiempo la *especificidad*. Esto obliga a seleccionar dicho valor de corte con mucho cuidado y presupone la elección apropiada del mejor indicador que representa el problema estudiado, dependiendo de esta elección la menor superposición posible de ambas poblaciones.

Cabe aquí destacar que las curvas ROC presentan la ventajosa propiedad de que son insensibles a los cambios en la proporción de instancias positivas y negativas que pueda haber en un lote de datos. Por su parte, el área bajo la curva ROC (AUC) exhibe también importantes propiedades estadísticas.

Resultados obtenidos con caso de estudio

El “caso de estudio” utilizado [7] consiste

en un lote de 102 muestras, que incluyen 55 proyectos trazados exitosamente (54%) y 47 proyectos con resultados negativos (46%), representados en lo sucesivo en las figuras por líneas verdes y rojas respectivamente. Con estos datos se establece la comparación entre los tres modelos presentados, para lo cual se consideraron las áreas encerradas bajo la curva ROC (AUC) y las exactitudes (Ecuación 10) obtenidas con cada caso.

Resultados de Modelo Vectorial “V”

Se toma como indicador o variable característica el módulo ρ del vector que define la posición de cada caso en el hiperespacio de ocho dimensiones (Ecuación 4), y al representarse la población de los 102 casos en función de ρ se obtiene la curva de distribución mostrada en la Figura 2.

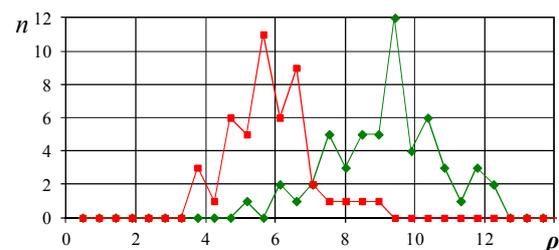


Figura 2: Curvas de distribución de experiencias de trazabilidad negativas y positivas en función de ρ

Como puede observarse, los casos negativos están distribuidos en el intervalo $3,7 < \rho < 9,0$, lo que representa el 37,4 % del intervalo de valores posibles (0 - 14,14). Por su parte, los casos positivos están distribuidos entre $5,1 < \rho < 12,3$, que representa el 50,9% del intervalo posible.

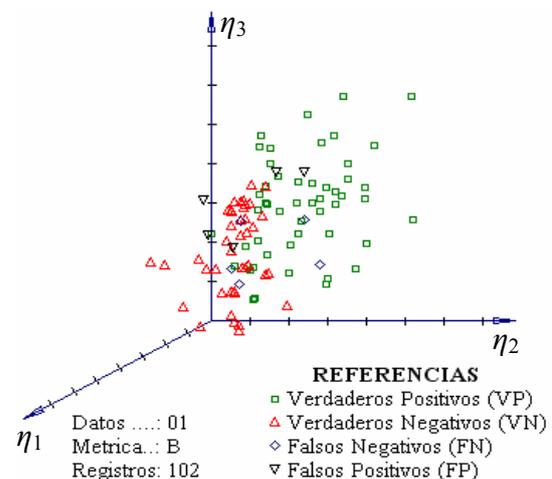


Figura 3: Representación de los 102 proyectos en un sistema de coordenadas ortogonales η_1, η_2, η_3

En la Figura 3 se muestra en una representación 3D la población de los proyectos considerados, sobre los ejes ortogonales η_1 , η_2 , η_3 . Aquí los proyectos han sido clasificados de acuerdo a lo previsto en la Tabla 3 considerando un cierto valor de corte ρ_c .

A continuación, en las Figuras 4.a y 4.b, se representa con esos mismos datos la curva de exactitud (Ecuación 10) y la Curva ROC.

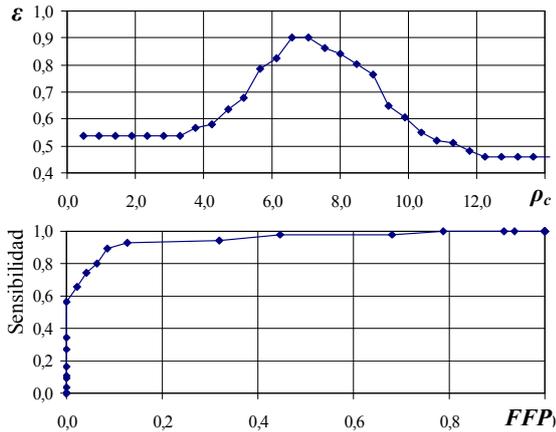


Figura 4: a) Representación de la exactitud ε en función de ρ_c y b) Curva ROC, ambos correspondientes a los resultados del Modelo "V"

En la Tabla 4 se reúnen los resultados que representan el desempeño del modelo "V": los intervalos de la curva de distribución, la máxima exactitud alcanzada (ε_{\max}), el valor de corte ($\delta = \rho_c / \rho$) en que ésta exactitud es alcanzada y el área bajo la curva ROC (AUC).

Tabla 4: Resumen de resultados del Modelo "V"

Modelo V	
I^- [%]	37,4
I^+ [%]	50,9
I_{\pm} [%]	27,6
ε_{\max} [%]	90,20
δ [%]	46,67 – 50,00
AUC	0,9501

Los intervalos de la curva de distribución que interesan son los ocupados por casos negativos (I^-), por casos positivos (I^+) y el intervalo en que se presentan ambos (I_{\pm}).

La definición de δ obedece a la necesidad de disponer de una referencia adimensional, que permita comparar las posiciones de los valores de corte en tres modelos cuyos

indicadores o variables características son de distinta naturaleza.

Resultados del Modelo de Áreas "A"

En este modelo se toma como indicador o variable característica al área A que encierra cada octógono representado sobre el gráfico de radar (Ec. 5). En la Figura 5 se presenta la curva de distribución de los 102 casos en función de A . En esta representación los casos negativos están presentes en el intervalo $2 < A < 28$, que corresponde a un valor de $I^- = 36,7\%$ y los casos positivos están en el intervalo $9 < A < 52$, con $I^+ = 60\%$. El intervalo con ambos resultados se presenta en $9 < A < 28$, por lo que $I_{\pm} = 26,7\%$

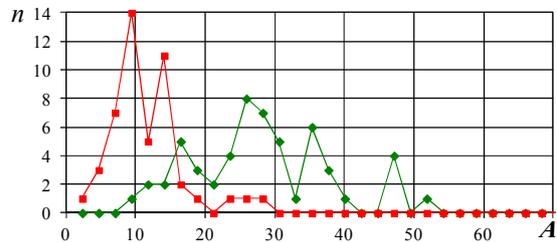


Figura 5: Curvas de distribución de experiencias de trazabilidad negativas y positivas en función de A

En la Figura 6 se muestran todos los proyectos considerados sobre un mismo diagrama de radar. Como se observa, la representación se aleja considerablemente de la imagen idealizada presentada en la Figura 1 y da testimonio de la complejidad del problema, justificándose la adopción del indicador "A" que permita su tratamiento numérico.

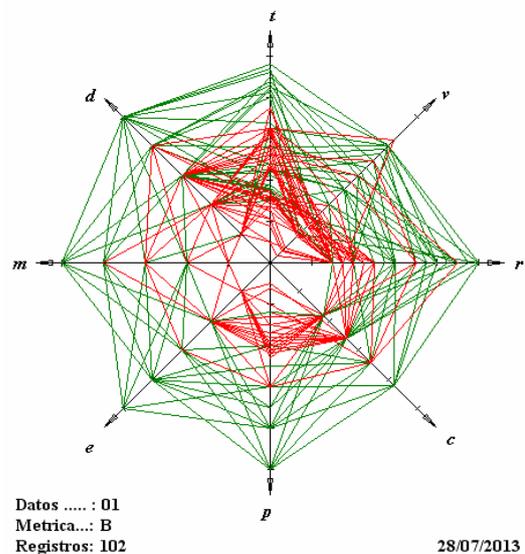


Figura 6: Representación de los 102 proyectos sobre los ocho ejes del diagrama de radar.

En las Figuras 7.a y 7.b mostradas a continuación se representan para el modelo “A” la curva de exactitud y la curva ROC respectivamente.

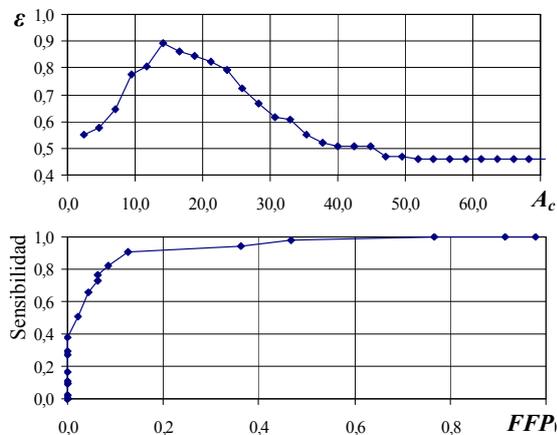


Figura 7: a) Representación de la exactitud ϵ en función del área A_c y b) Curva ROC, ambos correspondientes a los resultados del Modelo “A”

Por último, al igual que con el modelo anterior, en la Tabla 5 se resumen los resultados obtenidos.

Tabla 5: Resumen de resultados del Modelo “A”

Modelo A	
I^- [%]	36,7
I^+ [%]	60,0
I_{\pm} [%]	26,7
ϵ_{\max} [%]	89,22
δ [%]	20,00
AUC	0,9172

Resultados del Modelo de Radio Medio “R”

La complejidad del diagrama de radar de la Figura 6 llevó a la idea de algún tipo de normalización, que aunque distorsione la representación con respecto al significado de sus ejes, permita visualizar los casos y establecer un orden entre ellos. Después de estudiar algunas variantes se adoptó lo que se denominó el “Radio Medio” R (Ecuación 6), que surge de representar cada proyecto por un octógono que tenga la misma área del polígono del gráfico radar, pero con todos sus lados iguales.

A partir de esta idea se estableció una nueva curva de distribución, que es mostrada en la Figura 8, y los octógonos de los proyectos considerados son mostrados en la

Figura 9. Debe observarse que en esta representación los ejes han perdido su identidad. También puede observarse una clara correspondencia entre la ubicación de ambos tipos de proyectos en las dos representaciones.

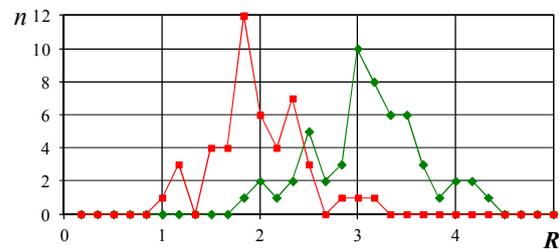
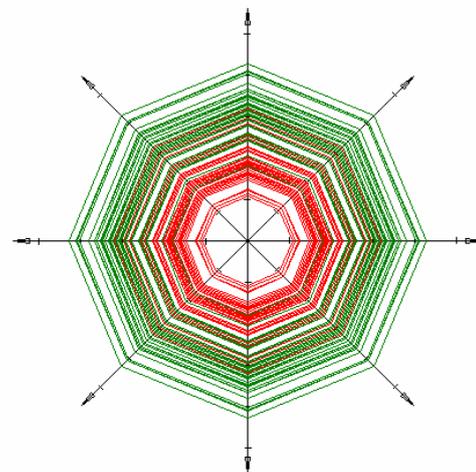


Figura 8: Curvas de distribución de experiencias de trazabilidad negativas y positivas en función de R



Datos: 01
Métrica....: B
Registros: 102

28/07/2013

Figura 9: Representación de los 102 proyectos mediante octógonos concéntricos.

En efecto, los proyectos negativos están a la izquierda en la curva de distribución y hacia el centro en el diagrama de octógonos, y los positivos en sentido contrario. Aún sin una definición del valor de corte, en este caso R_c , se perciben claramente los proyectos con tendencia a quedar encuadrados como falsos positivos o falsos negativos.

Volviendo a la curva de distribución, los casos negativos están presentes en el intervalo $1 < R < 3,16$, que corresponde a un valor de $I^- = 43,2\%$ y los casos positivos están en el intervalo $1,8 < R < 4,4$, con $I^+ = 52\%$. El intervalo con ambos resultados se presenta en $1,8 < R < 3,16$, que implica un valor $I_{\pm} = 27\%$.

Para este modelo “R” se representan en la Figura 10.a la curva de exactitud, en la Figura 10.b la curva ROC y en la Tabla 6 un resumen de los resultados obtenidos.

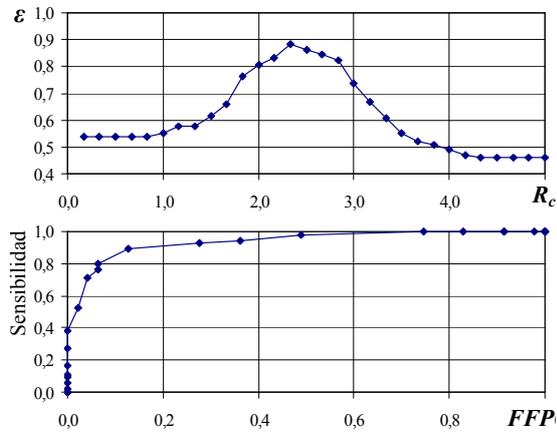


Figura 10: a) Representación de la exactitud ϵ en función de R_c y b) Curva ROC, ambos correspondientes a los resultados del Modelo “R”

Tabla 6: Resumen de resultados del Modelo “R”

Modelo R	
I- [%]	43,2
I+ [%]	52,0
I± [%]	27,2
ϵ_{\max} [%]	88,24
δ [%]	46,66
AUC	0,9375

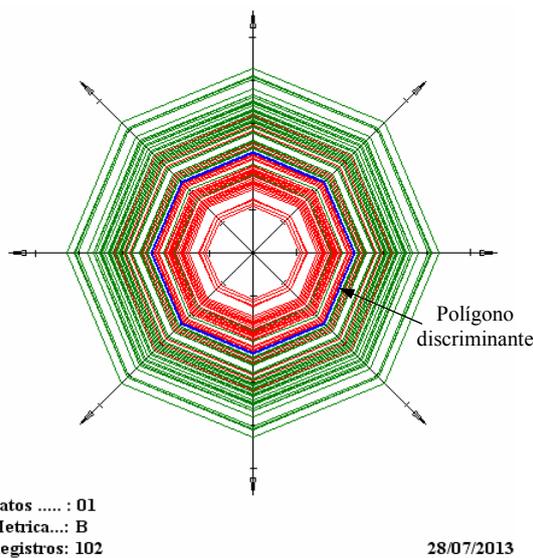


Figura 11: Representación de los 102 proyectos mediante octógonos concéntricos y el polígono discriminante que separa ambas poblaciones.

A título ilustrativo se muestra nuevamente

en la Figura 11 el diagrama de octógonos, esta vez con el polígono discriminante que corresponde a ϵ_{\max} que es representado en color azul. Los polígonos rojos por afuera del polígono discriminante son los casos falsos positivos y los polígonos verdes por dentro del polígono discriminante son los falsos negativos.

Análisis comparativo de los resultados

Como ya fue anticipado, para la comparación de los tres modelos se acude a los resultados de exactitud (Ecuación 10) y los del área bajo la curva ROC. Ésta última conduce a la determinación del *coeficiente de Gini (G)*, que es una medida de dispersión estadística que fue propuesta por el italiano Corrado Gini y se aplica al estudio de desigualdades en diversos campos, tales como la sociología, economía e ingeniería, entre otros. La relación de *G* con *AUC* es directa, tal como se muestra en Ecuación 11:

$$G = 2.AUC - 1, \quad 0 \leq G \leq 1 \quad (11)$$

Por último, se recurre al índice de Youden (γ) que es tradicionalmente utilizado para comparar herramientas de diagnóstico [12], considerando conjuntamente a la sensibilidad y especificidad de un modelo. Tal como está definido por la Ecuación 12, γ establece el mejor desempeño posible de un clasificador y el punto de corte correspondiente.

$$\gamma = |\text{sensibilidad} + \text{especificidad} - 1|_{\max} \quad (12)$$

Obsérvese que $0 \leq \gamma \leq 1$, donde mayores valores de γ están asociados a mejores desempeños de un modelo, y $\gamma = 1$ representa un clasificador perfecto. Por el contrario, un modelo incapaz de separar las dos poblaciones es representado por $\gamma = 0$.

Los resultados obtenidos, presentados con anterioridad, fueron reordenados para representar en la Figura 12 la sensibilidad, especificidad e índice de Youden de los tres modelos en función de δ . En este gráfico la curva azul (rombos) corresponde a la evolución de la sensibilidad, la curva marrón (cuadrados) a la evolución de la especificidad y los valores de la curva violeta (triángulos) son el índice de Youden y se determinan con la Ecuación 12.

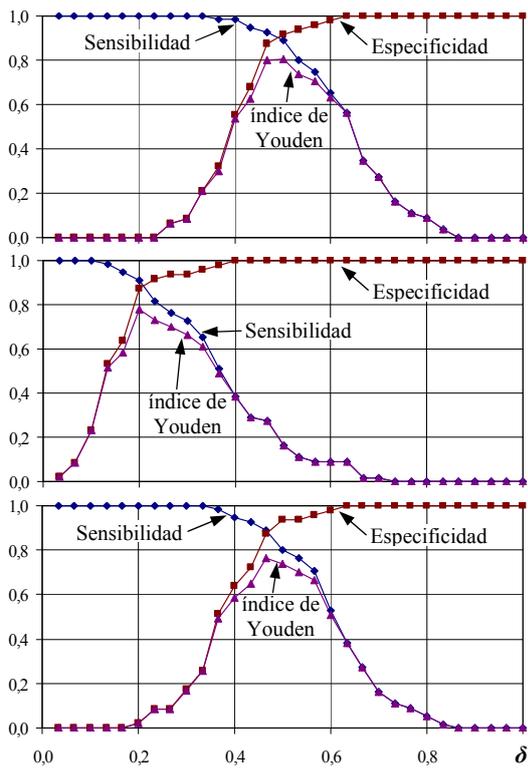


Figura 12: Representación de la sensibilidad (*se*), especificidad (*es*) e índice de Youden (γ) de los tres modelos en función de δ , que define los puntos de corte en el intervalo $0 \leq \delta \leq 1$

Se presenta a continuación la Tabla 7 con el fin de facilitar el análisis comparativo de los modelos estudiados.

Tabla 7: Resumen de los resultados de los tres modelos clasificadores estudiados: “V”, “A” y “M”

Indicadores evaluados	Modelos comparados		
	V	A	R
I^- [%]	37,4	36,7	43,2
I^+ [%]	50,9	60,0	52,0
I_{\pm} [%]	27,6	26,7	27,2
ϵ_{\max} [%]	90,20	89,22	88,24
δ [%]	46,6– 50,0	20,00	46,66
<i>AUC</i>	0,9501	0,9172	0,9375
<i>G</i>	0,9002	0,8344	0,8750
γ	0,8058	0,7814	0,7632
δ_{γ} [%]	50,0	20,00	46,66

Se trata de un resumen unificado que incluye: *i*) los intervalos que ocupan las muestras en la curva de distribución (I^- , I^+ , I_{\pm}), *ii*) la máxima exactitud (ϵ_{\max}) que brinda cada modelo, *iii*) la posición adimensional del valor de corte (δ) en que

la máxima exactitud es alcanzada, *iv*) el área bajo la curva *ROC* (*AUC*), *v*) el coeficiente de Gini (*G*), *vi*) el mejor índice de Youden (γ) que brinda cada modelo y *vii*) el valor de corte que lo posibilita (δ_{γ}).

El análisis de los resultados obtenidos con los tres modelos sobre el caso de estudio permite hacer las siguientes consideraciones:

- 1) Las curvas de distribución tienen apariencias claramente distintas. Sin embargo, en los tres modelos los datos de casos positivos y negativos ocupan espacios similares (I^- e I^+) en relación a la magnitud máxima de cada indicador utilizado.
- 2) Con la misma referencia, la zona en que se superponen ambas poblaciones (I_{\pm}) es prácticamente la misma en los tres modelos. Esta gran similitud deja la sospecha de que es consecuencia de los propios datos y en menor medida del modelo de clasificación, tema que será objeto de mayor estudio.
- 3) Las máximas exactitudes (ϵ_{\max}) alcanzadas por los tres modelos son similares, con la mayor correspondiendo al modelo “V” y la más baja al “R”
- 4) Sin embargo, los valores de corte que dan lugar a las exactitudes máximas son similares en los modelos “V” y “R”, pero muy diferentes al del modelo “A”.
- 5) Considerando el área encerrada bajo la curva *ROC* (*AUC*), el mejor modelo es el “V” y el de peor desempeño el “A”.
- 6) Como es de esperar, el coeficiente de Gini permite llegar a las mismas conclusiones que con el *AUC*, con la ventaja de hacerlo con un indicador ampliamente difundido.
- 7) El índice de Youden permite llegar a las mismas conclusiones que con la máxima exactitud: el mejor clasificador es el modelo “V” y el de peor desempeño el “R”.
- 8) Los mejores desempeños considerando la máxima exactitud y el índice de Youden se presentan, en cada modelo, con los mismos valores de corte.
- 9) El mejor modelo para el caso de estudio considerado es el vectorial “V”, tanto por

brindar el punto de mayor exactitud como el mejor desempeño global (AUC).

10) Si bien no tan buenos como el “ V ”, los modelos “ A ” y “ R ” permiten una utilización efectiva de los gráficos de radar, que de otra manera hubiesen quedado relegados a una mera representación conceptual del problema, pero de escaso valor práctico.

11) Las máximas diferencias porcentuales que se presentan entre los tres clasificadores muestran los siguientes valores:

$$\begin{aligned}\Delta \varepsilon_{\max} / \varepsilon_{\max} &= 2,2 \% \\ \Delta AUC / AUC_{\max} &= 3,5 \% \\ \Delta \gamma_{\max} / \gamma_{\max} &= 5,3 \%\end{aligned}\quad (13)$$

12) A raíz de lo comentado en 8 referido a los mejores desempeños de ε_{\max} y γ , se analizó la relación $\varepsilon_{\max} / \gamma$ obteniéndose para los tres modelos los valores 1,12 – 1,14 y 1,16, que muestra un muy leve incremento a medida que el desempeño es más pobre.

El análisis realizado permite concluir que los tres modelos de clasificación evaluados tienen un buen comportamiento, siendo el “ V ” el que supera a los demás según todos los parámetros de evaluación considerados. También se comprobó que las características específicas de cada modelo no tienen un impacto significativo en los correspondientes desempeños.

Conclusiones y trabajo futuro

Al procurar entender las causas que obstaculizan la aplicación efectiva de la trazabilidad en la industria del software se ingresó en un mundo complejo y apasionante, que en cierta medida responde a la estrictez de las matemáticas y al mismo tiempo es impactado por las conductas, muchas veces ambiguas, de los seres humanos. En ese contexto se identificaron ciertos factores que se consideraron determinantes, se propusieron métricas para cuantificarlos y se vienen desarrollando modelos para procurar reproducir los escenarios en los que se aplican los sistemas de trazabilidad y sus resultados. Tan pronto se pusieron a

prueba los primeros modelos surgió la necesidad de compararlos, y develar el interrogante sobre los que son más efectivos para la predicción de la trazabilidad de los proyectos de software. En este trabajo se propusieron tres modelos y se compararon sus desempeños con un caso de estudio adoptado como referencia. Se llegó a la conclusión que el módulo del vector que define un proyecto en un hiperespacio de ocho dimensiones es el parámetro característico que conduce al mejor desempeño en la predicción del resultado que puede esperarse de cierto proyecto. También que los tres modelos estudiados brindaron resultados que no son muy diferentes entre sí. El caso de estudio utilizado contiene una variedad de situaciones y fue concebido para cubrir todo el espectro que definen los factores adoptados. Es decir, se contemplaron proyectos y productos software de diferente naturaleza y magnitud, como así también organizaciones de diferente tamaño y madurez. Sin embargo, no puede descartarse que el caso de estudio adoptado presente limitaciones por no contemplar ciertas combinaciones especiales de los factores usados. Por ello, los próximos pasos estarán orientados a corroborar estas conclusiones con otros casos de estudio, para lo cual se viene trabajando paralelamente en obtener más datos de casos reales en la cantidad, calidad y variedad necesaria. Una vez identificado el mejor clasificador, será necesario concentrar la atención el punto en el que presenta el mejor desempeño y éste es otro campo a explorar, en el que el AUC y el índice de Youden desempeñarán un rol muy importante. Por último, es necesario puntualizar que se sigue trabajando con una familia de modelos apoyados en redes neuronales multicapas de perceptrones, que representan una alternativa superadora de los que aquí fueron presentados.

La posibilidad de disponer de un modelo efectivo de predicción de trazabilidad de proyectos de software justifica con creces el esfuerzo que se viene realizando.

Referencias

1. Blaauboer, F., Sikkel, K., Aydin, M.: Deciding to adopt requirements traceability in practice. Proc. of the 19th Int. Conf. on Advanced Infor. Systems Engineering. Springer-Verlag (2007).
2. Kannenberg, A., Saiedian, H.: Why Software Requirements Traceability Remains a Challenge. CrossTalk: The Journal of Defense Software Engineering. July/August, 14-19 (2009).
3. Ramesh, B.: Factors influencing Requirements Traceability Practice. Communications of the ACM. 41(12), 37-44. (1998)
4. Giró, J., Vázquez, J., Meloni, B., Constable, L., Jornet, A.: Aseguramiento de la Trazabilidad en Proyectos de Desarrollo de Sistemas de Software. Proyecto de Investigación, Secretaría de Ciencia y Tecnología, Código SCyT 1214/10. (2010)
5. Giró, J., Vázquez, J., Meloni, B., Constable, L., Jornet, A.: Modelos para anticipar la factibilidad de que un proyecto de desarrollo de software sea trazable. Workshop de Ingeniería de Software, CACIC 2011. Universidad Nacional de La Plata, 837-846 (2011).
6. Giró, J., Vázquez, J., Meloni, B., Constable, L., Jornet, A.: Hacia una respuesta al interrogante de si será factible trazar un cierto proyecto de desarrollo de software. Informe Técnico 2012/01, Proyecto 1214, SCyT, FRC, UTN (2012).
7. Giró, J., Vázquez, J., Meloni, B., Constable, L., Jornet, A.: Uso del Análisis ROC para anticipar la conveniencia de trazar proyectos de software. Workshop de Ingeniería de Software, CACIC 2012. Universidad Nacional del Sur, Ciudad de Bahía Blanca (2012).
8. Giró, J., Vázquez, J., Meloni, B., Constable, L.: Evaluación de variantes en modelo destinado a anticipar la conveniencia de trazar proyectos de software. Trabajo aceptado bajo el código 5759 a ser presentado el día Jueves 24 de Octubre en el X WIS del CACIC 2013, CAECE, Ciudad de Mar del Plata (2013).
9. Powers, D.: Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. Technical Report SIE-07-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia (2007).
10. Fawcett, T.: An introduction to ROC analysis. Elsevier ScienceDirect, Pattern Recognition Letters, 27 - 861-874 (2006).
11. Shin Y., Huffman Hayes J., Cleland-Huang J.: A framework for evaluating traceability benchmark metrics. TR:12-001, DePaul University, School of Computing (2012).
12. Biggerstaff, B.: Comparing diagnostic tests: a simple graphic using likelihood ratios. Statistics in Medicine 19 (5) 649-663 (2000).

Datos de Contacto

J.Giró, juanfgiro@gmail.com

J.Vazquez, jcjvazquez@gmail.com

B.Meloni, bemeloni@gmail.com

L.Constable, leticiaconstable@gmail.com

Dto. de Ingeniería en Sistemas de In-formación, FRC, UTN, Maestro López esq. Cruz Roja Argentina, Ciudad de Córdoba.