

DETERMINACION DEL MODELO DE META-ANALISIS PARA EXPERIMENTACION EN INGENIERIA DEL SOFTWARE

Amatriain, H., Fernandez, E., Dieste, O., Pollo-Cattaneo, F., García-Martínez, R.

Grupo de Ingeniería de Software Experimental

Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo 28660 Boadilla del
Monte. Madrid. España.
Tel + 34 91 336 5011
odieste@fi.upm.es

Grupo Investigación en Sistemas de Información

Depto. Desarrollo Productivo y Tecnológico
Universidad Nacional de Lanús
29 de Septiembre 3901 (1826) Remedios de
Escalada, Lanús. Argentina.
Tel +54 11 6322-9200 Ext. 194
rgarcia@unla.edu.ar

Grupo de Estudio en Metodologías de Ingeniería de Software

Facultad Regional Buenos Aires.
Universidad Tecnológica Nacional.
Medrano 951 (C1179AAQ) Ciudad Autónoma
de Argentina. Buenos Aires
Tel +54 11 4867-7511
fpollo@posgrado.frba.utn.edu.ar

CONTEXTO

Este proyecto de investigación se desarrolla en el marco de la cooperación existente entre el Grupo de Ingeniería de Software Experimental (GrISE) de la Facultad de Informática de la Universidad Politécnica de Madrid, el Grupo de Investigación en Sistemas de Información (GISI) del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús y el Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional.

ESTADO ACTUAL DE CONOCIMIENTO SOBRE EL TEMA

La síntesis cuantitativa [Goodman, 1996] de experimentos, más conocida como Meta-Análisis [Glass, 1976], consiste en combinar los resultados de varios estudios experimentales con el objeto de generar nuevas piezas de conocimiento. Estas nuevas piezas de conocimientos serán más generales y fiables que los resultados obtenidos por los estudios individuales, ya que dichas piezas de conocimiento están sustentadas por una mayor cantidad de evidencia empírica. La síntesis cuantitativa es habitual en disciplinas con una fuerte componente empírica, tales como la medicina, la psicología o la física. En Ingeniería del Software (IS), se ha extendido el uso de la síntesis cuantitativa desde que fue propuesta por [Basilli et al, 1996]. El primer trabajo de meta-análisis que se conoce es el desarrollado por Miller [1999] que intentó combinar 4 experimentos en 1999, y el mayor trabajo de este tipo es el desarrollado por [Dyba et al, 2007] que combinó 15 experimentos.

Si todos los experimentos incluidos en un meta-análisis fueran desarrollados con el mismo nivel de fiabilidad, bastaría con un simple promedio para determinar el resultado general. Ahora bien, en la práctica esto no es así, hay trabajos que son más fiables que otros (por ejemplo por contener mayor cantidad de sujetos experimentales), por este motivo el resultado final debe obtenerse mediante un promedio ponderado que de mayor peso a los estudios más fiables y menor a los que a priori podrían estar asociados a un mayor nivel de error [Glass, 1976].

Existen dos variantes bien diferenciadas para realizar el promedio ponderado de los resultados:

- 1) el modelo de efectos fijo, que asume que la variación entre los resultados de los experimentos se debe únicamente al error experimental, pero existe un valor de efecto único para toda la población [Hedges y Olkin, 1985], cuya función de estimación es:

$$d^* = \frac{\sum d_i / \sigma_i^2(d)}{\sum 1 / \sigma_i^2(d)} \quad (1)$$

d^* es el tamaño de efecto global
 $\sum d_i / \sigma_i^2(d)$ es la suma de los efectos individuales dividido por la varianza
 $\sum 1 / \sigma_i^2(d)$ es la suma de la inversa varianza
 $\sigma_i^2(d)$ es la varianza total

- 2) el modelo de efectos aleatorios, que se base en la premisa de que la variación los resultados se debe no solo al error experimental, sino también a ciertas variables no controladas que los afecta y, por ende, no existe un único efecto global,

sino que este depende de los estudios que hacen parte del proceso de síntesis [Hedges y Olkin, 1985]. La función de estimación es:

$$\Delta = \frac{\sum d_i / \gamma^2_i}{\sum 1 / \gamma^2_i} \quad (2)$$

Δ es el efecto global
 $\sum d_i / \gamma^2_i$ es la sumatoria de los efectos individuales
 $\sum 1 / \gamma^2_i$ es la sumatoria de la inversa de las varianzas entre-estudios e intra-estudios

Como puede verse en las funciones 1 y 2 la diferencia entre ambos modelos radica en que para el modelo de efecto fijo no se requiere estimar la varianza entre estudios porque supone que esta debería anularse entre los distintos resultados, mientras que para el modelo de efectos aleatorios es fundamental determinar cuál es el nivel de variación entre los resultados (la varianza entre estudios) ya que este parámetro permitirá mitigar el ruido generado por las variaciones. Bajo estas condiciones, se dice, que si un meta-análisis incluyera experimentos con valores similares, decir, no existe ruido producido por variables externas, ambos métodos deben dar igual resultado [Borenstein, 2007].

Como todo método estadístico el meta-análisis se ve afectado por dos tipos de errores: el de tipo I, que consiste en indicar que ambos tratamientos son iguales cuando en realidad no lo son; y el de tipo II, que consisten en indicar que uno de los tratamientos es mejor que otro cuando en realidad no lo es [Noortgate y Onghena, 2003].

En la práctica el error de tipo I es fijado de antemano (habitualmente como $\alpha=0,05$) y el error de tipo II queda supeditado, en gran medida, al tamaño del experimento (la cantidad de sujetos que el mismo incluya) que en IS muchas veces es menor a lo que la teoría estadística sugiere. Este hecho, en IS, muchas veces se debe a la falta de presupuesto económico, tiempo, recursos humanos calificados y, porque no decirlo, falta de cultura experimental, y provoca que se asuma que dos tratamientos (técnicas o métodos) se comportan de igual forma cuando en realidad no lo hacen.

Para la estadística, antes de decidir qué modelo de meta-análisis aplicar se debe determinar qué grado de heterogeneidad existe entre los

experimentos (nivel de variación entre los resultados), para ello existen varios métodos, siendo el más difundido y conocido es el denominado método Q. El problema es que este método carece de potencia estadística cuando se aplica a pocos experimentos [Schmidt y Hunter, 2003]. Este hecho provoca que la determinación del modelo de agregación en IS, donde en general los meta-análisis se aplican a pocos experimentos, deba hacerse en gran medida en base a la experiencia del investigador. En tal sentido el modelo de efectos aleatorios sería el más se adapte a la IS, un contexto experimental donde los cambios en la cultura de las organizaciones y el contexto tecnológico varían constantemente, ahora bien, según algunos autores, este modelo podría no ser aplicable debido a que el cálculo de la varianza entre estudios puede tener un elevado nivel de error cuando la cantidad de experimentos incluidos en el meta-análisis es bajo [Borenstein, 2007] como sucede en IS.

OBJETIVOS DE INVESTIGACIÓN

El objetivo de este trabajo es determinar, hasta qué punto el error producido por la baja cantidad de experimentos y pocos sujetos experimentales, situación descriptiva de la Ingeniería de Software, afecta a la fiabilidad y potencia estadística de los modelos de manera similar a como se hizo en [Dieste et al, 2011], intentando:

- (a) determinar cómo afecta la cantidad de experimentos a la fiabilidad de los modelos de meta-análisis;
- (b) determinar cómo afecta la cantidad de sujetos por experimentos a la fiabilidad de los modelos de meta-análisis;
- (c) determinar cómo afecta la cantidad de experimentos a la potencia estadística de los modelos de meta-análisis;
- (d) determinar cómo afecta la cantidad de sujetos por experimentos a la potencia estadística de los modelos de meta-análisis.

En base a estos resultados se podrá determinar qué modelo aporta el resultado más fiable en función de las características de los

experimentos que van formar parte del meta-análisis.

METODOLOGIA DE TRABAJO

El desarrollo de este proyecto utilizará la metodología propia de la investigación documental, del estudio de casos, de técnicas de análisis comparativo y de síntesis de comparaciones. Se preveen las siguientes tareas:

- Se desarrollará una simulación de Montecarlo que incluya a experimentos de igual tamaño y donde existe un tamaño de efectos poblacional único. El objetivo de esta simulación es determinar si el modelo de efectos aleatorios se comporta igual que el modelo de efecto fijo cuando la variación entre los experimentos es nula.
- Se desarrollará una simulación de Montecarlo que incluya a experimentos de diferente tamaño y donde existe un tamaño de efectos poblacional único. El objetivo de esta simulación es determinar si el modelo de efectos aleatorios se comporta igual que el modelo de efecto fijo cuando la variación entre los experimentos es nula y el pose del resultado final esta afecto por algún experimento.
- Se desarrollará una simulación de Montecarlo que incluya a experimentos de igual tamaño y donde no existe un tamaño de efectos poblacional único. El objetivo de esta simulación es determinar cuan sensible es al modelo de efecto fijo a la heterogeneidad.
- Se desarrollará una simulación de Montecarlo que incluya a experimentos de diferente tamaño y donde no existe un tamaño de efectos poblacional único. El objetivo de esta simulación es determinar cómo afecta el ruido producido por la heterogeneidad y experimentos de diferentes tamaños a ambos modelos.

RESULTADOS OBTENIDOS/ESPERADOS

Se espera determinar con precisión cuando es factible aplicar cada uno de los modelos de meta-análisis en le campo de experimentos en Ingeniería de Software, no solo porque el contexto experimental lo requiera sino porque

el método tiene la capacidad de arrojar un resultado mas seguro.

FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por dos investigadores formados y tres investigadores en formación. En el marco de este proyecto se están desarrollando una tesis doctoral y una tesis de maestría.

BIBLIOGRAFÍA

- Basili, V. R., Green, S., Laitenberger, O., Lanubile, F., Shull, F., Sörumgård, S., Zelkowitz, M.; 1996; The empirical investigation of perspective-based reading, *International Journal on Empirical Software Engineering*, Vol. 1, No. 2; pp. 133–164.
- Borenstein, M.; Hedges, L.; Rothstein, H.; 2007; *Meta-Analysis Fixed Effect vs. random effect*; www.Meta-Analysis.com
- Dieste, O., Fernández, E., García, R., Juristo, N. (2011) “Comparative analysis of meta-analysis methods: when to use which?” 6th EASE Durham (UK). En prensa.
- Dyba, T., Aricholm, E.; Sjoberg, D.; Hannay J.; Shull, F.; 2007; Are two heads better than one? On the effectiveness of pair programming. *IEEE Software*;12-15.
- Glass, G; 1976; Primary, secondary, and meta-analysis of research. *Educational Researcher* 5: 3-8
- Goodman C.; 1996; *Literature Searching and Evidence Interpretation for Assessing Health Care Practices*; SBU; Stockholm.
- Hedges, L.; Olkin, I.; 1985; *Statistical methods for meta-analysis*. Academic Press.
- Miller, J; 1999; Can Results from Software Engineering Experiments be Safely Combined? *IEEE METRICS*, 152-158
- Noortgate, W and Onghena p.; 2003; Estimating the mean effect size un meta-analysis: Bias, precision, and mean squared error of different weighting methods. *Behavioral research methods, instruments and computers*; 35:504-511
- Schmidt, F. and Hunter, J. (2003) “*Handbook of Psychology, Research Methods in Psychology*”, Chapter 21, “*Meta-Analysis*”, Schinka, J., Velicer, W., Weiner, I. Editors, Volume 2