

On the Effectiveness of Concern Metrics to Detect Code Smells: An Empirical Study

Autores: Juliana Padilha, Juliana Pereira, Eduardo Figueiredo, Jussara Almeida, Alessandro Garcia e Cláudio Sant'Anna.

Apresentação: Paulo H. Mariano

Introdução

- ▶ O artigo descreve um estudo empírico sobre a utilização de métricas de interesses (concern metrics) na detecção de três bad smells específicos. São analisados os resultados em comparação com outras métricas.



Introdução

- ▶ *Concern*(interesse, preocupação) é qualquer propriedade importante ou área de interesse de um sistema.
- ▶ Regras de negócio, persistência e segurança são exemplos típicos de *concerns*.



Introdução

- ▶ A separação inadequada de *concerns* prejudica a modularidade do projeto afetando também a manutenção do software em questão.
- ▶ A detecção dessas falhas não é algo trivial e exige um apoio eficaz.



Introdução

- ▶ Métricas de software são os principais meios para se avaliar a capacidade de manutenção de um sistema de informação.
- ▶ Medição de software também é vista como uma solução comum para encontrar sintomas de determinadas falhas no projeto, como por exemplo encontrar Bad Smells.



Métricas de Software e Bad Smells

- ▶ Métricas de software tem desempenhado um papel importante na compreensão e análise de sistemas de informação.
- ▶ As métricas de software foram divididas em três conjuntos: métricas tradicionais, métricas de interesse (*concerns*) e métricas híbridas.



Métricas Tradicionais

- ▶ São as métricas comumente utilizadas.
- ▶ Métricas Orientadas a Objetos propostas por Chidamber e Kemerer.
- ▶ Métricas bem documentadas na literatura de engenharia de software.



Métricas Tradicionais

Table 1. Definitions of Traditional Metrics

Metric	Definition
Coupling between Objects (CBO)	Number of classes from which a class calls methods or accesses attributes.
Lack of Cohesion in Methods (LCOM)	Divides pairs of methods that do not access common attributes by pairs that do access.
Lines of Code (LOC)	Total number of lines of code.
Number of Attributes (NOA)	Number of attributes defined in a class.
Number of Methods (NOM)	Number of methods defined in a class.
Weighted Methods per Class (WMC)	Number of methods and their parameters in a class



Métricas Tradicionais

- ▶ Motivos para a seleção das métricas:
 - ✓ Encontrar combinações envolvendo essas métricas pode ser algo relevante nesse estudo.
 - ✓ Selecionar um número reduzido de métricas.
 - ✓ Parecem ajudar nas tarefas de manutenção de software.



Métricas de Interesse(Concern)

- ▶ Foram definidos com o objetivo de capturar propriedades de modularidade associado com a realização das preocupações em artefatos de software.
- ▶ O objeto é a identificação de falhas em projetos causadas pela má modularização dos interesses.



Métricas de Interesse(Concerns)

Table 2. Definitions of Concern-based Metrics

Metric	Definition
Concern Diffusion over Components (CDC)	Number of classes whose main purpose is to contribute to the implementation of a concern and the number of other classes that access them.
Concern Diffusion over Operations (CDO)	Number of methods whose main function is to implement a concern.
Concern Diffusions over LOC (CDLOC)	Number of transition points for each concern through the lines of code. Transition points are points in the code where there is a “concern switch”.
Number Concerns per Component (NCC)	Number of concern in each class.



Métricas de Interesse(Concerns)

- ▶ Motivos para a seleção dessas métricas:
 - ✓ Tem sido utilizadas com sucesso em vários estudos relacionados com manutenção de software.
 - ✓ Nenhum estudo sistemático foi realizado para avaliar o suporte a detecção de bad smells.



Métricas Híbridas

- ▶ Combinação das métricas anteriores.





Bad Smells

Bad Smells

- ▶ O artigo investiga o uso de métricas de interesses para detectar três bad smells: Divergent Change , Shotgun Surgery e God Class.
- ▶ Esses Bad Smells foram escolhidos por serem comuns em sistemas de informação e estarem relacionados a má modularização dos interesses.



Divergent Change

- ▶ Este bad smell ocorre quando uma classe é frequentemente alterada de maneiras diferentes por razões diferentes.
- ▶ Por exemplo, temos que mudar três métodos de uma classe cada vez que obtemos uma nova base de dados ou temos que mudar outros quatro métodos cada vez que um novo cálculo financeiro é incluído.



Shotgun Surgery

- ▶ Este bad smell é o oposto do Divergent Change.
- ▶ Shotgun Surgery ocorre cada vez que fazemos um tipo de mudança que leva a uma série de pequenas mudanças em muitas classes diferentes.



God Class

- ▶ Esse bad smell representa um objeto que sabe muito ou que faz muitas coisas. Uma classe que cresceu muito em sua lógica e faz quase tudo no sistema.
- ▶ Uma classe que implementa muitos interesses, e por isso tem muitas responsabilidades.





Configurações do Estudo



Configurações do Estudo

- ▶ Este estudo tem como objetivo avaliar a eficácia das métricas de interesses na detecção de bad smells.
- ▶ O estudo se baseia em métricas tradicionais como linha de base para o experimento.
- ▶ É realizada uma análise comparativa entre as métricas tradicionais e as métricas de interesses, a fim de identificar se as métricas de interesses ajudam na detecção dos três bad smells especificados.



Configurações do Estudo

- ▶ **Sistemas Analisados: HealthWatcher e MobileMedia**
- ▶ Foram escolhidos por já terem sido utilizados em outros estudos relacionados a facilidade de manutenção.
- ▶ Acesso aos desenvolvedores e especialistas dos softwares.



Configurações do Estudo: HealthWatcher

- ▶ É um sistema web para registro e gestão de queixas ao sistema público de saúde.
- ▶ Possui cerca de 6 mil linhas de código.
- ▶ Alguns interesses(concerns) implementados foram: regras de negocio, concorrência, distribuição, manipulação de exceções e persistência.



Configurações do Estudo: MobileMedia

- ▶ Aplicativo móvel de manipulação de fotos, vídeos e músicas
- ▶ Possui cerca de 4 mil linhas de código.
- ▶ Usada a 7^o versão do sistema.
- ▶ Alguns interesses(concerns) implementados foram: classificação, favoritos, manipulação de exceção, segurança e persistência.



Configurações do Estudo:

Lista de referencias dos Bad Smells

- ▶ Foi realizada uma análise sistemática do código fonte dos dois sistemas para a detecção das classes com os bad smells citados.
- ▶ Dois especialistas de cada sistema participaram dessa análise para a construção da lista de referencias.



Configurações do Estudo:

Lista de referencias dos Bad Smells

Table 3. Code Smell Reference List for Health Watcher and MobileMedia

System	Smell	Classes in the Reference List
Health Watcher	Divergent Change	EmployeeRecord, HealthWatcherFacade, HealthUnitRecord, PersistenceMechanism, IFacade, HealthWatcherFacadeInit, IPersistenceMechanism, ServletInsertEmployee, ComplaintRecord, ServletSearchComplaintData, ServletUpdateComplaintData, ServletUpdateHealthUnitData
	Shotgun Surgery	PersistenceMechanism, ComplaintRecordRDB, EmployeeRepositoryRDB, IComplaintRepository, HealthUnitRepositoryRDB, IPersistenceMechanism, IHealthUnitRepository, IEmployeeRepository
	God Class	HealthWatcherFacade, HealthWatcherFacadeInit, PersistenceMechanism
Mobile Media	Divergent Change	ImageMediaAccessor, MediaController, MediaAcessor, MediaListController
	Shotgun Surgery	ControllerInterface, MediaAccessor, ScreenSingleton



Configurações do Estudo: Indivíduos

- ▶ O estudo envolveu um conjunto de 54 indivíduos , definidos com os nomes de S1 a S54, de duas instituições diferentes: UFMG no Brasil e Lancaster no Reino Unido.
- ▶ UFMG: 11 profissionais de TI, 4 doutorandos e 12 estudantes de graduação.
- ▶ Lancaster: 14 doutorandos e 13 alunos de graduação.



Configurações do Estudo: Indivíduos

- ▶ Indivíduos foram organizados de forma que cada grupo trabalhou com apenas um conjunto de métricas: métricas tradicionais, métricas de interesse e métricas híbridas.
- ▶ O estudo foi realizado utilizando os designs OO de ambos sistemas.
- ▶ HealthWatcher foi usado pelos participantes de Lancaster.
- ▶ MobileMedia foi usado pelos participantes da UFMG.



Configurações do Estudo: Indivíduos

- ▶ Antes de realizar o experimento foi feito um questionário para saber o nível de cada participante do estudo.
- ▶ Os indivíduos responderam a perguntas sobre seu nível de conhecimento em relação à diagrama de classes, programação Java e métricas de software.



Configurações do Estudo: Indivíduos

- ▶ 13 indivíduos não responderam ao questionário.
- ▶ Cerca de 60% dos indivíduos têm de moderado a alto conhecimento sobre diagrama de classes e programação Java.
- ▶ Cerca de 70% dos indivíduos tem de moderado a alto conhecimento em pelo menos um dos tópicos pesquisados.



Configurações do Estudo: Indivíduos

Table 4. Background Data of Subjects

		Divergent Change	Traditional	Concern	Hybrid	No Answer
Knowledge	Class Diagram	S5 - S6	S9 - S11	S14 - S24	S1, S2, S3, S7, S8, S12, S13, S18	
	Java Programming	S5 - S6	S9 - S11	S14 - S24		
	Measurement	-	S9	S16, S20, S22, S24		
	Academic Experience	S4, S6	S9	S19, S21-S24		
	Work Experience	S5	S10,S11	S14 - S17, S20		
Shotgun Surgery						
Knowledge	Class Diagram	S28, S29	S31, S32	S34 - S37	S25, S26, S30, S33, S38	
	Java Programming	S28, S29	S31, S32	S34 - S37		
	Measurement	-	S31	S35, S36		
	Academic Experience	S27, S29	S31	S39, S41-S44		
	Work Experience	S28	S32	S33 - S37, S40		
God Class						
Knowledge	Class Diagram	S46	S48 - S50	S51- S54	-	
	Java Programming	S45, S46	S48 - S50	S51- S54		
	Measurement	-	S49 - S50	S52 -S54		
	Academic Experience	S46, S47	S49, S50	S52-S54		
	Work Experience	S45	S48	S51		



Tarefas Experimentais



Tarefas Experimentais

- ▶ O estudo foi precedido de uma sessão de 30 minutos de treino para a familiarização das métricas e o código fonte dos sistemas.
- ▶ Cada individuo recebeu um documento contendo:
 - Uma breve explicação e uma visão parcial do sistema por meio de um diagrama de classe.
 - Uma descrição dos interesses envolvidos nos respectivos sistemas.
 - Diretrizes e passos que deviam ser seguidos.
 - Perguntas que deveriam ser respondidas.
 - Informações que deveriam ser registradas.



Tarefas Experimentais

- ▶ Foi pedido o registro do tempo gasto para a realização das tarefas.
- ▶ Identificar quais métricas foram relevantes para identificar os Bad Smells.





Resultados

Métricas de Avaliação de Resultados: Recall e Precisão

- ▶ Foram utilizadas três valores para avaliar os resultados (com base na tabela 3 que contem as classes identificadas com bad smells) :
 - Verdadeiro Positivo(TP)
 - Falso Positivo(FP)
 - Falso Negativo(FN)



Métricas de Avaliação de Resultados: Recall e Precisão

- ▶ Verdadeiro Positivo(TP): Quantifica o número de bad smells corretamente identificados.
- ▶ Falso Positivo(FP): Quantifica o número de bad smells erradamente identificados.
- ▶ Falso Negativo(FN): Quantifica o numero de bad smells não assinalados.



Métricas de Avaliação de Resultados: Recall e Precisão

- Recall mede a fração de classes relevantes listados por um indivíduo. As classes relevantes são classes que aparecem na lista de referência (TP + FN).
- Precisão mede a proporção de bad smells detectados corretamente pelo total de classes listadas. (TP + FP).

$$\text{Recall (R)} = \frac{TP}{TP + FN}$$

$$\text{Precision (P)} = \frac{TP}{TP + FP}$$



Métricas de Avaliação de Resultados: Recall e Precisão

- ▶ O estudo focou principalmente no recall, visto que é uma medida mais completa.
- ▶ Alto recall significa que o individuo foi capaz de identificar a maioria dos bad smells.
- ▶ Alta precisão indica que verdadeiros positivos(TP) foram mais identificados do que falsos positivos(FP).



Suporte das métricas de interesses para a detecção de Divergent Change

- ▶ Considerados três dados nessa análise: Recall, Precisão e Tempo(minutos gastos para completar a tarefa).
- ▶ 24 indivíduos deviam identificar o bad smell Divergent Change nos sistemas de destino.
- ▶ Métricas de interesse e métricas híbridas obtiveram melhores resultados que as métricas tradicionais.



Suporte das métricas de interesses para a detecção de Divergent Change

- ▶ O grupo que utilizou métricas de interesses obteve um recall de 62%.
- ▶ Dois em cada cinco indivíduos identificaram todas as classes com o bad smell Divergent Change(100% de recall).



Suporte das métricas de interesses para a detecção de Divergent Change

- ▶ Grupo com métricas tradicionais obtiveram apenas 33% de recall.
- ▶ No grupo com métricas híbridas houve uma variação de 0 a 100% de recall(SI9 e SI6), sendo a média de 41%.



Suporte das métricas de interesses para a detecção de Divergent Change

- ▶ Esses resultados mostram que mesmo sendo utilizadas de maneira isolada, as métricas de interesses são efetivas em na detecção de Divergent Change.



Suporte das métricas de interesses para a detecção de Divergent Change

- ▶ Resultados do estudo para a detecção do bad smell Divergent Change.

Table 5. Results for Divergent Change

Group	Traditional						Concern						
Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11		
R(%)	17	17	17	33	25	25	100	100	33	25	50		
P(%)	67	50	40	50	17	25	63	100	100	25	29		
T(min)	15	15	40	38	41	36	26	29	29	15	33		
Group	Hybrid												
Subject	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24
R(%)	75	8	25	50	100	25	50	0	50	25	50	25	50
P(%)	100	50	75	25	67	33	40	0	67	17	40	17	50
T(min)	40	31	23	36	27	39	24	11	18	19	13	13	12



Detecção de Shotgun Surgery

- ▶ Utilizando as métricas listadas nenhum individuo se destacou na detecção desse bad smell.
- ▶ Apenas um individuo em cada grupo alcançou mais de 60% de recall.
- ▶ S28 alcançou 67% de recall com métricas tradicionais.
- ▶ S30 alcançou 75% utilizando métricas de interesses.
- ▶ S35 alcançou 67% recall utilizando métricas híbridas.



Detecção de Shotgun Surgery

- ▶ O grupo que utilizou métricas de interesses foram um pouco melhor, alcançaram a media de 44% de recall.
- ▶ No entanto, as taxas de detecção são muito baixas para quase todas as métricas, sugerindo que as métricas utilizadas não são efetivas na detecção desse bad smell.



Detecção de Shotgun Surgery

- ▶ Resultados do estudo para a detecção do bad smell Shotgun Surgery.

Table 6. Results for Shotgun Surgery

Group	Traditional					Concern		
Subject	S25	S26	S27	S28	S29	S30	S31	S32
R(%)	13	13	0	67	33	75	25	33
P(%)	25	33	0	25	25	35	40	25
T(min)	6	10	27	12	14	13	28	14

Group	Hybrid											
Subject	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42	S43	S44
R(%)	13	50	67	33	33	33	0	0	33	0	0	0
P(%)	25	80	6	33	25	33	0	0	20	0	0	0
T(min)	35	14	19	15	4	10	14	9	21	3	7	5



Detecção de Shotgun Surgery

- ▶ Quase todos os indivíduos tiveram baixa precisão.
- ▶ Mais da metade das detecções estavam incorretas, independente da métrica utilizada.
- ▶ Os indivíduos gastaram menos tempo na detecção desse bad smell.
- ▶ Isso pode indicar que os indivíduos quando não tem as ferramentas adequadas para detectar um bad smell eles não gastam muito tempo procurando.



Detecção de God Class

- ▶ Os dados encontrados sugerem que a utilização de métricas tradicionais de forma isolada, não são efetivas na identificação de God Class.
- ▶ Dois indivíduos(S45 e S46) do grupo com métricas tradicionais marcaram apenas 33% em recall e precisão.
- ▶ Em contraste, dois de três indivíduos do grupo com métricas de interesses e três de quatro indivíduos do grupo híbrido atingiram 100% de recall.



Detecção de God Class

- ▶ Portanto, a utilização conjunta das métricas de interesses e das métricas tradicionais parecem ter sucesso em identificar esse bad smell em particular.



Detecção de God Class

- ▶ Resultado do estudo para a identificação de God Class.

Table 7. Results for God Class

Group	Traditional			Concern			Hybrid			
Subject	S45	S46	S47	S48	S49	S50	S51	S52	S53	S54
R(%)	33	33	67	100	67	100	33	100	100	100
P(%)	33	33	67	75	100	75	50	100	60	75
T(min)	18	25	27	37	66	43	22	53	51	35



Analise Estatística e Discussão

Análise Estatística e Discussão

- ▶ Essa sessão busca responder três perguntas a respeito do estudo:
 - RQ1. Qual a precisão do desempenho das métricas de interesses em comparação com as métricas tradicionais?
 - RQ2. O background dos indivíduos impacta na eficiência de detecção dos bad smells?
 - RQ3. Existe uma combinação de métricas que aumenta o recall de detecção de bad smells?



Comparando métricas de interesses e métricas tradicionais.

▶ Respondendo a primeira questão:

Qual a precisão do desempenho das métricas de interesses em comparação com as métricas tradicionais?



Comparando métricas de interesses e métricas tradicionais.

- ▶ Para responder a essa pergunta foi investigado se o tipo de sistema(Health Watcher e MobileMedia) influencia na detecção dos bad smells.



Comparando métricas de interesses e métricas tradicionais.

- ▶ Os resultados para o sistema HealthWatcher observados são significativamente melhores (75% maior recall, em média). Em outras palavras, a detecção de bad smells usando as métricas de interesses levam a um maior recall nesse sistema.



Comparando métricas de interesses e métricas tradicionais.

- ▶ Enquanto que o sistema utilizado tem impacto sobre a detecção de bad smells usando métricas de interesses, a detecção usando métricas tradicionais e híbridas não é significativamente influenciada por isso.



Comparando métricas de interesses e métricas tradicionais.

► Tabela com os resultados:

Table 8. Confidence Intervals (CI) for the average recall in Health Watcher and MobileMedia

Systems	Health Watcher (HW)			Mobile Media (MM)		
	Traditional (T)	Concern (C)	Hybrid (H)	Traditional (T)	Concern (C)	Hybrid (H)
All	(13.26, 35.34)	(54.59, 95.40)	(20.98, 91.02)	(18.70, 46.49)	(22.94, 49.32)	(19.08, 43.32)
DC	(11.6, 30.4)	(12.5, 142.9)	(-22.7, 94.8)	(23.9, 26.5)	(-41.4, 116.4)	(27.2, 57.9)
SS	(-4.0, 21.3)	(-107.9, 207.9)	(-85.3, 148.3)	(-57.3, 157.3)	(28.9, 37.9)	(6.4, 33.4)
GC	(11.2, 77.4)	(56.9, 121.1)	(43.8, 123)	-	-	-



Comparando métricas de interesses e métricas tradicionais.

- ▶ Foi concluído que as métricas de interesses produziram significativamente maior recall, em relação a métricas tradicionais para o sistema HealthWatcher.
- ▶ Para o sistema MobileMedia, os resultados médios são melhores para métricas de interesses.
- ▶ As métricas de interesses tiveram um desempenho superior ao das métricas híbridas nos dois sistemas.



Comparando métricas de interesses e métricas tradicionais.

- ▶ O artigo afirma que, em geral, métricas de interesses são mais adequados, entre aquelas analisadas, para a detecção dos três tipos de bad smells estudados.
- ▶ Utilização de métricas de interesses foi consistentemente melhor em comparação com métricas tradicionais nos casos de detecção Divergent Change e God Class.
- ▶ No entanto, a diferença entre as duas métricas para a detecção de Shotgun Surgery não é estatisticamente significativa.



Comparando métricas de interesses e métricas tradicionais.

- ▶ Estes resultados indicam que a precisão do conjunto de métrica depende em grande parte da adequação de cada métrica para quantificar uma propriedade explicitamente mencionada na definição do bad smell.
- ▶ Exemplo:

God Class é caracterizada pela "Elevada quantidade de membros da classe que possuem muitas responsabilidades" . Esta propriedade é melhor identificada com métricas de interesses.



Comparando métricas de interesses e métricas tradicionais.

- ▶ Dados também sugerem que a detecção de Divergent Change com apenas métricas tradicionais parece mais difícil em comparação com as métricas de interesse.
- ▶ A explicação pode estar relacionada ao fato de que este bad smell está intimamente relacionado à má separação de interesses.



Comparando métricas de interesses e métricas tradicionais.

- ▶ 10 de 18 indivíduos em ambos os grupos alcançaram 68% de recall em média, utilizando métricas de interesses ou métricas híbridas .



Background dos Indivíduos

- ▶ Essa sessão responde a seguinte questão:
O background dos indivíduos impacta na eficiência de detecção dos bad smells?



Background dos Indivíduos

- ▶ Foi avaliado o impacto do conhecimento dos indivíduos e o tempo gasto por eles sobre utilizando métricas de interesses.



Background dos indivíduos

- ▶ Os resultados encontrados indicam que, o recall tende a aumentar com a experiência de trabalho e o tempo gasto na detecção.
- ▶ 96% do recall pode ser atribuído ao fator tempo. Indivíduos com pouca experiência, mas que dedicaram um tempo maior a detecção dos bad smells com métricas de interesses foram bem sucedidos.
- ▶ 4% do recall pode ser atribuído a experiência profissional como desenvolvedor.



Métricas reunidas em conjunto

- ▶ Essa sessão responde a seguinte questão:
Existe uma combinação de métricas que aumenta o recall de detecção de bad smells?



Métricas reunidas em conjunto

- ▶ Foram relatadas as métricas consideradas uteis para identificar os bad smells.
- ▶ A fim de determinar quais métricas foram usadas em conjunto para detectar os bad smells, foi realizada uma análise dos indivíduos que usaram as mesmas métrica e obtiveram alta taxa de recall.
- ▶ A análise foi restrita a um numero de recall superior a 30%.



Métricas reunidas em conjunto

- ▶ Métricas consideradas uteis para a detecção de Divergent Change.

Table 9. Metrics Considered Useful for Divergent Change

Metrics	NCC	LCOM	CDC	LOC
Subjects who used this metric	S7, S8, S9, S11, S12, S14, S15, S16, S20, S23, S24	S1, S2, S4, S6, S12, S13, S14, S15, S17, S22, S24	S8, S10, S23	S2, S17, S20
Average of recall	60%	34%	50%	31%



Métricas reunidas em conjunto

- ▶ NCC parece a métrica mais eficaz (entre os analisados) para detectar Divergent Change.
- ▶ Indivíduos que indicaram NCC como não sendo útil alcançaram menos de 11% de recall.
- ▶ Além disso, os indivíduos que indicaram NCC e LCOM como sendo útil alcançaram 50% de recall em média.
- ▶ Indivíduos que indicaram NCC e LCOM como não sendo útil alcançaram 0% de recall; como é o caso de S19.



Métricas reunidas em conjunto

- ▶ A maioria dos indivíduos teve desempenho ruim ao detectar Shotgun Surgery.
- ▶ CBO foi considerada útil por onze indivíduos. No entanto, esses indivíduos atingiram apenas 15% da recall em média.
- ▶ Cinco indivíduos indicaram CDC como sendo útil e alcançaram 23% de recall em média.
- ▶ NCC foi considerada útil por quatro indivíduos que atingiram 42% da recordação em média.



Métricas reunidas em conjunto

- ▶ NCC obteve melhores resultados em termos de recall.
- ▶ Todos que usaram NCC, solo ou em combinação com outras métricas, pontuaram mais do que 30% de recall.
- ▶ No entanto, esses resultados não permitem concluir que essas métricas são boas para detectar Shotgun Surgery.



Métricas reunidas em conjunto

- ▶ Algumas métricas consideradas uteis para a identificação de Shotgun Surgery.



Table 10. Metrics Considered Useful for Shotgun Surgery

Metrics	CBO	CDC	NCC
Subjects who used this metric	S25-S29, S37, S39, S40, S42-S44	S30, S31, S33, S40, S43	S32, S35, S36, S37
Average of Recall	15%	23%	42%



Métricas reunidas em conjunto

- ▶ CBO e LCOM foram considerados úteis para detectar God Class. Os indivíduos que utilizaram estas métricas alcançaram cerca de 67% de recall em média.
- ▶ WMC, LOC e CDLOC alcançaram recall acima de 85%.
- ▶ Estes resultados sugerem que métricas de tamanho, como LOC, WMC e CDLOC são bons indicadores de God Class.
- ▶ A utilização conjunta de WMC com LOC parece ser a melhor combinação de métricas.



Métricas reunidas em conjunto

- ▶ Métricas consideradas uteis para a detecção de God Class.

Table 11. Metrics Considered Useful for God Class

Metrics	CBO	LCOM	WMC	LOC	CDLOC
Subjects who used	S46, S47, S51, S54	S45, S51, S53, S54	S47, S52, S53	S52, S53, S54	S48, S49, S53
Average of recall	67%	67%	89%	100%	89%





Ameaças a validade e Trabalhos relacionados



Ameaças a validade e Trabalhos relacionados

- ▶ As conclusões obtidas são restritas para as métricas envolvidas, os bad smells descritos e os softwares escolhidos.



Ameaças a validade e Trabalhos relacionados

- ▶ O artigo descreve um quadro experimental.
- ▶ As métricas de interesses dependem do quão preciso foi o mapeamento dos interesses relacionados ao sistema alvo.
- ▶ A fim de atenuar esta ameaça, foram utilizados mapeamentos produzidos pelos desenvolvedores originais.



Ameaças a validade e Trabalhos relacionados

- ▶ Outros pesquisadores realizaram três experimentos para avaliar a utilidade de métricas de interesses para identificar módulos sujeitos a erros.
- ▶ Seus experimentos avaliaram seis métricas; dois deles são também utilizados no experimento em questão, CDC e CDO.



Ameaças a validade e Trabalhos relacionados

- ▶ Eles encontraram de moderada a forte correlação entre a métricas de interesses e defeitos de módulos.
- ▶ O objetivo do estudo é diferente, devido ao fato de não estar focados em análise de propensão a erros.
- ▶ Esse experimento complementa e amplia as descobertas do outro grupo de pesquisa, pois observa que a métricas também poderiam servir como indicadores confiáveis de bad smells.



Conclusões e trabalhos futuros

Conclusões e trabalhos futuros

- ▶ A avaliação da capacidade de manutenção de software é em grande parte dependente da disponibilidade de métricas que detectam com precisão bad smells.
- ▶ Métricas de interesses são cada vez mais utilizadas em estudos empíricos.



Conclusões e trabalhos futuros

- ▶ Os resultados revelaram que métricas de interesses são claramente úteis para detectar Divergent Change e God Class.
- ▶ Mostrou também que a experiência dos desenvolvedores não tem influência sobre a eficácia na detecção de bad smells e sim o tempo gasto pelos mesmos.



Conclusões e trabalhos futuros

- ▶ Além disso, observou-se que a eficácia de cada conjunto de métrica depende em grande medida da adequação de cada métrica para quantificar uma propriedade explicitamente mencionada na definição do bad smell.



Conclusões e trabalhos futuros

- ▶ Planeja-se realizar mais estudos empíricos para analisar o papel das métricas de interesses em diferentes níveis de abstração.



FIM

