

ПОБУДОВА ЗАЛЕЖНОСТЕЙ МІЖ ПРЯМИМИ ТА НЕПРЯМИМИ МЕТРИКАМИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Статтю присвячено опису застосування предметно-орієнтованого методу побудови залежностей між метриками програмного забезпечення для визначення залежностей між прямими та непрямыми метриками. Розглянуто побудову залежності непрямої метрики «розуміння», яка відповідає аналогічній властивості програмного забезпечення від прямих метрик. Проведено розрахунок ступеня впливу прямої метрики на метрику «розуміння», побудовано функцію залежностей між найбільш впливовою прямою метрикою та непрямою метрикою.

Ключові слова: метрики, прямі метрики, непрямі метрики, побудова залежностей.

Вступ

Все більшої актуальності набувають питання дослідження програмного забезпечення. Вирішенням цих питань займається емпірична інженерія програмного забезпечення [1].

Ключовим методом дослідження програмного забезпечення є вимірювання. Головний інструмент, який при цьому застосовується, – метрики. За допомогою метрик оцінюють властивості складових розробки програмного забезпечення (продуктів, процесів). Існують прямі та непрямі метрики [2]. Прямі метрики можна отримати при вимірюванні. На основі прямих метрик важко проводити оцінку властивостей програмного забезпечення. Для цього використовують непрямі метрики. Їх не можна виміряти, але можна отримати аналітичним чи емпіричним шляхом на основі прямих метрик. Головна задача, яка при цьому постає, – як визначати прямі метрики, від яких залежать непрямі, та яка залежність буде існувати між ними [3].

Огляд останніх досліджень

Для побудови залежностей використовують два підходи – статистичний аналіз [4] та нейронні мережі [5].

Вирішення задачі по визначенню залежностей між метриками за допомогою нейромереж потребує не лише знань про значення метрик (вхідних величин), а вимагає проведення навчання нейромережі [5]. Навчання нейромережі проводиться на уже отриманих раніше даних. Оскільки питання визначення залежностей між метриками вивчали тільки в реальному контексті (досліджувалися деякі конкретні метрики), а визначення залежностей між різними прямими та непрямыми метриками в цілому не досліджували, сформулювати правила для навчання ней-

ромережі важко. Для вирішення задачі по визначенню залежностей між метриками за допомогою статистичного аналізу достатньо знати тільки значення метрик, а залежності будуються статистичними методами [4]. Тому в цьому випадку необхідно застосовувати статистичний аналіз.

Статистичний аналіз проводиться за допомогою відповідних математичних програмних середовищ, до яких належать MatLab, MathCad, Maple, Mathematica, MS Excel. Також можна використати статистичні програмні середовища загального призначення: Statistica, SPSS, SAS, Systat, Minitab, Statgraphics, або програмні середовища спеціального призначення: SYSTAT, S-plus, STATA, PRISM, STADIA, Олимп, Класс-Мастер, Статистик-Консультант. Досвід їхнього використання свідчить, що математичні програмні середовища та статистичні програмні середовища загального призначення для вирішення поставленої задачі потребують додаткового програмування з використанням статистичних алгоритмів. Середовища для емпіричних досліджень в програмному забезпеченні немає.

Отже, для побудови залежностей між прямими та непрямыми метриками програмного забезпечення потрібне середовище та метод, який буде використовуватися в середовищі.

Для вирішення головної задачі було розроблено предметно-орієнтований метод побудови залежностей між метриками програмного забезпечення. На основі методу було реалізовано за-сіб [6].

У статті розглядається застосування предметно-орієнтованого методу побудови залежностей між метриками програмного забезпечення на прикладі побудови залежності між непрямою метрикою «розуміння» та рядом прямих метрик.

Мета статті

Основною метою статті є дослідження практичного застосування розробленого методу побудови залежностей між метриками через побудову залежності непрямої метрики «розуміння» від прямих метрик та визначення прямих метрик, від яких залежить непряма метрика.

Застосування предметно-орієнтованого методу

Суть предметно-орієнтованого методу побудови залежностей між метриками програмного забезпечення полягає в тому, що побудова залежностей відбувається статистичними методами з урахуванням високої точності вимірювань програмного забезпечення без похибок та гіперболічною спадною залежністю між значеннями метрик і кількістю виміряних програм.

Загальну схему методу представлено на мал. 1. На його основі будуються залежності між непрямою метрикою «розуміння» та низкою прямих метрик. Метою дослідження є визначення прямих метрик, від яких залежить досліджувана непряма метрика, та вид залежності непрямої метрики від прямих.

Отримання метрик для дослідження. Прямі метрики для досліджень було отримано шляхом вимірювань програмного забезпечення. Вимірювання прямих метрик програмного забезпечення проходило під час експерименту з визначення законів розподілів метрик [6]. Було виміряно 50 прямих метрик.

Непряма метрика «розуміння» є деякою функцією від однієї чи декількох прямих метрик. Оскільки для непрямої метрики «розуміння» поки не визначено функції, на основі якої можна було б розрахувати метрику, вона була отримана шляхом експертного оцінювання програмного забезпечення [7]. Оцінку експерт міг вибирати із запропонованих: «погано», «задовільно», «добре», «відмінно». Експертне оцінювання проходило тільки для того програмного забезпечення, яке було до цього виміряне.

Об'єм вибірки. Важливим критерієм при отриманні метрик є їхня кількість. Необхідний обсяг метрик прямо впливає на результати побудови залежностей. Мінімальна кількість метрик, які необхідні для побудови залежностей між метриками, була визначена експериментальним шляхом.

Для цього було проведено дослідження статистичних характеристик низки метрик при збільшенні вибірки. Дослідження проводилося в зворотному порядку. З усього обсягу було розраховано статистичні характеристики для декількох вибірок. Ці вибірки були отримані з загальної кількості метрик шляхом зменшення ви-

бірки на 10 %. Наприклад, для метрики WMC було отримано 5000 значень. Для цієї вибірки було розраховано статистичні характеристики. Далі вибірку було зменшено на 10 %, і статистичні характеристики було пораховано для 4500 значень метрики. Далі вибірку було зменшено ще на 10 %, і статистичні характеристики було пораховано для цієї вибірки. Таким чином вибірка зменшувалася до половини початкової вибірки.

З'ясувалося, що для визначення кількості значень метрик достатньо провести декілька контрольних розрахунків статистичних характеристик, щоб виявити, чи потрібно збільшувати вибірку. Крім цього, було визначено, що найсуттєвішим показником є медіана. Це пов'язано з відсутністю нормального розподілу серед метрик. А також стало зрозуміло, що більше 4-х вибірок брати із запропонованої недоцільно, так як вони в повній мірі відображають загальні закономірності. Ключовим показником тут є точність. Для різної заданої точності вибірка може змінюватися. Тому цей параметр повинен задавати сам дослідник.

Якщо співвідношення між сусідніми замірами не менше, ніж задана точність, то цю вибірку можна вважати повною для побудови залежностей. Тобто збільшувати кількість вимірів не потрібно. Описане співвідношення має вигляд:

$$\gamma \leq \frac{\tilde{x}_1}{\tilde{x}_2} \approx \frac{\tilde{x}_2}{\tilde{x}_3} \approx \frac{\tilde{x}_3}{\tilde{x}_4}$$

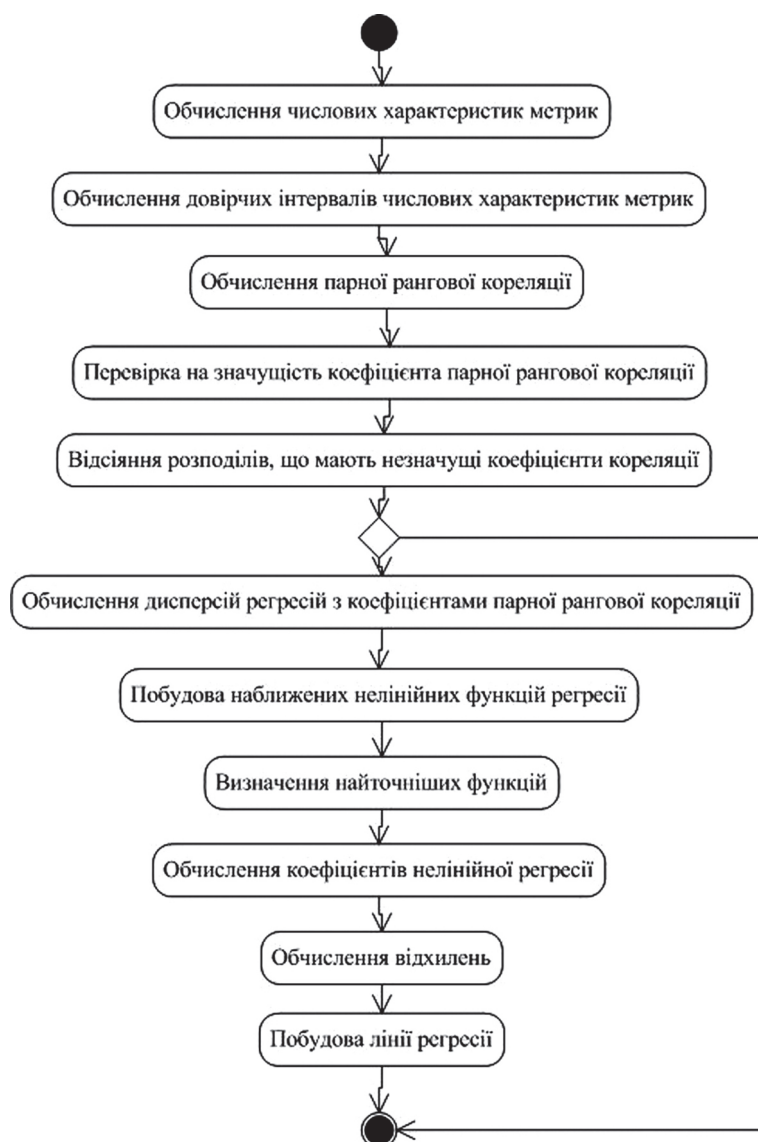
де γ – задана точність, $\tilde{x}_i$ – медіани кожної з вибірок.

Відповідно, якщо медіани відрізняються для кожної з вибірок на задану точність, вимірний обсяг метрик приймається для подальших розрахунків. Якщо ні – пропонується збільшити об'єм вибірки.

Підтвердженням отриманих результатів слугують статистичні характеристики деяких вимірювань й досліджених метрик [8].

Обчислення числових характеристик метрик. Відповідно до розробленого методу, спочатку було отримано числові характеристики прямих метрик та непрямої «розуміння» (RA): математичне сподівання, середнє квадратичне відхилення, коефіцієнти експесу й асиметрії. Деякі значення наведено в табл. 1.

Обчислення кореляції та відсіювання незалежних метрик. Обчислення парної рангової кореляції відбувається шляхом розрахунків коефіцієнтів Спірмена та Кендала [9]. Деякі розраховані коефіцієнти наведено в табл. 2. Крім цього, в таблиці подано значущість отриманих коефіцієнтів, яка говорить про наявність чи відсутність залежності.



Мал. 1. Предметно-орієнтований метод побудови залежностей між метриками програмного забезпечення

Таблиця 1. Початковий та центральний моменти метрик

Познач. метрики	Мат. сподівання	Середнє квадрат. відхилення	Коеф. асиметрії	Коеф. ексцесу	Пояснення метрики
AVRPF	58.8417	17.9308	-0.429	2.4917	Середнє значення виключень в модулі
TM	2.5140	2.4356	1.7780	6.0558	Загальна кількість функцій
TLOC	170.9636	142.7548	1.2357	3.8945	Загальне число непорожніх рядків, включаючи коментарі та рядки заголовків
TNCSS	121.5091	111.8934	1.3023	4.0540	Загальне число вихідних конструкцій. Не включає порожні рядки, рядки коментарю, чи заготовочні рядки
TVG	16.1000	19.7032	2.1608	7.5764	Загальна цикломатична складність у всіх модулях
MFO	8.9912	9.3785	1.5897	5.7260	Максимальне число викликів інших функцій в модулі
MFI	1.0000	1.7008	2.4493	10.2992	Максимальне число викликів модуля
AVGFI	0.3838	0.5715	1.5512	5.0540	Загальне число викликів інших функцій в модулях
‘RA’	2.9279	1.3823	-0.2189	0.6029	«розуміння»

Таблиця 2. Коефіцієнти кореляції для непрямої метрики «розуміння» та прямих метрик

Назва метрики	Коефіцієнт кореляції Кендала	Коефіцієнт кореляції Спірмена	Значущість
AVRPF	0.0912	0.1368	+
TM	0.3986	0.5979	+
TLOC	0.0693	0.1040	+
TNCSS	0.0263	0.0395	-
TVG	-0.2482	-0.3724	+
MFO	0.0161	0.0241	-
MFI	0.1941	0.2912	+
AVGFI	0.1327	0.1991	+

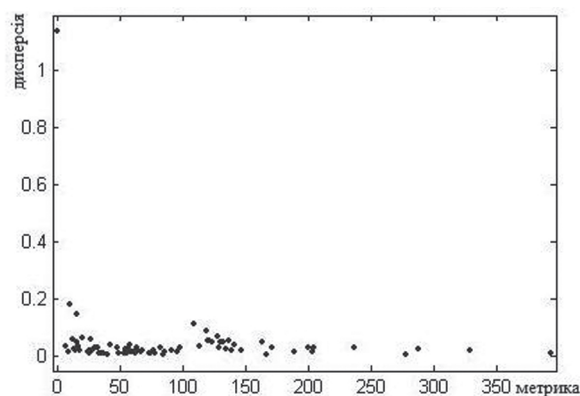
Як видно з табл. 2, непряма метрика «розуміння» залежить від більшості прямих метрик, про що свідчить значущість. Із загальної кількості 50 прямих метрик було відсіяно 5. Для інших 45 метрик будувалися регресії.

Побудова регресій. Побудова регресії проводилася для кожної прямої метрики окремо. Далі буде наведено приклад побудови регресії для непрямої метрики «розуміння» та прямої метрики «кількість непустих рядків» (TLOC).

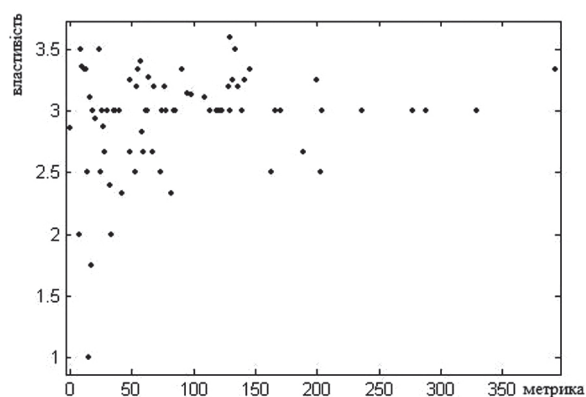
Спочатку проводиться розрахунок дисперсій. Дисперсії представлені у графічному виді на мал. 2. Як бачимо, значних відхилень по кожному з вимірювань немає, тому наступним етапом буде побудова регресії. Перед побудовою регресії було визначено кореляційне поле для метрик (мал. 3). Кореляційне поле має складну конфігурацію, що свідчить про нелінійну функціональну залежність між метриками. Далі, згідно з [9], проводимо побудову декількох наближених регресій, розраховуємо їхні коефіцієнти та вибираємо найбільш точну.

Побудовані типові функції регресії та їхні відхилення наведено в табл. 3. У функціях: x – математичне сподівання, умовні математичні

сподівання експертних оцінок, a , b , c – коефіцієнти.



Мал. 2. Значення дисперсій



Мал. 3. Кореляційне поле метрик

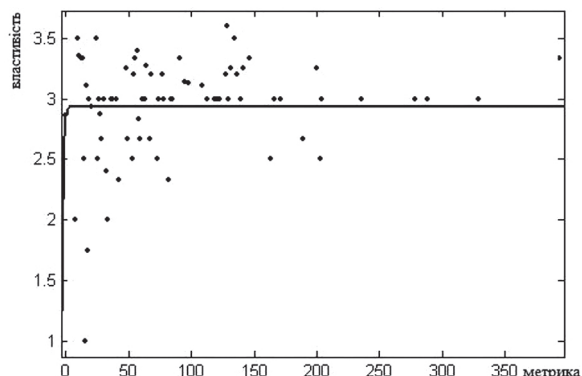
За основу для вибору оптимальної функції слугує відхилення.

Проаналізувавши значення відхилень можна зробити висновок, що найменше – 0.2906 для функції $y=1/(a \exp(-x)+b)$ (мал. 4). Тобто це функція, яка найточніше виражає залежність

Таблиця 3. Коефіцієнти та відхилення функцій регресії

№ п/п	Функція	Коефіцієнт а	Коефіцієнт b	Коефіцієнт c	Відхилення
1	$y = ax+b$	0.001161	2.83	0	0.2917
2	$y=1/(ax+b)$	-0.0001219	0.352	0	0.2913
3	$y=x/(ax+b)$	0.329	0.4844	0	0.3142
4	$y = a \ln x+b$	1.252	1.127	0	5.7184
5	$y=b \exp(ax)$	0.0003761	2.836	0	0.2914
6	$y = ax^3+b$	0.996e-009	2.918	0	0.2912
7	$y=c+b \exp(ax)$	-11.4	-0.07609	2.938	0.2909
8	$y=ax^2+bx+c$	-6.025e-006	0.002988	2.75	0.3016
9	$y=1/(ax^2+bx+c)$	-0.006006	0.5697	0.376	2.8139
10	$y=x/(ax^2+bx+c)$	-5.971e-005	0.3377	0.3557	0.3210
11	$y=a/(1/x+b)$	2.064	0.6792	0	0.3142
12	$y=1/(a \exp(-x)+b)$	0.009149	0.3403	0	0.2906

між метрикою «кількість непустих рядків» та «розуміння».



Мал. 4. Функція регресії між метриками

Висновки

У результаті проведеної роботи було виведено закономірність при визначенні обсягу вибірки метрики. З'ясувалося, що цю величину мож-

на розрахувати на основі перевірки математичного сподівання декількох менших вибірок.

Було визначено, що більшість метрик не мають нормального закону розподілу. Це пов'язано з особливостями даних програмного забезпечення, оскільки існує великий обсяг програмного коду з невеликими значеннями метрик. Підтвердження цього є у [8].

Проведені дослідження свідчать про те, що на непряму метрику «розуміння» впливає багато прямих метрик. Серед усіх прямих метрик було вибрано одну, для якої й визначалася функція залежності. Як з'ясувалося, залежності між прямими метриками та непрямою метрикою «розуміння» мають нелінійний характер, про що свідчать коефіцієнти кореляції та побудована регресія між прямою метрикою «кількість непустих рядків» та непрямою «розуміння».

Результати досліджень для ряду об'єктно-орієнтованих метрик підтверджуються в [8].

Подальшу роботу буде направлено на визначення закономірностей між ключовими прямими та непрямыми метриками.

1. Torii K. Ginger 2: An Environment for Computer-Aided Empirical Software Engineering / Koji Torii, Ken-ichi Matsumoto, Kumiyo Nakakoji, Yoshiro Takada, Kaduyuki Shima // IEEE Transactions on Software Engineering. – 1999, July/August. – Vol. 25, No. 4. – P. 475–486.
2. Fenton E. Norman. Software Metrics : A Rigorous and Practical Approach / Norman E. Fenton, Shari Lawrence Pfleeger. – Cambridge University Press, 1996. – 638 p.
3. Shull F. Guide to Advanced Empirical Software Engineering / Forrest Shull, Janice Singer, Dag I. K. Sjøberg. – Springer-Verlag New York, Inc., 2008. – 394 p.
4. Вентцель Е. С. Теория вероятностей : Учеб. для вузов / Вентцель Е. С. – 7-е изд. стер. – М. : Высш. шк., 2001. – 575 с. : ил.
5. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. – Пер. с англ. – М. : Мир, 1992. – 118 с.
6. Дишлевий О. П. Пакет статистичного аналізу для емпіричної інженерії програмного забезпечення / О. П. Дишлевий // Наука і молодь. Прикладна серія. Збірник наукових праць. – 2009. – № 9. – С. 104–108.
7. Хоменко В. А., Дышлевый А. П. Метод экспертного оценивания свойств повторно используемых компонентов программного обеспечения // Матеріали VIII Міжнародної науково-технічної конференції «Авіа-2007». Т. 1. – К. : НАУ, 2007. – С. 13.169–13.172.
8. Lanza M. Object-oriented metrics in practice: Using software metrics to characterize, evaluate, and improve the design of object-oriented systems / Michele Lanza, Radu Marinescu. – Springer-Verlag Berlin Limited, 2006. – 205 p.
9. Бабак В. П. Статистична обробка даних / Бабак В. П., Білецький А. Я., Приставка О. П., Приставка П. О. : Монографія. – К. : «МІВВІЦ», 2001. – 388 с.

O. Dyshlevyy

BUILDING DEPENDENCES BETWEEN DIRECT AND INDIRECT SOFTWARE METRICS

The article is dedicated to describing the application of subject-oriented method of building dependences between software metrics for determine the dependencies between direct and indirect metrics. It was considered building depends indirect metrics «understanding», which corresponds to similar property of software from a number of direct metrics. Level of impact on the metric «understanding» is calculated for each direct metric, there is determined the most influential metric and there is built functions of dependencies between the most powerful direct and indirect metric.

Keywords: software metrics, direct metrics, indirect metrics, building dependences.