

SUN'IY INTELLEKT (LLM) YORDAMIDA EKVIVALENT MUTANTLARNI FILTRLASH VA SHAXSIYLASHTIRILGAN TESKARI ALOQA GENERATSIYALASH

Xamrayev Xurshidbek Komiljon o'g'li

University of Business and Science stajor – o'qituvchisi

xamrayevbek96@gmail.com

Annotatsiya: Mazkur maqolada mutatsion testlashning ikki asosiy to'sig'i — ekvivalent mutantlarni aniqlashning yechib bo'lmamas muammosi va yuqori hisoblash xarajatlarini katta til modellari (LLM) yordamida hal etish imkoniyati ko'rib chiqilgan. LLM-modulning ikki vazifasi — ekvivalent mutantlarni filtrlash va talaba profiliga moslab tushuntiruvchi teskari aloqa generatsiyalash loyihalangan. Tahlil shuni ko'rsatdiki, LLM lar dastur semantikasini chuqurroq qamrab ekvivalent mutant aniqlashda istiqbolli natija beradi, biroq “gallyutsinatsiya” va noto'g'ri takliflar xavfi saqlanib qoladi. Shu bois inson nazorati (human-in-the-loop) zarurligi asoslangan.

Kalit so'zlar: sun'iy intellekt, katta til modeli, LLM, ekvivalent mutant, mutatsion testlash, shaxsiylashtirilgan teskari aloqa, gallyutsinatsiya, human-in-the-loop.

Аннотация: В статье рассматривается возможность решения двух основных препятствий мутационного тестирования — неразрешимой проблемы выявления эквивалентных мутантов и высоких вычислительных затрат — с помощью больших языковых моделей (LLM). Спроектированы две функции LLM-модуля: фильтрация эквивалентных мутантов и генерация объясняющей обратной связи с учётом профиля студента. Анализ показал, что LLM, глубже охватывая семантику программы, дают

перспективные результаты в выявлении эквивалентных мутантов, однако сохраняется риск «галлюцинаций» и неверных рекомендаций. В связи с этим обоснована необходимость контроля со стороны человека (human-in-the-loop).

Ключевые слова: *искусственный интеллект, большая языковая модель, LLM, эквивалентный мутант, мутационное тестирование, персонализированная обратная связь, галлюцинация, human-in-the-loop.*

Abstract: *This article examines the possibility of addressing two main obstacles of mutation testing — the undecidable problem of detecting equivalent mutants and high computational costs — using large language models (LLMs). Two functions of an LLM module are designed: filtering equivalent mutants and generating explanatory feedback tailored to the student’s profile. The analysis showed that LLMs, by more deeply capturing program semantics, yield promising results in detecting equivalent mutants; however, the risk of “hallucinations” and incorrect suggestions remains. Accordingly, the need for human oversight (human-in-the-loop) is substantiated.*

Keywords: *artificial intelligence, large language model, LLM, equivalent mutant, mutation testing, personalized feedback, hallucination, human-in-the-loop.*

Kirish (Introduction)

Mutatsion testlashning ta’limda qo’llanilishini cheklovchi ikki asosiy to’siq mavjud. Birinchisi — ekvivalent mutantlar muammosi: asl dastur bilan barcha kirishlarda bir xil ishlaydigan, ammo sintaktik jihatdan farqli mutantlar; bu muammo uch o’n yildan beri yechib bo’lmay (undecidable) deb tan olingan va real loyihalarda ularning ulushi 4%–39% ni tashkil etadi (Tian et al., 2024). Ikkinchisi — har bir mutantni

alohida ishga tushirish tufayli yuzaga keladigan yuqori hisoblash xarajatlari (Jia & Harman, 2011).

Katta til modellari (LLM) kod bilan bog'liq masalalarda dastur semantikasini chuqur qamrab, bu muammolarni hal etishda istiqbolli vosita sifatida namoyon bo'lmoqda. Tadqiqotning maqsadi — LLM yordamida ekvivalent mutantlarni filtrlash va talabaga shaxsiylashtirilgan teskari aloqa beruvchi modulni loyihalash va asoslashdan iborat.

Metodlar (Methods)

Tadqiqotda LLM ning ekvivalent mutant aniqlashdagi samaradorligi bo'yicha empirik ishlar tahlil qilindi (Tian et al., 2024); LLM-modulning ikki vazifasi loyihalandi — (1) mutant va asl kod juftligini semantik taqqoslab ekvivalent mutantlarni filtrlash, (2) talabaning kuchli va zaif tomonlari profiliga moslab tushuntiruvchi teskari aloqa generatsiyalash; generatsiya sifati ekspert baholash bilan tekshirildi (Roest, Keuning & Jeuring, 2024).

Natijalar (Results)

Tahlil va loyihalash quyidagi natijalarni berdi:

1) Filtrlash: LLM lar dastur semantikasini chuqurroq qamrab, ekvivalent mutantlarni filtrlashda an'anaviy usullarga nisbatan istiqbolli natija berdi; bu mutation score ning aniqligini va baholashning ishonchliligini oshiradi.

2) Shaxsiylashtirilgan feedback: GPT-4 tipidagi modellar asosida talaba xatolarini samarali ko'rsatuvchi, yechimni oshkor qilmaydigan shaxsiylashtirilgan teskari aloqa generatsiyalash mumkinligi tasdiqlandi.

3) Xarajatlarni kamaytirish: ekvivalent mutantlarni avtomatik filtrlash inson tomonidan qo'lda tekshirish hajmini kamaytirib, hisoblash va mehnat xarajatlarini qisqartiradi.

Shu bilan birga, LLM tomonidan generatsiya qilingan teskari aloqada noto'g'ri takliflar va "gallyutsinatsiya" (mavjud bo'lmagan muammoni ko'rsatish) xavfi aniqlandi (Roest, Keuning & Jeurig, 2024).

Muhokama (Discussion)

LLM-modul dissertatsiyaning integratsion komponenti bo'lib, mutatsion testlashning ikki tarixiy cheklovini yumshatadi: baholash xarajatini kamaytiradi va teskari aloqani shaxsiylashtiradi. Bu UIIP platformasida mutatsion baholashni amaliy va miqyoslanuvchan qiladi.

Asosiy cheklov — gallyutsinatsiya xavfi: LLM noto'g'ri yoki mavjud bo'lmagan xatoni ko'rsatishi talabani chalg'itishi mumkin. Shu sababli teskari aloqa to'liq avtomatlashtirilmasligi, balki o'qituvchi nazorati (human-in-the-loop) bilan birga qo'llanilishi zarur. Shuningdek, LLM natijalari modelga va so'rov (prompt) sifatiga bog'liq bo'lib, izchillikni ta'minlash qo'shimcha sozlashni talab qiladi.

Xulosa (Conclusion)

Sun'iy intellekt (LLM) mutatsion testlashning ekvivalent mutantlar va yuqori hisoblash xarajatlari muammolarini yumshatib, uni ta'limda qo'llashni amaliy qiladi. LLM-modul ekvivalent mutantlarni filtrlash va shaxsiylashtirilgan teskari aloqa generatsiyalash orqali baholash aniqligi va pedagogik qiymatini oshiradi. Gallyutsinatsiya xavfi tufayli inson nazorati saqlanib qolishi va prompt sifatini ta'minlash keyingi tadqiqot vazifasi bo'lib qoladi.

Foydalanilgan adabiyotlar (References)

1. Jia, Y., & Harman, M. (2011). An Analysis and Survey of the Development of Mutation Testing. *IEEE Transactions on Software Engineering*, 37(5), 649–678.
2. Roest, L., Keuning, H., & Jeuring, J. (2024). Evaluating the Application of Large Language Models to Generate Feedback in Programming Education. *arXiv:2403.09744*.
3. Tian, Z., Shu, J., Zhan, Y., et al. (2024). Large Language Models for Equivalent Mutant Detection: How Far Are We? *Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA)*. <https://doi.org/10.1145/3650212.3680395>
4. Open Source Language Models Can Provide Feedback: Evaluating LLMs' Ability to Help Students Using GPT-4-As-A-Judge. (2024). *Proceedings of ITiCSE 2024*. <https://doi.org/10.1145/3649217.3653612>
5. Messer, M., Brown, N. C. C., Kölling, M., & Shi, M. (2024). Automated Grading and Feedback Tools for Programming Education: A Systematic Review. *ACM Transactions on Computing Education*, 24(1), 1–43. <https://doi.org/10.1145/3636515>