

ОГЛЯД ТА АНАЛІЗ МЕТРИК ОЦІНЮВАННЯ ЯКОСТІ МАШИННОГО ПЕРЕКЛАДУ

© Яковина В., Масюкевич В., 2013

Наведено огляд основних підходів до машинного перекладу, проаналізовано існуючі метрики оцінювання якості машинного перекладу, виявлено їх переваги та недоліки.

Ключові слова: машинний переклад, оцінювання якості машинного перекладу, еталонний переклад.

This paper is devoted to the overview of main approaches of machine translation, to analysis of existing metrics for machine translation quality estimation. The advantages and disadvantages of these metrics have been revealed.

Key words: machine translation, machine translation quality estimation, reference translation.

Вступ

У сучасному суспільстві посилились процеси глобалізації та інформатизації, що привело до збільшення й спрощення інтернаціональних зв'язків і контактів. Розширення мережі Інтернет створює умови для інтернаціонального спілкування. Єдиною перешкодою є мовний бар'єр, подолати який за допомогою традиційного перекладу стає все складніше, оскільки обсяг перекладів постійно зростає, і навіть збільшення кількості перекладачів не може повністю задовольнити потреби в перекладах. Комп'ютеризація світу, удосконалення й поява нових версій програмних продуктів, спроби зробити програмне забезпечення максимально сумісним із низкою приладів та програм, інші нові досягнення індустрії та технічного прогресу виводять на ринок нові вироби, які мають бути описані й перекладені якомога більшою кількістю мов для донесення до кінцевого користувача і збільшення обсягів продажу. Машинний переклад є одним із засобів подолання мовного бар'єра. Він відіграє важливу роль в обміні науковими дослідженнями та поширенні інформації.

Системи машинного перекладу

Існує три основних підходи до машинного перекладу: машинний переклад на основі правил, статистичний і гібридний машинний переклади [1].

Машинний переклад на основі правил характеризується використанням і ручним створенням лінгвістичних правил. Ефективність систем машинного перекладу на основі правил визначається якістю двомовних словників та точністю заданих правил, а їх створення потребує довготривалої роботи [1].

Перші системи машинного перекладу створювались для конкретних пар мов і ґрунтувались на складних процесах моделювання мови, основу яких становили методи аналізу, трансферу, синтезу й інтерлінгви. Системи машинного перекладу першого покоління працювали за методом прямої заміни слів мови оригіналу словами мови перекладу. Системи другого покоління аналізували структури мови оригіналу, а потім на основі трансферу синтезували їх в еквівалентні структури мови оригіналу [1]. Третім поколінням були системи машинного перекладу на основі формальної мови-посередника – інтерлінгви. Ця концепція передбачала перетворення слів на мову-посередника, яка є універсальною мовою, створеною для системи, незалежною від залучених у процес перекладу мов. Цей підхід використовує два методи: аналізу і синтезу [2, 3].

Системи машинного перекладу на основі правил ґрунтуються на різних рівнях лінгвістичного опрацювання мовної пари [1]:

1. Морфологічному: лематизація лексичних одиниць, пошук лексичних одиниць у словнику, аналіз морфем, розпізнавання контекстного граматичного класу лексичних одиниць, відмінків, флексій тощо.

2. Синтаксичному: розпізнавання типів синтаксичних структур, реляційних зв'язків між окремими елементами синтаксичної структури тощо.

3. Семантичному: виокремлення лексичного значення багатозначних лексичних одиниць та афіксів, визначення їхньої семантичної функції, синтез їхньої синтаксичної однозначності на основі семантичного аналізу [1].

Системи машинного перекладу на основі правил не потребують доступу до баз паралельних текстів, їх можна налаштовувати, що поліпшує якість перекладу спеціалізованих текстів [4]. Системи на основі правил можуть мати справу з багатьма мовними явищами і зручні в супроводі. Проте винятки в граматиці додають труднощів [2], що потребує розроблення нових алгоритмів і покращення раніше створених [4].

Основним недоліком систем машинного перекладу на підставі правил є те, що для підвищення якості цих систем потрібно покращувати раніше створені й розробляти нові алгоритми, що є дуже ресурсоємним, тому сучасні системи машинного перекладу ґрунтуються переважно на статистичних або гібридних методах. Основою статистичних систем машинного перекладу є автоматична екстракція схожих сегментів мовних пар з двомовних повнотекстових корпусів, які нараховують мільярди слововживань. Гібридні системи машинного перекладу [5] створюються на базі наявних систем машинного перекладу на основі правил з додаванням до них статистичних методів. Отже, навчання статистичних і гібридних систем машинного перекладу базується на двомовних корпусах текстів і не потребує глибокого й складного контрастивного лінгвістичного аналізу. Це дало змогу знизити вартість систем машинного перекладу, що привело до їх стрімкого поширення [1].

Статистичні технології не потребують спеціальних лінгвістичних алгоритмів. Недоліком статистичного перекладу є те, що в ньому не враховано граматичних правил. Іноді у разі використання статистичного перекладача замість зв'язного тексту отримують ніяк не погоджені слова. Іншою проблемою є те, що при використанні цієї системи необхідна дуже велика кількість паралельних текстів. Чим більша база паралельних текстів, тим більша ймовірність знайти необхідний збіг [4].

Лінгвістичні особливості, двозначність, відсутність універсальності граматики і лексики – причини того, що системи машинного перекладу не досягають 100 % точності перекладу [2]. Існує ряд типових помилок, що виникають у ході машинного перекладу. До граматичних проблем перекладу належить неправильне розпізнавання типів зв'язків між членами речення, що, своєю чергою, порушує порядок членів речення, заміна одного члена речення на інший, системі досить складно розпізнавати родову та відмінкову форми. До лексичних проблем належить передусім неправильний принцип відбору словникових відповідників, повна або часткова неперекладність, неправильний переклад термінології [6].

Оцінювання якості перекладу

Оцінювання якості машинного перекладу є важливою сферою дослідження для визначення ефективності та для оптимізації роботи існуючих систем машинного перекладу.

Для оцінювання машинного перекладу часто використовують оцінювання перекладу людиною. Дві з найпоширеніших метрик оцінки людиною – плавність і адекватність [7, 8]. Плавність вимагає вільного володіння мовою перекладу, щоб робити висновок про те, чи вихідні дані системи є плавними (текст читається як написаний носієм мови), незалежно від того, чи зміст вихідного тексту є точним перекладом вхідних слів.

Адекватність не враховує рівня плавності вихідних даних системи, натомість настільки, наскільки це можливо, вимірює, чи необхідну інформацію у вхідному тексті можна отримати з

вихідного тесту. Вимоги до анотатора щодо адекватності суворіші, ніж для плавності, оскільки анотатор повинен бути двомовним для того, щоб робити висновки про те, чи інформація зберігається при перекладі. На практиці анотатор, який вільно володіє тільки мовою, на яку перекладається, також може анотувати адекватність за допомогою набору якісних перекладів вхідних даних, виконаних людиною [9, 10].

Автоматичні метрики оцінки машинного перекладу розроблено у зв'язку з високими витратами, відсутністю повторюваності, суб'єктивністю і повільністю оцінки перекладу з використанням суджень людини і бажанням застосувати автоматичне налаштування параметрів системи. Такі метрики загалом оцінюють якість вихідного перекладу машини, порівнянням вихідних даних системи машинного перекладу з набором еталонних перекладів [9]. Еталонний переклад – це переклад, який здійснив кваліфікований перекладач і який визнаний якісним.

Метрика WER (Word error rate) [9] визначається як відстань Левенштейна [11] між словами вихідних даних системи і словами еталонного перекладу, поділена на довжину еталонного перекладу. Відстань Левенштейна обчислюють, використовуючи динамічне програмування для знаходження оптимального збігу між вихідними даними машинного перекладу й еталонним перекладом. Кожне слово машинного перекладу прирівнюють до 1 чи 0 слів еталонного перекладу, і навпаки. Випадок, коли еталонне слово прирівнюється до 0, називається видаленням, тоді як прирівнювання слова машинного перекладу до 0 називається вставкою. Якщо слова машинного перекладу й еталонного перекладу збігаються, то це називається заміною. WER – це сума замін (S), вставок (I) і видалень (D), поділена на кількість слів у еталонному перекладі (N) [9]:

$$WER = \frac{S + I + D}{N}.$$

Недоліком метрики WER є те, що цей показник не може адекватно поєднати знання з декількох еталонних перекладів, а також не може змоделювати перестановку слів чи фраз у перекладі.

Метрика MWER (multi-reference WER) [12] – застосування метрики WER до більш ніж одного еталонного перекладу – визначається мінімальним значенням WER між машинним перекладом і кожним еталонним перекладом.

Метрика PER (Position-independent error rate) [13] – це спроба усунення недоліку метрики WER стосовно перестановки порядку слів. Еталонний і машинний переклад трактуються як набір слів, так, що слово з еталонного перекладу може співвідноситись з словом машинного перекладу незалежно від його позиції. Тому PER машинного перекладу гарантовано буде меншим або дорівнює WER машинного перекладу. Недоліком є те, що правильний переклад не відрізняється від того, в якому перестановки здійснені в неправильному порядку.

Метрика BLEU (Bilingual Evaluation Understudy) [14] є поточним стандартом для автоматичної оцінки машинного перекладу. Ключовою характеристикою метрики BLEU є пряме використання кількох еталонних перекладів. Оцінка метрикою BLEU вихідних даних системи визначається підрахунком кількості n-грам, або послідовності слів, у вихідних даних системи, які трапляються в наборі еталонних перекладів. BLEU є метрикою, орієнтованою на точність, тому що показує, яка частина перекладу системи є коректною, а не вимірює, чи еталонний переклад повністю відтворено у вихідному перекладі системи. BLEU можна було б обійти, продукуючи дуже короткі вихідні дані системи, що складаються тільки з дуже коректних n-грам, якби не використання штрафного фактора стислості, який зменшує значення метрики BLEU, якщо вихідний переклад системи коротший за еталонний.

$$p_n = \frac{\sum_{c \in \{can\}} \sum_{n-gram \in c} Cnt_{clip}(n-gram)}{\sum_{c' \in \{can\}} \sum_{n-gram' \in c'} Cnt_{clip}(n-gram')}; \quad (1)$$

$$BP = f(x) = \begin{cases} 1, & \text{if } c > r \\ e^{\left(1 - \frac{r}{c}\right)}, & \text{if } c < r \end{cases} \quad (2)$$

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right). \quad (3)$$

Рівняння (1) показує обчислення коефіцієнта точності BLEU (p_n) для n -грам довжиною n , де Can – речення в корпусі тестів, Cnt (n -gram) – кількість входжень n -грами в машинному перекладі, і $Cnt_{clip}(n$ -gram) – максимальна кількість входжень n -грами в еталонному перекладі. Рівняння (2) показує розрахунок штрафного фактора стислості BLEU (BP), де c – довжина машинного перекладу і r – довжина еталонного перекладу. Ці умови об'єднуються, як показано в рівнянні (3), для розрахунку загального показника BLEU, де N , як правило, дорівнює 4, а w_n зазвичай встановлюють $1/N$.

Недоліком метрики BLEU є відсутність відтворення у формулюванні. BLEU створено і використовується для великих корпусів тестів. Значення BLEU для одного речення вважаються ненадійними. Також недоліком є неврахування збігу синонімів і неможливість відслідкувати велику кількість правильних порядків слів. Тобто семантична якість перекладу не визначається метрикою BLEU [9].

METEOR (Metric for evaluation of translation with explicit ordering) [15] – настроювана, орієнтована на збіг метрика, тоді як BLEU – метрика, орієнтована на точність. На відміну від BLEU, METEOR обчислює як точність, так і відтворення, поєднуючи їх для обчислення параметризованого гармонійного середнього:

$$F_{mean} = \frac{P \cdot R}{aP + (1-a)R}.$$

Метрика METEOR використовує декілька етапів встановлення відповідності між словами машинного перекладу й еталонного перекладу для того, щоб зіставити два рядки [9]:

1. Точне встановлення відповідності – визначаються рядки, що є ідентичними в еталонному і машинному перекладі.

2. Встановлення відповідності основ – проводиться стемінг (скорочення слова до основи) і визначаються слова з однаковим коренем в еталонному і машинному перекладі.

3. Встановлення відповідності синонімів – визначаються слова, що є синонімами відповідно до WordNet (семантичний словник для англійської мови, де слова розділено на групи синонімів та позначені семантичні зв'язки між ними) [16].

На кожному наступному етапі встановлення відповідності розглядаються тільки слова, для яких не встановлена відповідність на попередньому етапі.

На відміну від метрики BLEU, метрика METEOR не забраковує довші відповіді й враховує рівень лінгвістичного знання у вигляді встановлення відповідності коренів і синонімів, даючи змогу встановити еквівалентність між машинним перекладом і еталонним перекладом. METEOR не має однієї ключової характеристики BLEU: прямого використання декількох рекомендованих перекладів [9].

Метрика GTM (General text matcher) [17] намагається змоделювати рух фрази під час перекладу використовуючи максимальний розмір збігу для обчислення якості перекладу. Метрика знаходить найдовшу послідовність слів, які збігаються в машинному перекладі й рекомендованому перекладі. Розмір збігу визначається формулою

$$size(M) = \sqrt{\sum_{r \in M} length(r)^2},$$

де M – множина знайдених збігів, r – довжина еталонного перекладу. Точність і відтворення обчислюють як розмір збігів, поділений на довжину машинного чи рекомендованого перекладу відповідно. Показник GTM є середнім гармонійним значенням точності та відтворення.

Як і метрика BLEU, GTM не враховує лінгвістичного знання, і вважає, що слова в машинному і рекомендованому перекладах збігаються, якщо вони ідентичні. Якщо одне слово в перекладі неправильне, то значення GTM залежить від місця помилки в реченні. Помилка посередині речення найбільше впливає, а помилка в кінці речення має мінімальний вплив, зменшуючи довжину збігу на 1 [9].

Метрика TER (Translation edit rate) [18] розглядає зміну порядку розміщення фраз, чого не враховує метрика WER, дозволяючи переміщення блоків слів, яке називають зсувом. Проводиться

грубий пошук слів для зсуву, а також вводять подальші обмеження на слова для зсуву. Ці обмеження призначені для моделювання способу, в який редактор-людина вибирає слова для зсуву.

Обмеження зсуву потрібні для зменшення складності обчислення моделі й кращого моделювання якості перекладу. Метрика TER використовує такі обмеження:

1. Зсуви вибираються “жадібним” алгоритмом, який вибирає зсув, що найбільше зменшує WER між еталонним і машинним перекладом.

2. Послідовність зсунутих слів у машинному перекладі повинна точно відповідати послідовності слів у еталонному перекладі.

3. Слова, які зсуваються, повинні мати хоча б одну помилку, згідно з WER, перед зсувом. Це запобігає зсуванню слів, які збігаються.

4. Слова в еталонному перекладі, які мають бути зсунуті, мають також містити хоча б одну помилку. Це запобігає зсуванню слів, які вже збігаються.

Якщо TER використовується у випадку декількох еталонних перекладів, то машинний переклад оцінюється щодо кожного еталонного перекладу окремо. TER обчислюють за формулою [9]:

$$TER = \frac{SUB + INS + DEL + SHIFT}{\bar{N}},$$

де SUB, INS, DEL, SHIFT – кількість заміन, вставок, видалень, зсувів відповідно; \bar{N} – середня кількість слів еталонного перекладу. TER не використовує одночасно декілька еталонних перекладів, як BLEU, і не враховує зовнішні лінгвістичні знання, як METEOR [9].

Метрика CDER (Cover disjoint error rate) [19] використовує той факт, що кількість блоків у реченні дорівнює кількості інтервалів між блоками плюс один. Рух блоків можна еквівалентно описати як операцію довгого стрибка через інтервал між двома блоками. Значення довгого стрибка вважають сталим. Довгі стрибки поєднуються з класичними операціями редагування Левенштейна, а саме вставкою, видаленням, заміною і операцією збігу, яка має значення 0. Відстань довгого стрибка визначає мінімальну кількість операцій, потрібну для трансформації машинного перекладу в еталонний переклад. Як і відстань Левенштейна, відстань довгого стрибка можна зобразити, використовуючи сітку збігів [19].

Метрика HTER (Human-mediated translation error rate) [18] потребує використання одномовної людини, що створює цільові еталонні переклади, націлені на певні вихідні дані системи. Цільові еталонні переклади створюють, змінюючи вихідні дані системи з мінімальною кількістю редагувань, так щоб вони були плавними і зберігали зміст інших еталонних перекладів. Оскільки використовується мінімальна кількість редагувань для корекції вихідних даних машинного перекладу, створення цільових еталонних перекладів може вважатися вибором з набору можливих еталонних перекладів одного найближчого до машинного перекладу.

Цільові еталонні переклади можуть використовуватись для будь-яких метрик оцінювання якості, які використовують еталонні переклади [9]. Однак, оскільки цільові еталонні переклади створюються для кожних вихідних даних системи, їх не можна повторно використати для вихідних даних з різних систем і навіть вихідних даних з різних версій систем машинного перекладу.

Хоча метрика HTER побудована на судженнях людини, її найочевиднішим недоліком є те, що вона оцінює всі помилки однаково, тоді як деякі редагування, деякі помилки перекладу є тривіальними, а інші можуть бути згубними [9].

Метрика STM (Syntax tree based metric) [20]. Щоб правильно оцінити читабельність речення, для створення метрики STM використовують синтаксичні дерева на основі подібності дерев еталонного перекладу і машинного перекладу. Не можна сподіватись, що ціле синтаксичне дерево вихідних даних можна знайти в еталонному перекладі, проте цей підхід ґрунтується на частці піддерев, які є спільними для машинного та еталонного перекладу. Для кожного еталонного перекладу обчислюють частки піддерев з різною глибиною, а їх середнє арифметичне значення приймають як метрику на основі синтаксичних дерев:

$$STM = \frac{1}{D} \sum_{n=1}^D \frac{\sum_{t \in \text{subtrees}_n(\text{hyp})} \text{cont}_{\text{clip}}(t)}{\sum_{t \in \text{subtrees}_n(\text{hyp})} \text{cont}(t)},$$

де D – максимальна глибина піддерев; $count(t)$ – кількість появ піддерев t в синтаксичному дереві машинного перекладу; $count_{clip}(t)$ – скорочена кількість появ t у синтаксичних деревах еталонного перекладу [20].

Висновки

Для створення удосконалених систем машинного перекладу потрібні автоматизовані метрики оцінювання якості машинного перекладу. Огляд існуючих автоматизованих метрик, здійснений в цій роботі, показав, що:

– усі оглянуті метрики потребують набору еталонних перекладів, що унеможлиблює повністю автоматизовану оцінку нових текстів, які з'являються під час роботи системи машинного перекладу;

– відмінності в синтаксичних структурах речень потребують адаптації метрики STM до конкретної мовної пари.

1. Міщенко А.Л. Машинний переклад у контексті сучасного науково-технічного перекладу // Вісник ХНУ ім. В.Н. Каразіна. – Серія «Романо-германська філологія. Методика викладання іноземних мов». – № 1051. – С.172–180, 2013. 2. Tripathi S., Sarkhel J. K. Approaches to Machine Translation // Ann. of Library and Inform. Studies., Vol. 57. – P. 388–393, 2010. 3. Sanyal S., Borgohain R. Machine Translation Systems in India // Computing Research Repository, 2013 [Електронний ресурс]. – Режим доступу: <http://arxiv.org/ftp/arxiv/papers/1304/1304.7728.pdf>. 4. Франчук Н.П. Комп'ютерний переклад // Науковий часопис НПУ ім. М.П. Драгоманова. Серія №2. Комп'ютерно-орієнтовані системи навчання: зб. наук. праць / редрада. – К.: НПУ ім. М.П. Драгоманова, 2010. – № 8 (15). – С. 185–190. 5. Costa-jussà M. R., Banchs R., Rapp R., Lambert P., Eberle K., Babych B. Workshop on Hybrid Approaches to Translation: Overview and Developments // Proc. of the Second Workshop on Hybrid Approaches to Translation, p. 1–6, Sofia, Bulgaria, 2013. 6. Смірнова Т.В. Переваги та недоліки застосування машинного перекладу / Т.В. Смірнова // Мат. X міжнародної науково-технічної конференції "ABIA-2011". – Т.4. – К.: НАУ, 2011. – С. 36–37, 39. 7. White, J.S., T. O'Connell and F. O'Mara. 1994. The ARPA MT Evaluation Methodologies: Evolution, Lessons, and Future Approaches // Proc. of the First Conference of the Association for Machine Translation, p. 193–205. 8. Callison-Burch, C., C. Fordyce, P. Koehn, C. Monz and J. Schroeder. 2007. (Meta-)evaluation of machine translation // Proc. of the Second Workshop on Statistical Machine Translation, p. 136–158, Prague, Czech Republic, June. Association for Computational Linguistics. 9. Olive J. et al. Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation, Springer Science+ Business Media, 2011. 10. Стхамич Ю.С. Адекватність та еквівалентність перекладу в контексті комп'ютерної лінгвістики // Вісник Житомирського державного університету ім. І.Франка. – Філологічні науки. – № 66. – С.235–238, 2012. 11. Levenshtein, V. I. 1966. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. Soviet Physics Doklady, 10:707–710. 12. Nießén, S., F.J. Och, G. Leusch, and H. Ney. 2000. An evaluation tool for machine translation: Fast evaluation for MT research // In Proc. of the 2nd International Conference on Language Resources and Evaluation (LREC-2000). – P. 39–45. 13. Tillmann, C., S. Vogel, H. Ney, A. Zubiag and H. Sawaf. 1997. Accelerated DP Based Search For Statistical Translation // European Conference on Speech Communication and Technology. – P. 2667–2670, Rhodes, Greece, September. 14. Papineni, K., S. Roukos, T. Ward and W. Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation // Proc. of the 40th Annual Meeting of the Association for Comp. Linguistics, p. 311–318, Philadelphia, PA. 15. Banerjee, S. and A. Lavie. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments // Proc. of the ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization. – P. 65–72, Ann Arbor, Michigan, June. 16. Fellbaum, C. 1998. WordNet: An Electronic Lexical Database. MIT Press. 17. Turian, J.P., L. Shen and D.I. Melamed. 2003. Evaluation of machine translation and its evaluation // Proc. MT Summit IX. – P. 386–393, New Orleans, LA. 18. Snover, M., B. Dorr, R. Schwartz, L. Micciulla and J. Makhoul. 2006. A Study of Translation Edit Rate with Targeted Human Annotation // Proceedings of Association for

Machine Translation in the Americas (AMTA-2006). – P. 223–231, Cambridge, Massachusetts.
19. Leusch, G., N. Ueffing and H. Ney. 2006. CDER: Efficient MT Evaluation Using Block Movements // *Proc. of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006).* 20. Liu D., Gildea D. 2005. Syntactic Features for Evaluation of Machine Translation // *Proc.s of ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization.* – P. 25–32.

УДК 004.052; 519.62

П. Сердюк, О. Том'як

Національний університет "Львівська політехніка",
кафедра програмного забезпечення

АРХІТЕКТУРА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ВЕРИФІКАЦІЇ ТА АВТОМАТИЗОВАНОГО ТЕСТУВАННЯ МЕТОДІВ РОЗВ'ЯЗУВАННЯ ЗАДАЧ МАТЕМАТИЧНОЇ ФІЗИКИ

© Сердюк П., Том'як О., 2013

Розглянуто архітектуру програмного забезпечення для тестування та верифікації методів розв'язування систем рівнянь у часткових похідних, що описують фізичні процеси. Розроблено архітектуру процесора автоматизованого генерування задач, їх розв'язування та порівняльного аналізу з іншими методами аналізу задач математичної фізики. Проаналізовано роботу програмного забезпечення на оболонці COMSOL 4.2.

Ключові слова: автоматизоване тестування, диференціальні рівняння, математична фізика, верифікація.

The article deals with software architecture for testing and verification methods for solving systems of equations in partial derivatives describing physical processes. Developed automated generation of processor architecture problems, their solution and comparative analysis with other methods of analysis tasks mathematical physics. Analyzed with software on shell COMSOL 4.2.

Key words: automated testing, differential equations, mathematical physics, verification.

Вступ

Важливим аспектом розроблення нових методів розв'язування систем диференціальних рівнянь у часткових похідних є їх верифікація та перевірка. Однією з найскладніших проблем автоматизованого тестування є покриття тестовими випадками різного роду моделей. У цій статті пропонується автоматизоване генерування тестових моделей для верифікації математичних методів розв'язування задач математичної фізики.

Для верифікації та визначення збіжності методів розв'язування задач математичної фізики використовують послідовну схему генерації пакетів для розроблення тестових шаблонів послідовного ланцюга. Така техніка значно збільшує кількість можливих завдань і зменшує надмірну специфікацію змінних стану, що іноді виникає у разі використання стандартного алгоритму [1, 2].

Сьогодні розробляється багато методів для розв'язування задач математичної фізики та початково-крайових задач для лінійних і нелінійних інтегральних рівнянь часткових похідних [3, 4]. Для верифікації, аналізу збіжності та стійкості розв'язків у статті пропонується використати емпіричний метод вимірювання характеристик методу, який полягає в автоматизованому генеруванні вхідних даних для методів розв'язування та їх подальшого аналізу. У статті розглянуто архітектуру програмного забезпечення, яке складається з: трансляторів у математичні скриптові мови, генератора тестових конструкцій та модуля верифікації.

Задача генерування тестових конструкцій полягає у наданні вхідних даних за відповідними обмеженнями: просторовими, обмеженнями на фізичні величини та конструкції. Генерування