

УДК 081-13

Лопата Владислав Владиславович

студент

*Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»*

**АНАЛІЗ ЛЮДСЬКИХ ПЕРЕКЛАДІВ ТЕКСТІВ НА ВІДПОВІДНІСТЬ
ДО ТЕКСТУ ОРИГІНАЛУ**

**АНАЛИЗ ЧЕЛОВЕЧЕСКИХ ПЕРЕВОДОВ ТЕКСТОВ НА
СООТВЕТСТВИЕ ТЕКСТУ ОРИГИНАЛА**

**ANALYSIS OF HUMAN TRANSLATIONS OF TEXTS FOR
CORRESPONDENCE TO THE ORIGINAL TEXT**

Анотація. У статті розглянуто проблему відповідності тексту оригіналу до тексту перекладу та запропоновано метод співставлення перекладеного та оригінального текстів для виявлення показника точності перекладу.

Ключові слова: переклад, машинний аналіз, еквівалентність, BLEU.

Аннотация. В статье рассмотрена проблема соответствия текста оригинала в текст перевода и предложен метод сопоставления переведенного и оригинального текстов для выявления показателя точности перевода.

Ключевые слова: перевод, машинный анализ, эквивалентность, BLEU.

Summary. The article considers the problem of correspondence of the original text to the translated text and proposes a method of comparing the translated and original texts to identify the accuracy of the translation.

Key words: *translation, machine analysis, equivalence, BLEU.*

Вступ. Проблема якості перекладів продовжує залишатися актуальною, оскільки існуючі методи встановлення точності перекладу працюють на основі чітких показників і не використовують методи машинного навчання для покращення виявлення помилок у перекладацьких роботах. Існуючі методи аналізу текстів не є гнучкими у використанні, оскільки вони пропонують виявлення невідповідності перекладу лише виходячи з визначених теорією перекладів помилок, що в реальних випадках, через що, відверто помилкові переклади можуть такими методами розцінені як повністю вірні.

Постановка задачі. Задачею роботи є розробка системи, що дозволить проводити аналіз перекладів текстів та встановлювати показник точності перекладу

Аналіз існуючих способів побудови рекомендаційних систем. Переклад - дуже древній вид людської діяльності. Своїм корінням він іде в ті далекі часи, коли прамови почали розпадатися на окремі мови і виникла необхідність в людях, які знали кілька мов і здатних виступати в ролі посередників при спілкуванні представників різних мовних громад [1].

Еквівалентність перекладу - спільність змісту; смислова близькість оригіналу і перекладу. Межею перекладацької еквівалентності є максимально можливий ступінь збереження змісту оригіналу при перекладі, але в кожному окремому перекладі змістова близькість до оригіналу в різному ступені і різними способами наближається до максимальної [2].

Як правило, існує багато "ідеальних" перекладів даного вихідного речення. Ці переклади можуть відрізнятися у виборі слів або в порядку слів, навіть коли вони використовують ті самі слова. І все ж люди можуть чітко відрізнити хороший переклад від поганого [3].

На основі вище приведених даних можна зробити висновок про можливість побудови моделі аналізу перекладів на відповідність до оригіналу.

У роботі [1] в стиснутому вигляді пропонуються основні положення, на яких спираються при веденні перекладацької діяльності. З використанням цих положень можна побудувати аналітичну модель, яка перевіряла б наявність у тексті індикаторів відповідності або невідповідності основним поняттям теорії перекладу.

Зокрема, в роботі [1] вводиться поняття еквівалентності перекладу. Еквівалентність перекладу (від лат. *Aequus* - рівний, рівноцінний і *valentis* - чинний, ґрунтовний) - це спільність змісту (сміслова близькість), рівноцінність текстів оригіналу і перекладу.

У публікації пропонується класифікація національних реалій – це ключовий аспект, на який необхідно звертати увагу при встановленні спільності текстів, оскільки передача національних реалій є проблемою, з якою часто стикаються при перекладі творів етнічних культур відмінних від тих, що носять мову, на яку ведеться переклад.

В публікації не наведено вичерпного визначення спільності текстів, що не дозволяє на її основі повністю побудувати модель для аналізу тексту, але зі статі можна використати приклади перекладів для навчання нейронної мережі та згадану вище класифікацію національних реалій для покращення точності моделі, врахувавши можливість і неможливість передати точно певні реалії певних мов.

В роботі [2] приводиться загальна теорія перекладу, виходячи з якої можна встановити, чи відповідає виконаний переклад оригіналу. В роботі [2] приводяться приклади правильних і неправильних перекладацьких рішень, і на основі цих прикладів є можливим навчання нейронної мережі, яка б значно підвищила якість визначення відповідності перекладу до тексту оригіналу.

Публікація [2] пропонує визначення спільності текстів з використанням п'яти видів еквівалентності, тобто, спільність текстів наявна тільки тоді, коли між текстами наявний один з видів еквівалентності.

В публікації [2] відсутні аналітичні підходи до визначення спільності текстів, а лише надається визначення спільності. Виникає необхідність самостійно розробляти алгоритм визначення спільності на основі п'яти видів еквівалентності.

Робота [3] стосується математичних методів встановлення відповідності перекладів на прикладі встановлення відповідності машинного перекладу до перекладу, виконаного професіональним перекладачем. Метод BLEU, описаний в роботі, може бути використаним як підґрунтя для побудови математичної моделі встановлення відповідності тексту перекладу до тексту оригіналу. Такий підхід досі не використовувався в аналізаторах створеного людиною перекладу і є для цієї сфери новим інструментом, який призначений підвищити точність перекладу.

Робота [3] не розглядає використання машинного навчання в парі з методом BLEU для проведення аналізу тексту. Саме в парі з машинним навчанням метод BLEU є найкориснішим, оскільки не завжди можна знайти альтернативні варіанти перекладу, з якими б порівнювався перевірюваний, а нейронна мережа може прогнозувати варіант перекладу, що розширить діапазон випадків, де метод BLEU є застосовним.

Всі роботи, що розглянуті вище, не розглядають машинне навчання як спосіб визначення точності перекладу. Для предметного середовища даної роботи використання машинного навчання, оскільки його можливості будуть використані при побудові системи аналізу текстів.

Існуючі методи аналізу текстів не є гнучкими у використанні, оскільки вони пропонують виявлення невідповідності перекладу лише

виходячи з визначених теорією перекладів помилок, що в реальних випадках, через що, відверто помилкові переклади можуть такими методами розцінені як повністю вірні.

Підвищити точність перекладу шляхом впровадження аналізу тексту з використанням методів машинного навчання.

Опис моделі BLEU аналізу текстів перекладів на відповідність до оригіналу. Розглянемо два варіанти перекладу речення з китайської, наведені в публікації [2]:

Варіант 1: It is a guide to action which ensures that the military always obeys the commands of the party.

Варіант 2: It is to insure the troops forever hearing the activity guidebook that party direct.

Хоча вони, схоже, стосуються однієї теми, вони помітно відрізняються за якістю. Для порівняння пропонується розглянути три людські переклади того самого речення.

Посилання 1: It is a guide to action that ensures that the military will forever heed Party commands.

Посилання 2: It is the guiding principle which guarantees the military forces always being under the command of the Party

Посилання 3: It is the practical guide for the army always to heed the directions of the party.

Зрозуміло, що хороший переклад, варіант 1, поділяє багато слів і фраз із цими трьома перекладами з посилань, тоді як варіант 2 - ні. Варіант 1 поділяє фразу " It is a guide to action" з посиланням 1, "which " із посиланням 2, "ensures that the military" з посиланням 1, "always" з посиланнями 2 і 3, "commands" з посиланням 1 і нарешті "of the party" з посиланням 2. На противагу цьому, варіант 2 демонструє набагато менше збігів, і їхній ступінь менший.

Зрозуміло, що програма може дати варіанту 1 оцінку вище, ніж варіанту 2, просто порівнявши збіги n-грамів між перекладами кожного варіанту та довідкові переклади.

N-грама — послідовність з n елементів. З семантичної точки зору, це може бути послідовність звуків, складів, слів або букв. На практиці частіше зустрічається N-грами як ряд слів, стійкі словосполучення називають колокацію. Послідовність з двох послідовних елементів часто називають біграм, послідовність з трьох елементів називається триграма. Не менш чотирьох і вище елементів позначаються як N-грами, N замінюється на кількість послідовних елементів.

Далі у статі n-грамами будемо вважати послідовність з n слів.

Bilingual Evaluation Understudy, або скорочено BLEU – метод визначення спільності двох текстів з використанням математичної статистики.

Основним завданням програмування методу BLEU є порівняння n-грамів кандидата з n-грамами еталонного перекладу та підрахунку кількості збігів. Ці збіги не залежать від позиції. Чим більше збігів, тим кращий варіант перекладу. Для простоти спочатку буде розглянуто на обчисленні збігів уніграмів.

Далі буде розглянуто спосіб обчислення модифікованої точності n-граму на наборі тестів із кількома реченнями. Незважаючи на те, що зазвичай оцінюються системи машинного перекладу на основі цілих документів, основною одиницею оцінки є речення.

Вихідний рядок може перекладатися на багато цільових рядків, і в цьому випадку вважаємо набір таких рядків одним “реченням”. Спочатку необхідно обчислити n-грамові збіги речення до речення. Далі додати обрізану кількість n-грамів для всіх кандидатських речень і поділити на кількість n-грамів кандидата в тестовому корпусі, щоб обчислити змінений бал точності, для всього тестового корпусу.

$$p_n = \frac{\sum_{C \in \text{Candidates}} \sum_{n\text{-grams} \in C} \text{Count}_{clip}(n\text{-gram})}{\sum_{C' \in \text{Candidates}} \sum_{n\text{-grams} \in C'} \text{Count}_{clip}(n\text{-gram})}$$

де $\text{Count}(n\text{-gram})$ – кількість n-грам у перевірюваному тексті, $\text{Count}(n\text{-gram}')$ – кількість n-gram у еталонному тексті.

Щоб підтвердити, що модифікована точність n-грам розрізняє дуже хороші переклади та погані переклади, необхідно обчислити модифіковані числа точності для перекладу хорошого перекладача та стандартної (поганої) системи машинного перекладу, використовуючи 4 довідкові переклади. Середні результати точності показані на рисунку 1.

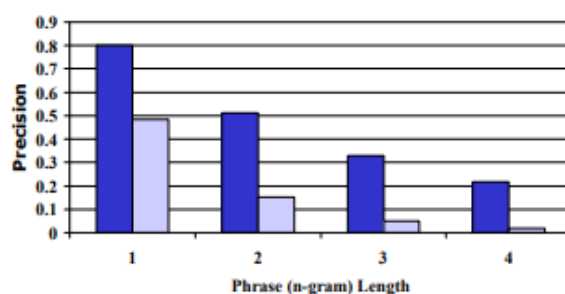


Рис. 1. Середні результати точності

Показник отриманий за методом BLEU змінюється в діапазоні від 0 до 1. Значення 1 отримують тільки ті переклади, що повністю збігаються з еталонним. З цієї причини навіть перекладач-людина не обов'язково отримає оцінку 1. Важливо зауважити, що чим більше посилальних перекладів на речення, тим вищий бал у вірного перекладу. Таким чином „грубі” порівняння оцінок з різною кількістю посилальних перекладів робити не можна: у тестовому корпусі, що складає близько 500 речень (40 загальних новин), перекладач-людина може отримати показник 0,3468 проти чотирьох посилань та 0,2571 проти двох посилань. В таблицях 1, 2 показано оцінки BLEU з 5 систем проти двох посилань у тому самому тестовому корпусі.

Таблиця 1

Показник Bleu, отриманий на 500 реченнях

S1	S2	S3	H1	H2
0.0527	0.0829	0.0930	0.1934	0.2571

Таблиця 2

Статистика на 20 блоках n-грам

	S1	S2	S3	H1	H2
Mean	0.051	0.081	0.090	0.192	0.256
StdDev	0.017	0.025	0.020	0.030	0.039
t	—	6	3.4	24	11

Описаний вище метод можна застосувати під час навчання нейронної мережі з метою визначення точності її роботи, а показник як один з показників точності перекладу.

Перевага цього методу полягає в тому, що він не залежить від визначених в теорії перекладу помилок і може виявити неточності в перекладі, який не класифікується як помилковий згідно до теорії перекладу.

Його недоліки полягають в тому, що він працює з еталонними перекладами, які є не для будь-якого тексту. В таких випадках доводиться використовувати машинні переклади.

У рамках наукового дослідження планується розробити інформаційну систему, що визначає точність перекладу за допомогою методу, описаного вище.

Висновки. У роботі розглянуто проблему відповідності тексту оригіналу до тексту перекладу та запропоновано метод співставлення перекладеного та оригінального текстів для виявлення показника точності перекладу.

Розглянуто метод BLEU для визначення показника точності перекладу. Визначено показники, за якими можна виявити факт наявності

у перекладі помилок, що допоможе перекладачам одразу відслідковувати в себе неточності і виправляти текст відповідно.

Література

1. Теория перевода: основные понятия и проблемы : [учеб. пособие] / М. Ю. Илюшкина ; [науч. ред. М. О. Гузикова] ; М-во образования и науки Рос. Федерации, Урал. федер. ун-т. Екатеринбург : Изд-во Урал. ун-та, 2015. 84 с.
2. Комиссаров В. Н. Общая теория перевода : учеб. пособие / В. Н. Комиссаров. М. : Че Ро, 1999. 136 с.
3. Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu IBM T. J. Watson Research Center Yorktown Heights, NY 10598, USA.