УДК 004.8:81’322.4 **DOI:** 10.31673/2412-9070.2026.017410
А. І. ПИЛИПЕНКО, канд. техн. наук, доцент;
ORCID: 0000-0002-6343-4469
Д. Є. ДАНИЛКО, студентка,
ORCID: 0009-0000-4176-4398
Київський національний університет імені Тараса Шевченка

МІЖМОВНЕ ВИЯВЛЕННЯ ОМОНІМІВ ІЗ ВИКОРИСТАННЯМ ЗМАГАЛЬНОГО ВИРІВНЮВАННЯ ВЕКТОРНИХ ПРЕДСТАВЛЕНЬ ТА ЗВАЖЕНОГО ЗА РІЗНИЦЕЮ РАНГІВ МЕТОДУ CSLS

У цій статті представлено новий метод виявлення міжмовних омонімів, який поєднує змагальне вирівнювання векторних представлень слів із зваженим за різницею рангів показником CSLS (Cross-domain Similarity Local Scaling). Запропонований підхід покращує якість вирівнювання для типологічно близьких, малоресурсних мов, використовуючи польсько-українські пари слів як приклад дослідження. Спочатку векторні представлення слів вирівнюються за допомогою некерованого змагального методу з подальшим уточненням через CSLS. Потім семантична подібність обчислюється за допомогою CSLS і додатково коригується шляхом застосування штрафів за різницю рангів на основі взаємних рангів найближчих сусідів. Експериментальні результати показують, що така модифікація CSLS суттєво підвищує здатність розрізняти три семантичні групи, зокрема ізолюючи омонімічні пари слів. Класифікатор на основі правил із оптимізованими порогами CSLS досягає середнього макро-значення F1 = 0.916, що підтверджує ефективність запропонованого методу.

Ключові слова: міжмовні омоніми; векторні представлення слів; CSLS; змагальне вирівнювання; некероване навчання; семантична подібність; різниця рангів; FastText.

Постановка проблеми

Актуальним завданням у багатомовній обробці природної мови є підвищення якості вирівнювання векторних представлень слів для типологічно близьких мов, таких як польська та українська. Зокрема, сучасним викликом у багатомовній обробці природної мови є виявлення міжмовних омонімів [1]. Це пари слів, які мають схожу форму у різних мовах, але відрізняються за значенням, що часто призводить до помилок у перекладі, семантичному пошуку або вивченні мов. Останні досягнення у вирівнюванні міжмовних векторних представлень зробили можливим проєкцію одномовних векторних просторів у спільний семантичний простір без необхідності використання паралельних корпусів. Зокрема, підхід змагального вирівнювання, запропонований Conneau та ін. (2018) [2], представив некерований метод зіставлення векторних представлень слів між мовами, що поєднує генеративні змагальні мережі з уточненням за допомогою показника Cross-domain Similarity Local Scaling (CSLS). Цей метод широко застосовується для таких завдань, як індукція двомовних лексиконів і некерований машинний переклад.

Удосконалюючи підхід до вимірювання та зважування подібності, ця робота спрямована на внесок у ширшу проблему лексичної дезамбігуації для малоресурсних, близькоспоріднених мовних пар, з потенційними застосуваннями у машинному перекладі, укладанні словників і багатомовному пошуку.

Аналіз останніх досліджень і публікацій

Було запропоновано кілька альтернативних методів вирівнювання міжмовних векторних представлень, кожен із яких має свої переваги, але також і певні обмеження щодо виявлення омонімічних пар слів, такі як Метод Meemi (Meeting in the middle) [3], підхід Zhou та ін. (2022) [4], метод Wada та ін. (2018) [5] та метод, заснований на RCSLS (Relaxed Cross-Domain Simila-

rity Local Scaling) [6]. Дані підходи не були використані у даному дослідженні, оскільки вони менш придатні для виявлення тонких міжмовних відмінностей у випадках, коли слова мають схожу форму, але різні значення.

Однак більшість існуючих систем зосереджені на вирівнюванні семантично еквівалентних слів і не враховують нюансів, коли формально подібні слова відрізняються за змістом, тобто є омонімами. Крім того, поширені метрики оцінювання, такі як косинусна подібність або необроблені значення CSLS, часто виявляються недостатніми для розпізнавання тонких семантичних відмінностей, особливо у випадках часткового чи асиметричного вирівнювання. Хоча CSLS допомагає зменшити проблему «hubness», властиву просторам великої розмірності, він не враховує двонаправлену узгодженість між словами вихідної та цільової мов.

Формулювання мети статті

З метою подолання зазначеного обмеження у даному дослідженні пропонується механізм зважування за різницею рангів, який коригує значення CSLS з урахуванням взаємної узгодженості рангів між парами слів. Такий підхід має на меті підвищити інтерпретованість і надійність оцінки семантичної подібності шляхом накладення штрафів за невідповідні або односторонні вирівнювання. Дослідження зосереджене на польсько-українських парах слів через їхню типологічну близькість і обмежену кількість якісних двомовних ресурсів.

Метою цієї статті було перевірити чи може поєднання змагального вирівнювання [7, 8], CSLS та зважування за різницею рангів надійно розрізняти міжмовні омоніми [1] від правильних перекладів і несхожих пар слів. Для аналізу було визначено три семантичні групи: (1) семантично еквівалентні переклади, (2) омоніми (схожі за формою, але різні за значенням), (3) семантично не пов'язані пари. У дослідженні використовувалися некеровані векторні представлення fastText, змагальне вирівнювання без використання англійської як посередника та класифікатор на основі правил для оцінки ефективності.

Виклад основного матеріалу

Змагальне вирівнювання векторних представлень із уточненням за допомогою CSLS

Вирівнювання на основі методу Relaxed Cross-Domain Similarity Local Scaling (RCSLS) [6] показало середню Top-1 точність на рівні 71% для перекладу найближчих сусідів (з найвищими значеннями для споріднених мов – 83-86%), однак для пари польська–українська ця точність склала лише 68%, незважаючи на спорідненість мов. Крім того, для слов'янських мов бракує високоякісних паралельних даних, тому ефективність керованого навчання може бути переоціненою.

З урахуванням вищезгаданих факторів було вирішено використати змагальне вирівнювання просторів векторних представлень з уточненням CSLS, що є некерованим методом, описаним Conneau та ін. (2017) [2]. У подальшому навчанні були використані одномовні векторні представлення FastText розмірністю 300 для української та польської мов. Кожна модель містила понад 1,5 мільйона векторів, тому з кожної було обрано по 300 000 найбільш уживаних слів. Польські представлення, позначені як $X = \{x_1, \ldots, x_n\}$, використовувалися як джерельна мова, а українські – як цільова $Y = \{y_1, \ldots, y_n\}$, де n – кількість векторів.

Після підготовки даних була ініціалізована лінійна матриця відображення $W \in R^{d \times d}$, де d – розмірність embedding-простору. Далі було навчено дискримінатор D_θ, метою якого було розрізнення між перетвореними джерельними векторними представленнями Wx_i та справжніми векторами цільової мови y_i.

$$L_D(\theta_D|W) = -\frac{1}{n}\sum_{i=1}^{n} logP_{\theta_D}(source = 1|Wx_i) - \frac{1}{m}\sum_{i=1}^{m} logP_{\theta_D}(source = 0|y_i), \qquad (1)$$

де L_D – функція втрат дискримінатора; θ_D – параметри нейронної мережі дискримінатора; W – лінійна матриця відображення; n – кількість векторів мови-джерела; m – кількість векторів цільової мови; $P_{\theta_D}(source = 1|Wx_i)$ – ймовірність, що вектор Wx_i є з мови-джерела; $P_{\theta_D}(source = 0|y_i)$ – ймовірність, що вектор y_i є з цільової мови.

Мета дискримінатора – максимізувати функцію втрат (або мінімізувати її негатив) для правильного розпізнавання походження векторів.

Далі здійснювалося навчання матриці відображення W, що мало за мету мінімізувати здатність дискримінатора розрізняти перетворені та справжні вектори:

$$L_W(W|\theta_D) = -\frac{1}{n}\sum_{i=1}^{n} logP_{\theta_D}(source = 0|Wx_i) - \frac{1}{m}\sum_{i=1}^{m} logP_{\theta_D}(source = 1|y_i), \quad (2)$$

де L_W – функція втрат для перетворення.

Таким чином, дискримінатор (1) навчається розрізняти простори, а відображення (2) – ускладнювати це розрізнення, наближаючи обидва простори. Описане навчання тривало 5 епох, досягши максимальної точності 53.9% (Top-1) на третій епосі. Крім того, кількість кроків навчання дискримінатора між оновленнями перетворення було зменшено з 5 до 2, що дозволило уникнути перенавчання і дало приріст точності на 2%. Подальше збільшення кількості епох не дало додаткового покращення.

Після завершення навчання було проведено етап уточнення. Для цього застосовувався метод CSLS (Cross-domain Similarity Local Scaling), який дозволив зіставити найчастотніші слова джерельної та цільової мов з метою побудови синтетичного словника. Важливо, що як якірні пари для словника використовувалися лише взаємні найближчі сусіди. Далі було обчислено найкращу ортогональну матрицю, яка вирівнює відповідні пари слів, використовуючи алгоритм Прокруста [8].

$$W^* = \underset{W \in O_d(R)}{argmin} ||WX - Y||_F = UV^T, \quad (3)$$

де W – ортогональна матриця; $Q_d(R)$ – ортогональна група дійсних матриць розміру $d \times d$ $U\Sigma V^T = SVD(YX^T)$, де SVD – це розклад за сингулярними значеннями, а Σ – діагональна матриця сингулярних значень, що є частиною етапу SVD, необхідного для отримання матриць U та V.

Етап уточнення (3) проводився 5 разів, досягнувши найвищої Top-1 точності [9, 10] (72.5%) на третій ітерації. Після цього було помічено незначне зниження ефективності, що свідчить про оптимальність цього порогу. Отриманий показник є на 6.6 відсоткових пунктів вищим за точність, досягнуту методом RCSLS [6] у початковому вирівнюванні FastText представлень.

Косинусна подібність і метод CSLS у виявленні міжмовних омонімів

Після того як векторні простори вихідної та цільової мов були вирівняні якомога точніше і було отримано якісні вирівняні векторні представлення слів, стало можливим визначити семантичну подібність між парами слів для виявлення міжмовних омонімів [1]. Для вирішення цієї задачі було застосовано метод Cross-domain Similarity Local Scaling (CSLS) [2], який зменшує значення подібності для слів, надто близьких до багатьох інших, і збільшує – для тих, що близькі лише до кількох:

$$CSLS(A, B) = 2 \cdot cos(A, B) - r_T(A) - r_S(B), \quad (4)$$

де A, B – вектори слів у парі потенційних міжмовних омонімів; $\cos(A, B)$ – косинусна подібність; $r_T(A)$ – середня подібність між вихідним словом A та його найближчими сусідами у цільовій мові; $r_S(B)$ – середня подібність між цільовим словом B та його найближчими сусідами у вихідній мові. Значення $r_T(A)$ і $r_S(B)$ обчислюються як середнє топ-k значень косинусної подібності між відображеним вихідним словом і його k найближчими сусідами у цільовій мові, відповідно.

Таким чином, метод CSLS (6) накладає штраф на подібності, які є високими лише через загальну близькість до великої кількості інших слів.

Далі було підготовлено набір із 150 польсько-українських пар – по 50 пар у кожній із трьох груп. Аналіз значень CSLS між трьома групами показав, що CSLS має сильну розрізнювальну здатність, особливо у відмежуванні слів групи 1 (значення від -0.1086 до 0.5629) від груп 2 (від -0.7805 до 0.2665) і 3 (від -0.8984 до -0.0269). Таке розділення засвідчило, що CSLS є ефектив-

ним для виявлення надійних міжмовних відповідників слів і зменшення впливу проблеми “hubness”, характерної для високовимірних векторних просторів [7].

Модифікація CSLS шляхом зважування за різницею рангів

Як видно з діапазонів отриманих значень CSLS, хоча група 1 була частково відокремлена, діапазони значень CSLS значною мірою перетиналися між групами 2 та 3. Це могло бути спричинено тим, що слова в обох групах, незважаючи на схожість у написанні або вимові, суттєво відрізняються за значенням. Єдина відмінна риса між групами полягає в тому, що слова у парах з групи 2, яка містить міжмовні омоніми, можуть бути помилково сприйняті людиною. Ця особливість ускладнює чітке розмежування пар груп 2 і 3.

З метою вирішення цієї проблеми було застосовано ваговий коефіцієнт різниці рангів. Цей метод дозволив скоригувати оцінку CSLS [2], аби відрізняти три групи слів за рівнем семантичної подібності через відмінності у рангах та систему штрафів. Зокрема, він дозволив знизити оцінку для пар з групи 3, оскільки слова у таких парах ніколи не є взаємними найближчими сусідами, що вирізняє їх від пар з групи 2. Спершу обчислювався ранг серед найближчих сусідів цільового українського слова (rank_pl2uk) після перекладу вихідного польського слова. Потім аналогічно визначався ранг польського слова (rank_uk2pl) після перекладу українського. Далі обчислювалася різниця рангів:

$$rank\_diff = log(1 + |rank\_pl2uk - rank\_uk2pl|)\,. \quad (5)$$

Після отримання значення різниці рангів стало можливим встановити штрафи для кожної групи на основі ступеня взаємної узгодженості у ранжуванні. Для визначення меж різниці рангів було використано набір із 150 польсько-українських пар. Після обчислення різниці рангів для кожної пари було виявлено, що для групи 1 (семантично подібні слова) різниця рангів коливалася у межах від 0 до 1; для групи 2 (міжмовні омоніми) – від 4 до 6.3; для групи 3 (слова з різним значенням, які не можуть бути сплутані) – переважно понад 8. Відповідно до цих спостережень були визначені наступні межі:

$rank\_diff \leq 1$ – group 1;

$1 < rank\_diff \leq 6.3$ – group 2;

$rank\_diff > 6.3$ – group 3.

Для групи 1 штраф виконував роль бонусу, збільшуючи оцінку CSLS [2] для пар із найвищим ступенем відповідності. Зокрема, множник варіювався від $1.5x$ для ідеально узгоджених пар до $1.2x$ для граничних випадків, що описується формулою:

$$rank\_penalty = 1.5 - 0.3 \times rank\_diff. \quad (6)$$

Це забезпечило більш гнучке оцінювання, віддаючи перевагу найбільш достовірним перекладним парам.

Для групи 2 застосовувався косинусний штраф, у якому найменше зниження ($0.05x$) припадало на центр діапазону ($rank\_diff \approx 3.65$), тоді як максимальне ($0.2x$) – на його межі:

$$rankpenalt = \min factor + \max\_factor - \min\_factor \times cosnorm\_dist \times \pi 2\,, \quad (7)$$

де min_factor – це найменший штрафний коефіцієнт (0.05); max_factor – це найбільший штрафний коефіцієнт (0.2); $norm\_dist = \frac{|rank\_diff - midpoint|}{half\_range}$ є нормалізованою відстанню до середини діапазону.

Такий підхід плавно знижував оцінки для менш узгоджених пар, зберігаючи адекватні результати для помірно послідовних відповідників. Косинусна функція була обрана для забезпечення плавності штрафної кривої.

У групі 3, де значення показника CSLS майже ніколи не було додатним, до значення CSLS застосовувався лінійно зростаючий штраф для різниць рангів, більших або рівних 6.3. Значення штрафу варіювалося від множника $1.5x$ до $2.0x$, залежно від того, наскільки невідповідними були ранги:

$$rank\_penalty = min\_factor + (max\_factor - min\_factor) \times scale, \quad (8)$$

де min_factor – це найменший штрафний коефіцієнт (1.5); max_factor – це найбільший

штрафний коефіцієнт (2.0); $scale = \frac{rank\_diff - min\_rank}{max\_rank - min\_rank}$ є нормалізованим значенням у межах від 0 до 1.

Це дозволило уникнути збереження низькоякісних пар перекладу шляхом зменшення їхніх показників подібності. Параметр scale використовувався для плавного регулювання сили штрафу – чим більше значення scale, тим суворішим ставав штраф.

Оскільки штрафи за різницю рангів для кожної з груп були визначені, з'явилася можливість застосувати їх до значень CSLS у межах пар слів кожної групи, множачи показники CSLS на відповідні штрафні коефіцієнти. Після застосування цієї процедури розмежування між групами покращилося порівняно з результатами, отриманими при використанні необроблених значень CSLS, група 1 – значення від -0.0016 до 0.9944, група 2 – значення від -1.0164 до 0.4736, група 3 – значення від -2.1927 до -0.4349. У результаті стало можливим надійно відрізнити міжмовні омоніми, що належать до групи 2. Отримані результати продемонстрували, що впровадження вагового коефіцієнта різниці рангів підвищило інтерпретованість, точність та надійність CSLS як міри впевненості при міжмовному вирівнюванні слів.

Результати та аналіз

Для вирівнювання польських та українських векторних представлень слів було використано змагальне вирівнювання векторних представлень із уточненням за допомогою CSLS [2] замість керованого методу RCSLS [6]. Використання методу CSLS підвищило точність на 4.263 при $k = 1$ найближчих сусідів, на 3.238 при $k = 5$ і на 3.006 при $k = 10$. Показник Top-1 precision [10], який є важливою метрикою, що відображає правильний взаємний переклад, збільшився до 72,5%, що на 6,6 відсоткових пунктів вище, ніж точність, досягнута за допомогою методу RCSLS [6] у початковому вирівнюванні FastText. Крім того, розподіл значень CSLS суттєво покращився при використанні моделей, вирівняних за допомогою змагального вирівнювання векторних представлень із уточненням CSLS [2].

Щодо методу визначення семантичної подібності у межах пар слів, додавання зважування за різницею рангів до показника CSLS сприяло кращому розмежуванню груп, надаючи можливість чітко виділити цільову групу 2, яка містить пари міжмовних омонімів. Щільність розподілу з і без урахування зважування за різницею рангів продемонстровано на рис. 1-2.

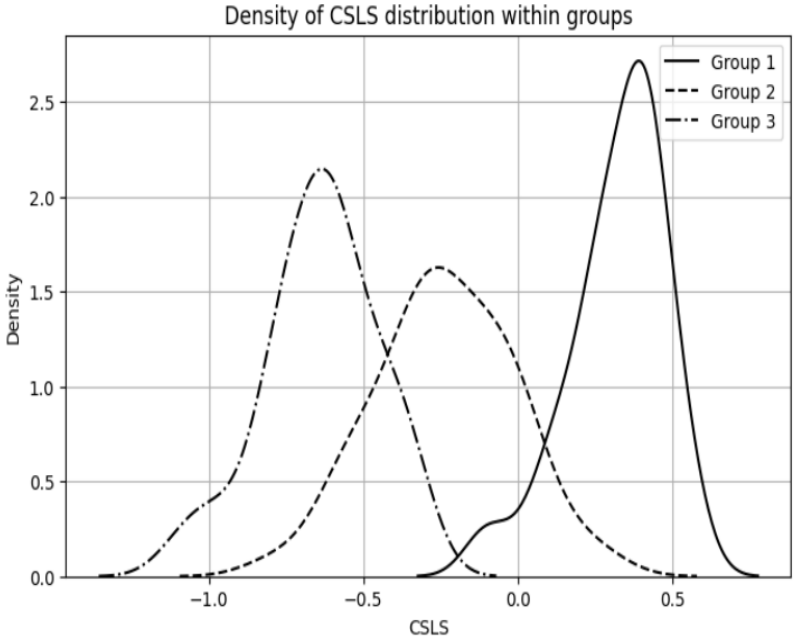

Рис. 1. Щільність розподілу значень CSLS між групами без зважування за різницею рангів

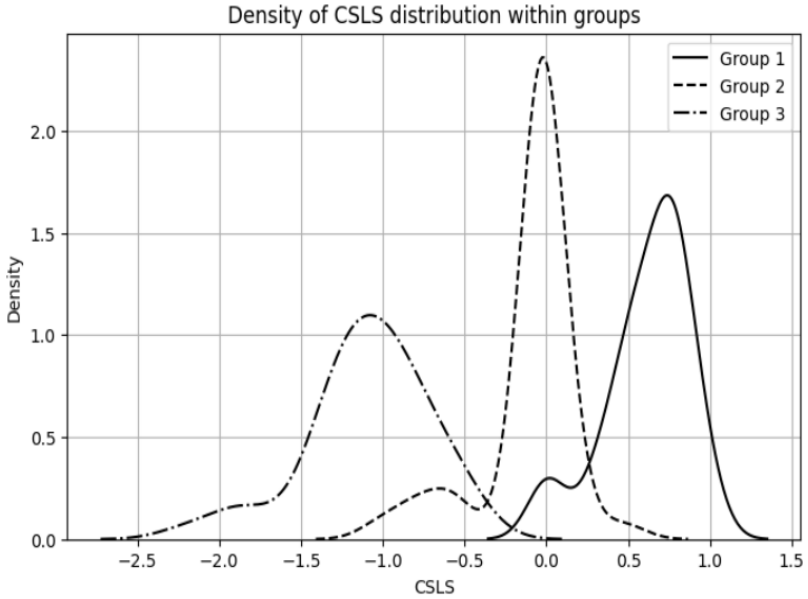

Рис. 2. Щільність розподілу значень CSLS між групами із зважуванням за різницею рангів

Таким чином, показник CSLS [2], покращений за допомогою зважування за різницею рангів, виявився найефективнішим методом для розмежування трьох семантичних груп. Як показано на рис. 1, 2, розподіли стали більш виразними, особливо для групи 2, яка відповідає міжмовним омонімам [1]. Для кількісної оцінки було реалізовано класифікатор на основі правил, який використовував порогові значення, оптимізовані для максимізації середнього макро-F1 показника. Пороги були визначені шляхом повного перебору спостережуваних значень CSLS у процесі класифікації на основі правил, у результаті чого були знайдені оптимальні межі: $t_1 = 0.1814$ та $t_2 = -0.4305$. Використовуючи ці пороги, класифікатор досяг макро-усередненого значення $F1 = 0.916$, що підтвердило надійність показника CSLS як ефективної межі прийняття рішень. Ці результати демонструють, що показник CSLS із зважуванням за різницею рангів не лише підвищує інтерпретованість результатів, але й забезпечує високоточну класифікацію пар слів за їхньою семантичною належністю.

Висновки

Отже, у цьому дослідженні було продемонстровано, що поєднання змагального вирівнювання векторних представлень із CSLS-метрикою, зваженою за різницею рангів, ефективно покращує виявлення міжмовних омонімів у типологічно близьких мовах, зокрема на прикладі польсько-української пари. Введення штрафів на різницю рангів дало змогу моделі точно розрізняти три семантичні групи, безпосередньо розв'язуючи задачу ідентифікації омонімів – слів, що виглядають подібними у різних мовах, але відрізняються за значенням. Запропонований метод робить внесок у розвиток підходів до міжмовного вирівнювання слів, інтегруючи структурну узгодженість у процес семантичного оцінювання. Він має практичну цінність для мов із низькими ресурсами та типологічно близьких мовних пар, де омонімія та лексична неоднозначність є частими й складними для автоматичного розпізнавання.

Список літератури

1. Sosnowski, Wojciech, and Maciej Jaskot. O Fałszywych Przyjaciołach Tłumacza Na Przykładzie Leksykonu Aktywnej Polskiej I Ukraińskiej Frazeologii. 1 Oct. 2017, www.researchgate.net/publication/320273226_O_falszywych_przyjaciolach_tlumacza_na_przykladzie_leksykonu_aktywnej_polskiej_i_ukrainskiej_frazeologii

2. Conneau, Alexis, et al. Published as a Conference Paper at ICLR 2018 WORD TRANSLATION without PARALLEL DATA. 2018. Available: https://openreview.net/pdf?id=H196sainb

3. Doval, Yerai, et al. "Meemi: A Simple Method for Post-Processing and Integrating Cross-Lingual Word Embeddings." Natural Language Engineering, vol. 29, no. 3, Cambridge University Press, Oct. 2021, pp. 746–68, https://doi.org/10.1017/s1351324921000280. Accessed 29 July 2025.

4. Zhou, Dong, et al. "Neural Topic-Enhanced Cross-Lingual Word Embeddings for CLIR." Information Sciences, vol. 608, Elsevier, June 2022, pp. 809–24, https://doi.org/10.1016/j.ins.2022.06.081

5. Wada, Takashi, and Tomoharu Iwata. "Unsupervised Cross-Lingual Word Embedding by Multilingual Neural Language Models." https://arxiv.org/abs/1809.02306

6. Joulin, Armand, et al. Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion. Association for Computational Linguistics, 2018, https://aclanthology.org/D18-1330.pdf

*7. Radovanovic, Milos, et al. "Hubs in Space: Popular Nearest Neighbors in High-Dimensional Data Hubs in Space: Popular Nearest Neighbors in High-Dimensional Data * Alexandros Nanopoulos Mirjana Ivanović." Journal of Machine Learning Research, vol. 11, 2010.*

8. Tsang, Sik-Ho. "Review — Word Translation without Parallel Data - Sik-Ho Tsang - Medium." Medium, 17 Sept. 2022, https://sh-tsang.medium.com/review-word-translation-without-parallel-data-ab20083bc246

9. Toliupa S., Pylypenko A., Tymchuk O., Kohut O. Simulated Datasets Generator for Testing Data Analytics Methods // 20th International Scientific Conference "Dynamical System Modeling and Stability Investigation", DSMSI 2023 - Volume 1: Mathematical Foundations of Information Technologies; Kyiv; Ukraine; Volume 3687, 2023, Pages 11-24. URL: https://ceur-ws.org/Vol3687/Paper_2.pdf

10. Sergio A. Alvarez, "An exact analytical relation among recall, precision, and classification accuracy in information retrieval", In Boston College, Boston, Technical Report, June 2002, Chestnut Hill, USA, Published by Boston College Libraries, Available: http://cs.bc.edu/~alvarez/APR/

A. Pylypenko, D. Danylko

CROSS-LINGUAL HOMONYM DETECTION WITH ADVERSARIAL EMBEDDING ALIGNMENT AND RANK-DIFFERENCE-WEIGHTED CSLS

The article presents a novel method for detecting cross-lingual homonyms – pairs of words in different languages that are similar in form but differ in meaning. Such pairs are a significant source of errors in machine translation and other multilingual Natural Language Processing (NLP) tasks. The research focuses on this problem for typologically close, yet low-resource language pairs, using Polish and Ukrainian as a case study. The proposed approach is implemented as a two-stage pipeline, which significantly improves the quality of aligning cross-lingual word vector representations and the subsequent computation of semantic similarity. In the first stage, monolingual FastText vectors for Polish and Ukrainian are aligned using unsupervised adversarial training, which projects both languages into a shared vector space without using parallel data or English as an intermediary. This alignment is further refined using the Procrustes algorithm, which leverages a synthetic dictionary built from mutual nearest neighbors identified by the Cross-domain Similarity Local Scaling (CSLS) method. The second stage, which is the main innovation of the work, involves refining the standard CSLS metric to compensate for the "hubness" problem in high-dimensional spaces. The authors introduce a rank-difference weighting mechanism that penalizes or encourages word pairs depending on the mutual consistency of their nearest neighbor ranks in both translation directions (from Polish to Ukrainian and vice versa). This correction creates a more sensitive similarity metric, allowing for the effective distinction of three semantic groups: true translations; cross-lingual homonyms, which could be misinterpreted by a human; and formally similar but semantically unrelated words. Experimental results on a specially compiled dataset of 150 Polish-Ukrainian word pairs show that CSLS with rank-difference weighting provides significantly better separation between these groups than standard cosine similarity or standard CSLS. In conclusion, the research contributes to the field of cross-lingual natural language processing by demonstrating that a combination of robust unsupervised alignment and a semantically-grounded, rank-weighted similarity metric makes it possible to effectively solve the complex task of homonym detection.

Keywords: cross-lingual homonyms; word embeddings; CSLS; adversarial alignment; unsupervised learning; semantic similarity; rank difference; FastText.

Надійшла до редакції: 03.10.2025
Прийнята до друку: 20.11.2025
Опубліковано: 27.02.2026