UDC: 004.934.2 https://doi.org/10.15407/jai2025.04.078

М. С. Клименко[1], **М. С. Шаш**[2]
[1]Інститут проблем штучного інтелекту МОН України і НАН України, Україна
7, Солом'янська вул., м. Київ, 03110
[2]Державний університет інформаційно-комунікаційних технологій, Україна
7, Солом'янська вул., м. Київ, 03110
[1]nik@ipai.net.ua
[2]max.shash@gmail.com
[1]https://orcid.org/0000-0003-4433-6641
[2]https://orcid.org/0009-0009-3274-5318

АНАЛІЗ МЕТОДІВ ОБЧИСЛЕННЯ СЕМАНТИЧНОЇ ВІДСТАНІ ДЛЯ ОЦІНКИ ЕФЕКТИВНОСТІ ПРИРОДНОМОВНИХ ЧАТ-БОТІВ

M. Klymenko[1], **M. Shash**[2]
[1]Institute of Artificial Intelligence Problems of the MES of Ukraine and the NAS of Ukraine, Ukraine
7, Solomyanska St., Kyiv, 03110
[2]State University of Information and Communication Technology, Ukraine
7, Solomyanska St., Kyiv, 03110
[1]nik@ipai.net.ua
[2]max.shash@gmail.com
[1]https://orcid.org/0000-0003-4433-6641
[2]https://orcid.org/0009-0009-3274-5318

ANALYSIS OF SEMANTIC DISTANCE CALCULATION METHODS FOR EVALUATION THE EFFECTIVENESS OF NATURAL LANGUAGE CHATBOTS

Анотація. У статті здійснено комплексний аналіз сучасних методів обчислення семантичної відстані між текстовими одиницями з метою оцінювання ефективності природномовних чат-ботів. Розглянуто еволюцію підходів до вимірювання семантичної подібності — від класичних лексико-статистичних методів і статичних векторних вбудовувань до контекстуалізованих моделей глибинного навчання, зокрема BERT та його похідних. Емпіричне дослідження проведено на діалоговому корпусі віртуального психологічного асистента, розробленого для надання психологічної підтримки. Ефективність методів оцінювалася за кількісними метриками класифікації намірів і відбору відповідей, а також за якісною експертною оцінкою адекватності відповідей. Отримані результати свідчать про суттєву перевагу контекстуалізованих моделей, зокрема SimSCE-BERT, над традиційними підходами, такими як word2vec і базовий BERT. Показано, що використання сучасних методів обчислення семантичної відстані сприяє підвищенню як технічної продуктивності чат-ботів, так і сприйманої користувачами якості взаємодії, що є критично важливим для масштабованих систем у сфері психологічної допомоги та інших прикладних доменах.

Ключові слова: семантична подібність, обробка природної мови, чат-боти, векторні вбудовування, BERT, психологічний асистент.

Abstract. The article presents a comprehensive analysis of modern methods for computing semantic distance between textual units in order to evaluate the effectiveness of natural language chatbots. The evolution of approaches to measuring semantic similarity is examined — from classical lexico-statistical methods and static vector embeddings to contextualized deep learning models, in particular BERT and its derivatives. The empirical study was conducted on a dialog corpus of a virtual psychological assistant designed to provide psychological support. The effectiveness of the methods was assessed using quantitative metrics for intent classification and response selection, as well as qualitative expert evaluation of response adequacy. The obtained results demonstrate a significant advantage of contextualized models, particularly SimSCE-BERT, over traditional approaches such as word2vec and the base BERT. It is shown that the use of modern methods for computing semantic distance contributes to improving both the technical performance of chatbots and the user-perceived quality of interaction, which is critically important for scalable systems in the field of psychological assistance and other applied domains.

Keywords: semantic similarity, natural language processing, chatbots, vector embeddings, BERT, psychological assistant.

Вступ

Застосування вимірювання семантичної відстані між фразами та текстами є одним із базисів завдань обробки природної мови (NLP), що дозволяє машинам кількісно оцінювати взаємозв'язок або схожість значення між текстовими одиницями. Основна потреба у вимірюванні семантичної відстані виникає з його ролі у покращенні розуміння тексту. Традиційні методи оцінки подібності текстів часто спираються на лексичну збіжність або спільні ключові слова, що може бути недостатньо для відображення справжніх семантичних відносин з урахуванням випадків синонімії, полісемії або омонімії [1]. Метрики семантичної відстані вирішують ці обмеження, фокусуючись на глибинному значенні, а не на поверхневих лексичних формах [2].

В завданні інформаційного пошуку вимірювання семантичної відстані дозволяє пошуковим системам знаходити документи, які семантично релевантні запиту, навіть якщо вони не мають точних співпадінь за ключовими словами. Це покращує релевантність і повноту результатів пошуку [3]. Аналогічно, в системах питання-відповідь семантична відстань допомагає ідентифікувати відповіді, які правильно адресують запитання користувача, незалежно від формулювання [4].

Класифікація та кластеризація текстів також значною мірою виграють від вимірювання семантичної відстані при роботі із великими наборами даних та управлінні інформацією. Для виявлення плагіату це допомагає виявляти випадки, коли семантичний зміст було скопійовано, навіть якщо формулювання було змінено. [3, 5]

У більш складних застосуваннях NLP, таких як машинний переклад та діалогові системи, семантична відстань відіграє важливу роль у забезпеченні того, щоб перекладений або згенерований текст зберігав оригінальне значення та намір. Наприклад, у машинному перекладі оцінка семантичної відстані між вихідним та цільовим реченням допомагає оцінити якість і точність перекладу.

Чат-боти є комплексним інтерфейсом взаємодії із мовними моделями, що спрямовані на вирішення задач різного роду. У даній роботі пропонується розглянути ефективність семантичного аналізу текстів на основі чат-боту психолога, що є результатом розробки Інституту проблем штучного інтелекту МОН України і НАН України у співпраці із Інститутом психології імені Г.С. Костюка НАПН України [6].

Огляд підходів до обчислення семантичної відстані

Сучасні методи вимірювання семантичної відстані між фразами та текстами переважно спираються на передові підходи, зокрема ті, що базуються на нейронних вбудовуваннях (embeddings) та моделях глибинного навчання [7, 8]. Ці підходи мають на меті кількісно оцінити семантичну подібність або пов'язаність між текстовими одиницями, виходячи за межі поверхневого лексичного перетину для захоплення глибших контекстуальних і концептуальних значень [9]. Ефективність цих методів можна порівнювати за їх здатністю точно відображати семантичні нюанси, обчислювальною ефективністю, масштабованістю та продуктивністю в різних прикладних завданнях обробки природної мови.

Базовою концепцією вимірювання семантичної відстані є використання вбудовувань слів, які представляють слова у вигляді щільних векторів у неперервному багатовимірному евклідовому просторі, де геометрична відстань корелює із семантичною подібністю [10, 11]. Ранні моделі, такі як Word2Vec, навчають ці вбудовування шляхом передбачення оточуючих слів або самого слова на основі його контексту [12]. Проте такі статичні вбудовування слів зазвичай призначають один вектор для кожного слова, не враховуючи полісемію та контекстно-залежну семантику [10, 13]. Наприклад, слово "bank" у англійській може означати фінансову установу або берег річки, і статичне вбудовування не здатне надійно розрізнити ці значення лише на основі

контексту, що графічно проілюстровано на рисунку 1 [14].

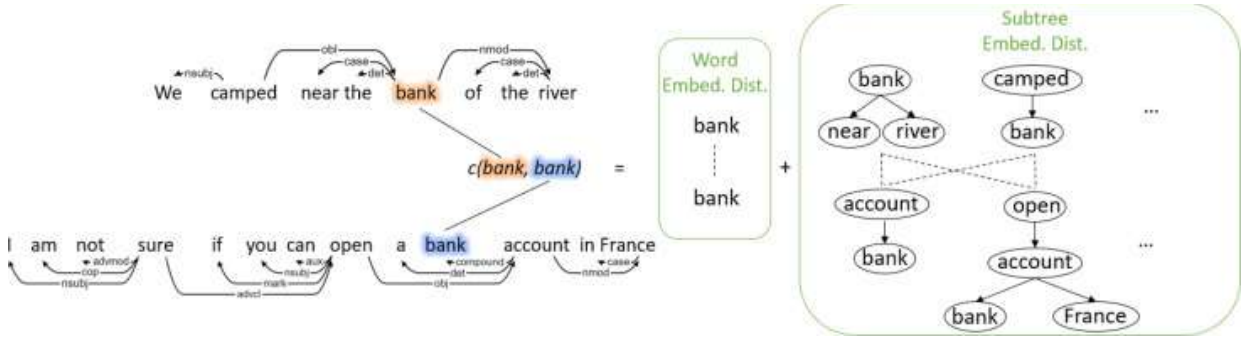

Рис. 1. Приклад ієрархічного аналізу семантичної відстані між словами “bank” у різному фразовому контексті [14]

Для подолання обмежень статичних вбудовувань слів дедалі більшої популярності набули контекстуалізовані вбудовування слів та вбудовування речень. Моделі на кшталт двонаправлених представлень енкодерів з трансформерів (Bidirectional Encoder Representations from Transformers. BERT) роблять значний крок уперед, оскільки на відміну від попередніх мовних моделей, попередньо навчають глибокі подання, одночасно враховуючи як лівий, так і правий контекст на всіх рівнях [15, 16, 17, 18, 19, 20]. Це дозволяє BERT генерувати контекстно-залежні вбудовування, у яких векторне подання слова змінюється залежно від його оточення в реченні. При застосуванні до задачі вимірювання текстової подібності BERT і його варіанти перетворюють речення або фрази у щільні векторні подання, після чого семантична відстань зазвичай вимірюється за допомогою косинусної подібності між цими векторами [16, 20].

Поширеною технікою є використання BERT для незалежної обробки двох вхідних речень із генерацією векторних подань для кожного з них. Ці вбудовування зазвичай усереднюються або агрегуються (pooling) для формування єдиного вектора на рівні речення, після чого обчислюється їх косинусна подібність для визначення семантичної пов’язаності [20]. Схема на рисунку 2 ілюструє цей процес, де Речення 1 та Речення 2 подаються на вхід моделі BERT, агрегуються для отримання векторів речень u та v, а потім обчислюється їх косинусна подібність.

Незважаючи на можливості BERT, його безпосереднє застосування для вимірювання подібності речень є обчислювально затратним, оскільки для кожного порівняння пар речень необхідно пропускати обидва речення через нейромережу [17]. Це призвело до розробки моделей на кшталт Sentence-BERT (SBERT), які здійснюють тонке налаштування BERT із використанням сіамських і триплетних мережевих структур для генерації семантично інформативних вбудовувань речень, що можуть ефективно порівнюватися за допомогою косинусної подібності. Такий підхід забезпечує ефективний пошук семантичної подібності, за якого вбудовування речень попередньо обчислюються один раз, а потім швидко порівнюються [17, 21].

Інші передові методи інтегрують додаткову лінгвістичну інформацію для покращення вимірювання семантичної подібності. Наприклад, SynWMD (Syntax-aware Word Mover’s Distance) удосконалює оригінальну метрику Word Mover’s Distance (WMD) шляхом урахування синтаксичної інформації з дерев розбору та важливості слів, що є критично важливим для оцінювання подібності речень [22]. Це усуває обмеження традиційної WMD, яка не враховує внутрішню контекстуальну та структурну інформацію в межах речення.

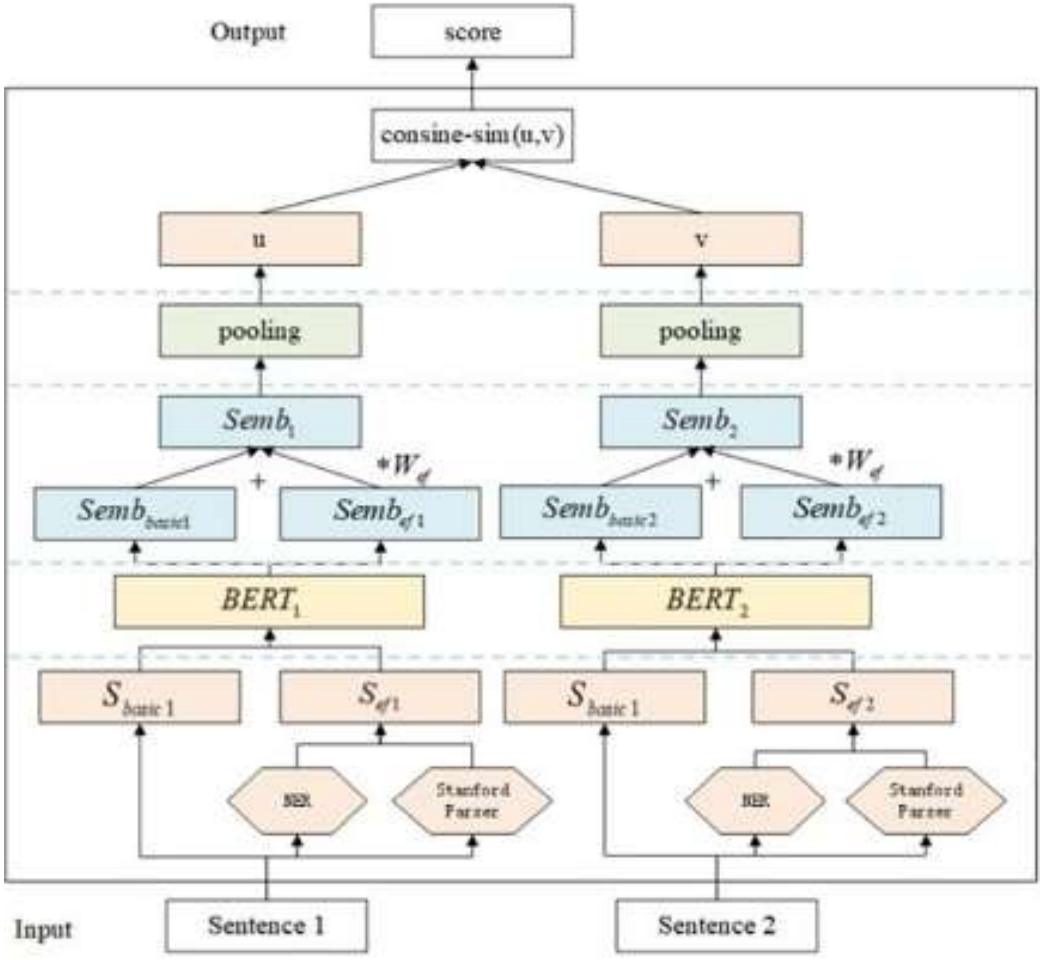

Рис. 2. Схема порівняння текстів за косинусною подібністю на основі векторного подання BERT [20]

Діаграма на рисунку 3 ілюструє порівняльну продуктивність різних методів вбудовування. Показано, що SimSCE-BERT зазвичай досягає менших середніх косинусних відстаней, що вказує на вищу подібність на різних наборах даних Semantic Textual Similarity (STS12, STS13, STS14, STS15, STS-B) порівняно з BERT (усереднення першого та останнього шарів) і word2vec. Це свідчить про те, що моделі, спеціально розроблені для вбудовування речень або ті, що враховують синтаксичну обізнаність, мають тенденцію формувати більш семантично узгоджені подання.

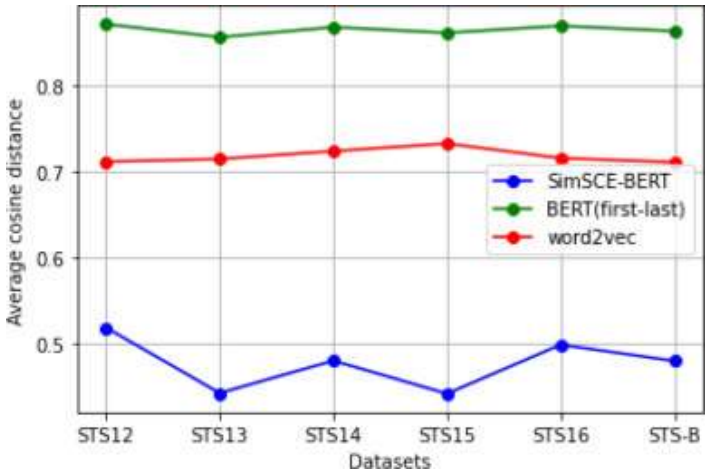

Рис. 3. Порівняння обчислень косинусної подібності наборів текстів векторними вбудовуваннями SimSCE-BERT, BERT та word2vec [14]

Крім того, дослідники вивчають методи інтеграції дрібнозернистої семантичної інформації (fine-grained) за межами рівня вбудовування речень. Наприклад, метод вимірювання подібності коротких текстів китайською мовою поєднує семантику на рівні речень і фраз, розв'язуючи проблеми синонімів і близьких за значенням слів, які можуть спотворювати вимірювання подібності [23]. Аналогічно цьому, контрастне навчання речень на основі оптимального транспорту безпосередньо описує відстань між реченнями як зважену суму відстаней між контекстуалізованими токенами, забезпечуючи більш інтерпретовану семантичну текстову подібність [24]. Використання механізмів уваги в сіамських архітектурах BERT також дозволяє надавати різну вагу окремим словам, що сприяє точнішому семантичному поданню [17].

Аналіз показників ефективності методів

Вимірювання семантичної відстані є критично важливим компонентом для чат-ботів, оскільки воно дозволяє їм розуміти наміри користувачів, генерувати релевантні відповіді та підтримувати зв'язні діалоги. Сформулюємо ключові характеристики, від яких залежить ефективність цих методів у застосуваннях чат-ботів.

Однією з основних характеристик, необхідних для функціонування чат-ботів, є *точність семантичного подання* [21, 25]. Чат-боти потребують методів, здатних точно захоплювати глибинний зміст запитів користувачів і потенційних відповідей. Це виходить за межі простого зіставлення ключових слів, оскільки семантично подібні фрази можуть використовувати повністю різну лексику [26]. Наприклад, запити користувача «Як поповнити баланс мобільного телефону?» та «Де я можу додати гроші на рахунок телефону?» мають бути розпізнані як семантично еквівалентні. Для оцінювання точності зазвичай використовуються еталонні набори даних STS, які зіставляють вихідні результати моделей із людськими оцінками подібності [25]. Наприклад, моделі S-BERT тонко налаштовані на задачах регресії для пар речень, тому досягли найкращих на сьогодні результатів у STS, формуючи вбудовування з високою кореляцією з людськими оцінками [17].

Обчислювальна ефективність є ще однією критично важливою характеристикою для чат-ботів, особливо з огляду на потребу у взаємодії в реальному часі [17]. Традиційні підходи, що використовують великі попередньо навчені мовні моделі, можуть бути обчислювально затратними. Обчислення семантичної подібності між запитом користувача та великим набором потенційних відповідей чат-бота може включати мільйони обчислень інференсу, що призводить до значних затримок. Наприклад, пошук найбільш подібної пари серед 10000 речень із використанням стандартної моделі BERT може займати приблизно 65 годин [17]. Для розв'язання цієї проблеми були розроблені моделі на кшталт Sentence-BERT (SBERT). SBERT попередньо обчислює вбудовування речень в офлайн-режимі, що дозволяє виконувати швидкі обчислення косинусної подібності під час інференсу, суттєво зменшуючи обчислювальні витрати [26]. Це дає змогу чат-ботам швидко знаходити найбільш релевантні заготовлені відповіді або генерувати контекстуально доречні репліки без помітних затримок. Оптимізовані стратегії агрегації, такі як діагональний пулінг з використанням механізму уваги (diagonal attention pooling, Ditto), також сприяють підвищенню ефективності шляхом зважування слів за їх важливістю для формування якісніших вбудовувань речень [27].

Здатність працювати з контекстом і полісемією є надзвичайно важливою для чат-ботів, щоб уникати хибних інтерпретацій. Природна мова насичена словами з кількома значеннями (полісемією), які змінюються залежно від контексту. Наприклад, слово «apple» може означати фрукт або технологічну компанію. Ранні статичні вбудовування слів не здатні впоратись з цією проблемою, призначаючи одному слову «apple» єдиний вектор незалежно від контексту його використання [14]. Контекстуалізовані вбудовування слів, отримані з моделей на кшталт BERT, генерують різні векторні подання одного й того самого слова залежно від оточуючих слів у реченні [28]. Це дозволяє чат-ботам точно розрізняти різні значення слів, що

призводить до більш точних обчислень семантичної відстані та доцільніших відповідей. Спеціалізовані методи, такі як SynWMD, додатково підсилюють цей ефект шляхом урахування інформації з синтаксичних дерев розбору та важливості слів, усуваючи обмеження традиційної метрики, яка не враховує внутрішню контекстуальну та структурну інформацію в межах речення.

Масштабованість також є критично важливою, оскільки чат-боти часто взаємодіють із великими базами знань або широкими колами користувачів, що потребує семантичного порівняння великих обсягів тексту. Обраний метод повинен бути здатним ефективно обробляти та порівнювати зростаючі масиви текстових даних без пропорційного збільшення обчислювальних ресурсів або погіршення продуктивності [13]. Хоча тонке налаштування попередньо навчених моделей може забезпечувати високу точність, воно не завжди добре масштабується для безперервного навчання або динамічного оновлення баз знань. Несупервізовані або напівсупервізовані методи, які потребують мінімальних обчислювальних ресурсів, зокрема ті, що спираються на композиційну семантику фраз, є привабливими з точки зору масштабування в умовах обмежених ресурсів [29].

Інтерпретованість є новою характеристикою, що набуває дедалі більшого значення для складних систем чат-ботів. Хоча моделі глибинного навчання можуть досягати високої точності, їх природа «чорної скриньки» часто ускладнює розуміння причин отримання певного показника семантичної подібності. Для критично важливих застосувань розуміння того, чому саме два тексти вважаються подібними або відмінними, допомагає в налагодженні систем, покращенні якості моделей і підвищенні довіри користувачів. Методи, які явно описують відстань між реченнями як зважену суму відстаней між контекстуалізованими токенами, наприклад підходи на основі контрастивного навчання речень із використанням оптимального транспорту, забезпечують більш інтерпретовану семантичну текстову подібність, висвітлюючи внесок окремих слів у загальний показник подібності [24, 30].

Переносимість і узагальнюваність є важливими для чат-ботів, призначених для роботи в кількох доменах або мовах. Надійний метод вимірювання семантичної відстані повинен демонструвати стабільну продуктивність у різних прикладних сферах, таких як обслуговування клієнтів або технічна підтримка, а також у різних мовах без потреби у масштабному повторному навчанні чи використанні великих доменно-специфічних наборів даних. Багатомовні вбудовування речень, здатні представляти численні мови, демонструють високу переносимість для zero-shot міжмовних завдань, дозволяючи чат-боту, навченому однією мовою, розуміти запити іншою мовою [31]. Врахування даної характеристики суттєво знижує витрати ресурсів на розгортання чат-ботів у багатомовних середовищах.

Емпіричне дослідження

Метою цього дослідження є проведення порівняльної оцінки методів семантичної подібності у рамках конвеєра обробки діалогів AI-орієнтованого чат-бота психологічної підтримки [6]. Порівняння у даній роботі робиться між розглянутими вище моделями word2vec, BERT та SimSCE-BERT й зосереджене на здатності моделей коректно вловлювати семантичний намір висловлювань користувачів і добирати контекстуально відповідні відповіді з підготовленої бази знань. Це завдання є критично важливим для чат-ботів, що функціонують у сфері ментального здоров’я, де запити користувачів часто є короткими, емоційно насиченими, імпліцитно сформульованими та семантично неоднозначними.

Дослідження проводилось з використанням експериментального діалогового корпусу, отриманого за результатами дослідної експлуатації платформи віртуального психологічного асистента. Корпус складається з 504 анонімізованих текстових взаємодій користувачів та відповідних категорій відповідей, верифікованих експертами. Кожен діалоговий зразок містить: висловлювання користувача (1–3 речення), анотовану мітку психологічного наміру, еталонну відповідь або кластер відповідей з

бази знань чат-бота. Набір даних було поділено на навчальну (70%), валідаційну (15%) та тестову (15%) підмножини.

Ефективність моделей оцінювалась за наступними метриками: точність класифікації намірів; значення Precision, Recall та F1-score (усереднені за макросхемою); середній взаємний ранг (Mean Reciprocal Rank, MRR) для відбору відповідей; якісна експертна оцінка адекватності відповідей, проведена професійними психологами. Усереднені результати обчислень наведені у таблиці 1.

Модель	Точність	F1-score	MRR
word2vec	0.62	0.59	0.54
BERT	0.78	0.76	0.72
SimSCE-BERT	0.86	0.85	0.83

Таблиця 1. Усереднені метрики оцінки моделей векторних вбудовувань для експериментального діалогового корпусу віртуального психологічного асистента

Модель word2vec демонструє обмежену стійкість під час обробки емоційно імпліцитних або метафоричних висловлювань, часто не здатна розрізняти семантично споріднені, але контекстуально різні вирази. Вбудовування BERT значно покращують семантичне представлення завдяки контекстуалізації, однак продуктивність знижується для коротких, недостатньо специфікованих висловлювань – поширеної характеристики комунікації психологічного дистресу. SimSCE-BERT демонструє найвищу семантичну стабільність, точно групуючи перефразовані вирази та зберігаючи семантичну узгодженість навіть для мінімалістичних користувацьких запитів (наприклад, «я почуваюся порожнім», «усе здається безглуздим»). За результатами аналізу експертів-оцінювачів встановлено, що відповіді, відібрані з використанням вбудовувань SimSCE-BERT, мали вищий рівень сприйманої емпатії та релевантності, особливо на ранніх етапах психологічної підтримки. Перевага SimSCE-BERT може бути пояснена оптимізацією для семантичної подібності на рівні речень та зниженою чутливістю до лексичної варіативності.

Дане дослідження обмежується текстовими взаємодіями та не враховує мультимодальні входи, такі як голосові або мімічні сигнали. Подальші дослідження будуть зосереджені на мультимодальних семантичних вбудовуваннях, доменно-специфічному донавчанні на україномовних психологічних корпусах та лонгітюдному оцінюванні результатів для користувачів.

Висновки

Потреба в більш надійних та контекстуально обізнаних методах для точного вимірювання семантичної відстані лишається у той час, як застосування ШІ стають дедалі складнішими та потребують глибшого лінгвістичного розуміння.

Для ефективного впровадження чат-ботів методи вимірювання семантичної відстані мають забезпечувати баланс між високою точністю відображення семантичного змісту, обчислювальною ефективністю для взаємодії в реальному часі, надійною обробкою контексту й полісемії, масштабованістю для роботи з великими наборами даних та, дедалі частіше, інтерпретованістю для розуміння процесів ухвалення рішень, а також високою переносимістю між доменами та мовами. Перехід від статичних вбудовувань слів до динамічних, контекстуалізованих і спеціалізованих вбудовувань речень, зокрема на основі трансформерних архітектур, відображає безперервні зусилля з оптимізації цих характеристик для практичних застосувань обробки природної мови, таких як чат-боти.

Емпіричне дослідження демонструє, що SimSCE-BERT забезпечує статистично та якісно значуще покращення порівняно з word2vec і BERT для завдань семантичної подібності в чат-ботах психологічної підтримки. Його інтеграція підвищує як технічну продуктивність, так і сприйману якість відповідей, що робить його перспективним рішенням для масштабованих систем допомоги у сфері ментального здоров’я.

Література

1. Wang, Y., Xue, T., & Yang, X. (2025). Exploring the relationship between features calculated from contextual embeddings and EEG band power during sentence reading in Chinese. Frontiers in Neuroscience, 19. https://doi.org/10.3389/fnins.2025.1656519

2. Peng Ding, P. D., Peng Ding, D. L., Dan Liu, Z. Z., Zhiyuan Zhang, J. H., & Jie Hu, N. L. (2022). A Novel Discrimination Structure for Assessing Text

Semantic Similarity. Internet Technology Journal, 23(4), 709–717.
https://doi.org/10.53106/160792642022072304006

3. Wang, J., & Dong, Y. (2020). Measurement of Text Similarity: A Survey. Information, 11(9), 421. https://doi.org/10.3390/info11090421

4. Dhagat, R., Rawal, A., & Soni, S. (2022). Comparative Evaluation of Semantic Similarity Upon Sentential Text of Varied (Generic) Lengths. In Lecture Notes in Electrical Engineering (pp. 107–122). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0284-0_9

5. Zhou, S., Xu, X., Liu, Y., Chang, R., & Xiao, Y. (2019). Text Similarity Measurement of Semantic Cognition Based on Word Vector Distance Decentralization With Clustering Analysis. IEEE Access, 7, 107247–107258.
https://doi.org/10.1109/access.2019.2932334

6. Шевченко, А. І., Панок, В. Г., Шевцов, А. Г., Слюсар, В. І., Малий, Р. І., Єрошенко, Т. В., & Назар, М. М. (2024). Розробка віртуального психологічного асистента зі штучним інтелектом у сфері охорони здоров'я. Клінічна та профілактична медицина, (8), 15-27. https://doi.org/10.31612/2616-4868.8.2024.02

7. Sharma, K. (2023). 30 Years of Research on Semantic Similarity Measurement. Center for Open Science. https://doi.org/10.31219/osf.io/qpb6d

8. Wang, Y. (2022). A Survey on Efficient Processing of Similarity Queries over Neural Embeddings (Version 1). arXiv.
https://doi.org/10.48550/ARXIV.2204.07922

9. Zhou, S., Xu, X., Liu, Y., Chang, R., & Xiao, Y. (2019). Text Similarity Measurement of Semantic Cognition Based on Word Vector Distance Decentralization With Clustering Analysis. IEEE Access, 7, 107247–107258.
https://doi.org/10.1109/access.2019.2932334

10. Colla, D., Mensa, E., & Radicioni, D. P. (2020). Novel metrics for computing semantic similarity with sense embeddings. Knowledge-Based Systems, 206, 106346. https://doi.org/10.1016/j.knosys.2020.106346

11. Pan, J.-S., Wang, X., Yang, D., Li, N., Huang, K., & Chu, S.-C. (2024). Flexible margins and multiple samples learning to enhance lexical semantic similarity. Engineering Applications of Artificial Intelligence, 133, 108275. https://doi.org/10.1016/j.engappai.2024.108275

12. der Brück, T. vor, & Pouly, M. (2024). Estimating Text Similarity based on Semantic Concept Embeddings. arXiv. https://doi.org/10.48550/ARXIV.2401.04422

13. Chang, H.-S., Agrawal, A., & McCallum, A. (2021). Extending Multi-Sense Word Embedding to Phrases and Sentences for Unsupervised Semantic Applications (Version 2). arXiv.
https://doi.org/10.48550/ARXIV.2103.15330

14. Wei, C., Wang, B., & Jay Kuo, C.-C. (2023). Synwmd: Syntax-aware word Mover's distance for sentence similarity evaluation. Pattern Recognition Letters, 170, 48–55.
https://doi.org/10.1016/j.patrec.2023.04.012

15. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Proceedings of the 2019 Conference of the North, 4171–4186. https://doi.org/10.18653/v1/n19-1423

16. Xiao, Z., Ning, X., & Duritan, M. J. M. (2025). BERT-SVM: A hybrid BERT and SVM method for semantic similarity matching evaluation of paired short texts in English teaching. Alexandria Engineering Journal, 126, 231–246.
https://doi.org/10.1016/j.aej.2025.04.061

17. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (Version 1). arXiv.
https://doi.org/10.48550/ARXIV.1908.10084

18. Xu, Y., Tian, J., Tang, M., Tao, L., & Wang, L. (2024). Document-level relation extraction with entity mentions deep attention. Computer Speech & Language, 84, 101574.
https://doi.org/10.1016/j.csl.2023.101574

19. Liu, N., Hu, J., & Liang, W. (2023). MIFINN: A novel multi-information fusion and interaction neural network for aspect-based sentiment analysis. Knowledge-Based Systems, 280, 110983.
https://doi.org/10.1016/j.knosys.2023.110983

20. Wang, T., Shi, H., Liu, W., & Yan, X. (2022). A joint FrameNet and element focusing Sentence-BERT method of sentence similarity computation. Expert Systems with Applications, 200, 117084. https://doi.org/10.1016/j.eswa.2022.117084

21. Herbold, S. (2023). Semantic similarity prediction is better than other semantic similarity measures. arXiv.
https://doi.org/10.48550/ARXIV.2309.12697

22. Wei, C., Wang, B., & Kuo, C.-C. J. (2022). Synwmd: Syntax-Aware Word Mover's Distance for Sentence Similarity Evaluation. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4145635

23. Shen, Z., & Xiao, Z. (2024). A Chinese Short Text Similarity Method Integrating Sentence-Level and Phrase-Level Semantics. Electronics, 13(24), 4868.
https://doi.org/10.3390/electronics13244868

24. Lee, S., Lee, D., Jang, S., & Yu, H. (2022). Toward Interpretable Semantic Textual Similarity via Optimal Transport-based Contrastive Sentence Learning. arXiv.
https://doi.org/10.48550/ARXIV.2202.13196

25. Li, R., Cheng, L., Wang, D., & Tan, J. (2023). Siamese BERT Architecture Model with attention mechanism for Textual Semantic Similarity. Multimedia Tools and Applications, 82(30), 46673–46694.
https://doi.org/10.1007/s11042-023-15509-4

26. Pang, S., Yao, J., Liu, T., Zhao, H., & Chen, H. (2020). A Text Similarity Measurement Based on Semantic Fingerprint of Characteristic Phrases. Chinese Journal of Electronics, 29(2), 233–241.
https://doi.org/10.1049/cje.2019.12.011

27. Chen, Q., Wang, W., Zhang, Q., Zheng, S., Deng, C., Yu, H., Liu, J., Ma, Y., & Zhang, C. (2023). Ditto: A Simple and Efficient Approach to Improve Sentence Embeddings. arXiv.
https://doi.org/10.48550/ARXIV.2305.10786

28. Zhou, K., Ethayarajh, K., & Jurafsky, D. (2021). Frequency-based Distortions in Contextualized Word Embeddings (Version 1). arXiv.
https://doi.org/10.48550/ARXIV.2104.08465

29. Wang, Z., Dou, J., & Zhang, Y. (2022). Unsupervised Sentence Textual Similarity with Compositional Phrase Semantics (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2210.02284

30. Opitz, J., & Frank, A. (2022). SBERT studies Meaning Representations: Decomposing Sentence Embeddings into Explainable Semantic Features (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2206.07023

31. Soricut, R., & Ding, N. (2016). Multilingual Word Embeddings using Multigraphs (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1612.04732

References

1. Wang, Y., Xue, T., & Yang, X. (2025). Exploring the relationship between features calculated from contextual embeddings and EEG band power during sentence reading in Chinese. Frontiers in Neuroscience, 19. https://doi.org/10.3389/fnins.2025.1656519

2. Peng Ding, P. D., Peng Ding, D. L., Dan Liu, Z. Z., Zhiyuan Zhang, J. H., & Jie Hu, N. L. (2022). A Novel Discrimination Structure for Assessing Text Semantic Similarity. Internet Technology Journal, 23(4), 709–717. https://doi.org/10.53106/160792642022072304006

3. Wang, J., & Dong, Y. (2020). Measurement of Text Similarity: A Survey. Information, 11(9), 421. https://doi.org/10.3390/info11090421

4. Dhagat, R., Rawal, A., & Soni, S. (2022). Comparative Evaluation of Semantic Similarity Upon Sentential Text of Varied (Generic) Lengths. In Lecture Notes in Electrical Engineering (pp. 107–122). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0284-0_9

5. Zhou, S., Xu, X., Liu, Y., Chang, R., & Xiao, Y. (2019). Text Similarity Measurement of Semantic Cognition Based on Word Vector Distance Decentralization With Clustering Analysis. IEEE Access, 7, 107247–107258. https://doi.org/10.1109/access.2019.2932334

6. Shevchenko, A. I., Panok, V. G., Shevtsov, A. G., Slyusar, V. I., Maly, R. I., Eroshenko, T. V., & Nazar, M. M. (2024). Development of a virtual psychological assistant with artificial intelligence in the field of health care. Clinical and Preventive Medicine, (8), 15-27. https://doi.org/10.31612/2616-4868.8.2024.02

7. Sharma, K. (2023). 30 Years of Research on Semantic Similarity Measurement. Center for Open Science. https://doi.org/10.31219/osf.io/qpb6d

8. Wang, Y. (2022). A Survey on Efficient Processing of Similarity Queries over Neural Embeddings. arXiv. https://doi.org/10.48550/ARXIV.2204.07922

9. Zhou, S., Xu, X., Liu, Y., Chang, R., & Xiao, Y. (2019). Text Similarity Measurement of Semantic Cognition Based on Word Vector Distance Decentralization With Clustering Analysis. IEEE Access, 7, 107247–107258. https://doi.org/10.1109/access.2019.2932334

10. Colla, D., Mensa, E., & Radicioni, D. P. (2020). Novel metrics for computing semantic similarity with sense embeddings. Knowledge-Based Systems, 206, 106346. https://doi.org/10.1016/j.knosys.2020.106346

11. Pan, J.-S., Wang, X., Yang, D., Li, N., Huang, K., & Chu, S.-C. (2024). Flexible margins and multiple samples learning to enhance lexical semantic similarity. Engineering Applications of Artificial Intelligence, 133, 108275. https://doi.org/10.1016/j.engappai.2024.108275

12. der Brück, T. vor, & Pouly, M. (2024). Estimating Text Similarity based on Semantic Concept Embeddings. arXiv. https://doi.org/10.48550/ARXIV.2401.04422

13. Chang, H.-S., Agrawal, A., & McCallum, A. (2021). Extending Multi-Sense Word Embedding to Phrases and Sentences for Unsupervised Semantic Applications. arXiv. https://doi.org/10.48550/ARXIV.2103.15330

14. Wei, C., Wang, B., & Jay Kuo, C.-C. (2023). Synwmd: Syntax-aware word Mover's distance for sentence similarity evaluation. Pattern Recognition Letters, 170, 48–55. https://doi.org/10.1016/j.patrec.2023.04.012

15. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Proceedings of the 2019 Conference of the North, 4171–4186. https://doi.org/10.18653/v1/n19-1423

16. Xiao, Z., Ning, X., & Duritan, M. J. M. (2025). BERT-SVM: A hybrid BERT and SVM method for semantic similarity matching evaluation of paired short texts in English teaching. Alexandria Engineering Journal, 126, 231–246. https://doi.org/10.1016/j.aej.2025.04.061

17. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv. DOI: 10.48550/ARXIV.1908.10084

18. Xu, Y., Tian, J., Tang, M., Tao, L., & Wang, L. (2024). Document-level relation extraction with entity mentions deep attention. Computer Speech & Language, 84, 101574. https://doi.org/10.1016/j.csl.2023.101574

19. Liu, N., Hu, J., & Liang, W. (2023). MIFINN: A novel multi-information fusion and interaction neural network for aspect-based sentiment analysis. Knowledge-Based Systems, 280, 110983. https://doi.org/10.1016/j.knosys.2023.110983

20. Wang, T., Shi, H., Liu, W., & Yan, X. (2022). A joint FrameNet and element focusing Sentence-BERT method of sentence similarity computation. Expert Systems with Applications, 200, 117084. https://doi.org/10.1016/j.eswa.2022.117084

21. Herbold, S. (2023). Semantic similarity prediction is better than other semantic similarity measures. arXiv. https://doi.org/10.48550/ARXIV.2309.12697

22. Wei, C., Wang, B., & Kuo, C.-C. J. (2022). Synwmd: Syntax-Aware Word Mover's Distance for Sentence Similarity Evaluation. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4145635

23. Shen, Z., & Xiao, Z. (2024). A Chinese Short Text Similarity Method Integrating Sentence-Level and Phrase-Level Semantics. Electronics, 13(24), 4868. https://doi.org/10.3390/electronics13244868

24. Lee, S., Lee, D., Jang, S., & Yu, H. (2022). Toward Interpretable Semantic Textual Similarity via Optimal Transport-based Contrastive Sentence Learning (Version 2). arXiv.

https://doi.org/10.48550/ARXIV.2202.13196

25. Li, R., Cheng, L., Wang, D., & Tan, J. (2023). Siamese BERT Architecture Model with attention mechanism for Textual Semantic Similarity. Multimedia Tools and Applications, 82(30), 46673–46694. https://doi.org/10.1007/s11042-023-15509-4

26. Pang, S., Yao, J., Liu, T., Zhao, H., & Chen, H. (2020). A Text Similarity Measurement Based on Semantic Fingerprint of Characteristic Phrases. Chinese Journal of Electronics, 29(2), 233–241. https://doi.org/10.1049/cje.2019.12.011

27. Chen, Q., Wang, W., Zhang, Q., Zheng, S., Deng, C., Yu, H., Liu, J., Ma, Y., & Zhang, C. (2023). Ditto: A Simple and Efficient Approach to Improve Sentence Embeddings. arXiv. https://doi.org/10.48550/ARXIV.2305.10786

28. Zhou, K., Ethayarajh, K., & Jurafsky, D. (2021). Frequency-based Distortions in Contextualized Word Embeddings (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2104.08465

29. Wang, Z., Dou, J., & Zhang, Y. (2022). Unsupervised Sentence Textual Similarity with Compositional Phrase Semantics. arXiv. https://doi.org/10.48550/ARXIV.2210.02284

30. Opitz, J., & Frank, A. (2022). SBERT studies Meaning Representations: Decomposing Sentence Embeddings into Explainable Semantic Features (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2206.07023

31. Soricut, R., & Ding, N. (2016). Multilingual Word Embeddings using Multigraphs. arXiv. https://doi.org/10.48550/ARXIV.1612.04732

The article has been sent to the editors 18.12.25.
After processing 20.12.25.
Submitted for printing 30.12.25.