УДК 004.94:004.78 DOI: 10.30837/0135-1710.2022.178.004

К.Е. ПЕТРОВ, Є.К. ВОРОБЙОВ, І.В. КОБЗЕВ

СИНТЕЗ МОДЕЛІ КЛАСИФІКАЦІЇ ДІАЛОГОВИХ АКТІВ НА ОСНОВІ ВИКОРИСТАННЯ РЕКУРЕНТНИХ НЕЙРОННИХ МЕРЕЖ

Запропоновано математичну модель класифікації діалогових актів, яка дозволяє враховувати контекст попередніх висловлювань в рамках взаємодії з користувачем. Модель побудовано на основі використання апарату рекурентних нейронних мереж та механізму уваги. Наведено архітектуру штучної нейронної мережі. Проведено її навчання та тестування з використанням набору даних Switchboard Dialogue Act Corpus. Наведено результати комп'ютерного моделювання, які демонструють працездатність та ефективність запропонованої моделі.

1. Вступ

Проблема взаємодії людини з комп'ютером існує з моменту появи обчислювальної техніки. На початку «спілкуватися» з електронною обчислювальною машиною могли лише програмісти. Згодом, по мірі розширення сфери використання комп'ютерної техніки та збільшення масштабів її застосування, користувачі почали втягуватись в процес безпосередньої взаємодії з комп'ютером, що призвело до появи масової категорії прямих кінцевих користувачів, що працюють в діалоговому режимі. Таким чином, сформувалась проблема обробки природної мови (Natural Language Processing, NLP) та її транслювання у машинне представлення.

На теперішній час обробка природної мови є однією з найбільш досліджуваних проблем в галузі штучного інтелекту. Напрями цих досліджень достатньо різноманітні, зокрема, сюди можна віднести такі завдання, як машинний переклад; інформаційний пошук; реферування та анотування текстів; класифікація текстів, в тому числі тематичне моделювання; створення чат-ботів; аналіз тональності текстів; видобування знань з текстів; автоматична генерація текстів тощо.

Ще одним з актуальних завдань цього напряму є автоматизація процесу ведення (генерації) діалогів типу «машина-людина». Прикладами такої автоматизації можуть слугувати персональні мобільні асистенти, системи типу «запитання-відповідь», чат-боти тощо. Подібні системи дозволяють користувачу вирішувати певні задачі через спілкування з комп'ютером природною мовою у форматі діалогу.

Однією з ключових задач генерації діалогів є класифікація діалогових актів (Dialog Act, DA). Під DA розуміється висловлювання в контексті розмовного діалогу, яке виконує певну функцію в ньому. Іншими словами, акт певної репліки - це мета, яку хоче досягти автор, висловлюючи її. Прикладами таких актів можуть слугувати запити інформації, твердження, згода тощо. Розпізнавання та класифікація DA суттєво підвищує якість та релевантність згенерованих відповідей, що, в свою чергу, підвищує якість усього діалогу.

2. Аналіз сучасних досліджень у галузі класифікації діалогових актів

Діалогові системи можуть бути умовно поділені на два основні класи.

До першого з них відносяться системи, орієнтовані на завдання, які мають на меті допомогти користувачеві виконати певні задачі (наприклад, знайти продукти, забронювати помешкання або столик у ресторані тощо). Основний підхід при реалізації таких діалогових систем полягає у тому, щоб розглядати відповідь в діалозі як конвеєр. Система спочатку намагається зрозуміти повідомлення, що передає людина, представляє його як внутрішній стан, а потім виконує деякі дії відповідно до правил, враховуючи поточний стан діалогу, і, нарешті, дія перетворюється на свою поверхневу форму - природну мову.

До другого класу можна віднести системи, не орієнтовані на виконання задачі (також відомі як чат-боти), які взаємодіють з людиною, щоб забезпечити розумні відповіді та розваги. Зазвичай вони зосереджені на спілкуванні з людиною у відкритих доменах. Незважаючи на те, що такі системи, на перший погляд, виконують розважальні функції, вони домінують у багатьох реальних застосунках. Для реалізації таких систем застосовуються два основні підходи - генеративні методи, такі як Seq2Seq [1], які генерують належні відповіді під час розмови, і методи на основі пошуку, які навчаються вибирати відповіді для поточної розмови з певної бази знань.

Обидва класи систем мають вміти генерувати коректні та змістовні відповіді на репліки співрозмовника. Першим та одним з головних кроків цього процесу є розуміння та опрацювання попередньої репліки або декількох реплік. Під розумінням репліки мається на увазі її лексичний та синтаксичний розбір, знаходження її сенсу та віднесення до попередніх висловлювань у діалозі. Для покращення розуміння реплік також може бути застосована їх класифікація - тобто віднесення реплік до певного класу в залежності від їх мети та сенсу [2]. Групування висловлювань у класи дозволяє застосовувати певні шаблони генерації відповідей в залежності від цих класів.

Класифікація загальної мети висловлювань користувача в розмові (таких, наприклад, як відкрите запитання, висловлювання думки або запит думки), також відомих як DA, є ключовим кроком у розумінні природної мови для розмовних агентів. Хоча класифікація DA була детально вивчена в розмовах між людьми, вона недостатньо вивчена для нових автоматизованих розмовних агентів з довільними темами. Більше того, незважаючи на значний прогрес у класифікації DA на рівні висловлювань, повне розуміння висловлювань діалогу вимагає знання контексту розмови.

При вирішені задачі класифікації DA необхідно розглянути два попередніх етапи, перш ніж можна буде застосувати певну техніку класифікації. Перший етап полягає у виборі набору тегів, а другий - у визначенні ознак, які будуть використовуватися класифікаторами.

Визначення набору тегів DA є важливим та одночасно достатньо складним етапом, оскільки цей процес є результатом компромісу між трьома суперечливими вимогами. По-перше, теги DA повинні бути достатньо конкретними, щоб кодувати детальні характеристики цільового завдання; по-друге, бути достатньо загальними, щоб бути корисними для різних завдань, або, принаймні, стійкими до мінливості та розвитку цільового застосування; по-третє, бути чіткими та легко відокремлюваними, щоб максимізувати узгодження між людьми, що анотують репліки.

Дослідження DA призвело до кількох схем анотації. Серед них - розмітка в кількох шарах (DAMSL) [3], яка слугує основою для анотації двох довідкових корпусів, SwDA [4] та MRDA [5]. DAMSL розроблявся як універсальний набір. Теги в DAMSL мають декілька рівнів класифікації. На першому рівні всі теги поділяються на чотири основні класи: комунікативний статус (Communicative Status) - фіксує, чи є висловлювання зрозумілим і чи було його успішно завершено; інформаційний рівень (Information Level) - характеристика сенсового змісту висловлювання; функції, орієнтовані на перспективу (Forward Looking Function) - визначають, як поточне висловлювання обмежує майбутні переконання та дії учасників і впливає на дискурс та функції зворотного огляду (Backward Looking Function), тобто визначає, як поточне висловлювання пов'язане з попереднім дискурсом. Загалом, ці класи вважаються ортогональними, і можна побудувати приклади для будь-якої можливої їх комбінації.

Ще одна схема DIT++, яка призвела до створення стандарту ISO 24617-2 [6], пропонує організацію, що заснована на різних вимірах (наприклад, управління діалоговими ходами, управління соціальними зобов'язаннями тощо), кожен з яких містить різні DA. Загалом DIT++ пропонує виділити більше 100 DA, а DAMSL - 226 DA, які зазвичай групуються в 42 мітки. Але така велика кількість класів не призводить до підвищення ефективності вирішення задачі автоматичної класифікації. У значній кількості досліджень пропонують обмежити кількість класів, використовуючи або загальні метакласи, або найчастіше використовувані. Наприклад, у [7, 8], які базуються на використанні DAMSL, застосовується набір тегів, скорочений до 5 найпоширеніших класів (твердження, зворотна відповідь, думка, залишення, згода).

Зменшення набору тегів також можна зробити для пристосування класифікації до потреб конкретної доменної діалогової системи. Це, наприклад, пропонується у [9], де описується комунікація в контексті реагування на катастрофи за допомогою роботів. В ній пропонується використовувати конкретний набір метакласів ISO, адаптованих до потреб системи, об'єднавши 20 найкорисніших DA у 8 метакласів (зв'язок, повідомлення, твердження, запит, запитання, підтвердження, скасування, заперечення).

Розглянемо детальніше зміст другого етапу - виділення ознак для класифікації DA, тобто властивостей, які використовуються класифікаторами. Такі інформаційні властивості можна умовно поділити на декілька типів, описаних нижче.

Лексична інформація. Кожне висловлювання складається з послідовності слів. Як правило, DA висловлювання можна частково вивести зі списків слів, які утворюють це висловлю-

вання. Наприклад, запитання часто містять питальне слово, яке рідко зустрічається в інших класах DA. Лексичну інформацію зазвичай фіксують уніграми слів.

Синтаксична інформація, пов'язана з порядком слів у висловлюванні. Наприклад, у французькій та чеській мовах відносний порядок появи підмета і присудка може використовуватися для розрізнення декларацій та питань. N-грами слів часто використовуються при розпізнаванні DA для моделювання деякої локальної синтаксичної інформації.

Семантична інформація. DA також залежить від значень висловлювання та слів, які його складають. Прикладами можуть слугувати як широкі тематичні категорії, такі як «погода», «спорт», так і точні інтерпретації на основі фреймів, наприклад, «показати рейси з Лондона до Парижа 12 березня».

Просодія, і зокрема, мелодика висловлювання. Зазвичай запитання мають зростаючу мелодику в кінці висловлювання, тоді як висловлювання часто характеризуються дещо спадною мелодикою.

Контекст DA, тобто будь-який DA залежить від попередніх (і наступних). Найважливішим контекстом є попередні вислови. Наприклад, відповіді «так» або «ні», найімовірніше, будуть слідувати за запитанням. Послідовність DA також називається історією діалогу.

Тепер розглянемо детальніше методи, що використовуються для класифікації DA.

Основні типи підходів автоматичного розпізнавання DA, що розглядаються в літературі, можна широко класифікувати на баєсівські та небаєсівські підходи.

Баєсівські підходи ґрунтуються на баєсовому висновуванні [10]. Вони відносяться до найбільш розповсюджених та досліджених підходів. Розглянемо основні моделі, що використовують цей підхід.

Лексичні n-грам моделі [11] та n-грам моделі на основі послідовності реплік [12], в яких історія діалогу зазвичай моделюється за допомогою статистичної граматики дискурсу, яка представляє попередню ймовірність послідовності актів .

Приховані марковські моделі (ПММ) [13], де використовуються моделі n-го порядку, що означає, що кожен DA залежить від n попередніх DA (так само, як і для n-грам). Потім кожен стан ПММ моделює один DA, а спостереження відповідають ознакам рівня висловлювання. Імовірності переходу навчаються на проанотованому наборі даних. Розпізнавання DA здійснюється за допомогою деякого алгоритму динамічного програмування. Для моделювання історії діалогів успішно використовуються ПММ зі словами та просодичними ознаками. Так у [13] використовуються інтонаційні події та функції нахилу, такі як F0 (падіння/підйом), енергія, тривалість тощо. Запропонована модель досягає 64 % точності на корпусі DCIEM [14] з 12 класами DA. Існують також приклади поєднання ПММ з нейронними мережами [15], в якому був отриманий результат близько 76 % точності на іспанському корпусі CallHome.

Баєсові мережі. Приклад їх застосування для вирішення задачі розпізнавання DA наведено в [16]. Використовуються два типи ознак: ознаки висловлювання (слова у висловлюванні w_i) та ознаки контексту (попередній DA $C_{\tau-1}$). Автори порівнюють дві різні баєсові мережі, щоб розпізнати DA. Точність розпізнавання складає приблизно 64 % на підмножині корпусу MRDA і зі зменшеним набором класів.

Небаєсовські підходи також успішно використовуються в області розпізнавання DA, але вони не настільки популярні, як баєсовські. Прикладами таких підходів є дерева рішень, навчання на основі пам'яті та на основі трансформації, нейронні мережі, такі як багатошаровий перцептрон або мережі Кохонена тощо.

Моделі на основі дерев рішень (або дерев класифікації та регресії). У випадку розпізнавання DA рішення зазвичай стосуються особливостей висловлювання. Під час ухвалення кожного рішення щодо віднесення DA до відповідного класу реалізується процес порівняння значення деякої ознаки з пороговим.

Навчання на основі пам'яті (MBL) є застосуванням теорії міркувань на основі пам'яті в галузі машинного навчання. Ця теорія базується на припущенні, що можна обробляти новий зразок, порівнюючи його із збереженими представленнями попередніх зразків. Так у [17] розглядається використання MBL для автоматичної розмітки DA на корпусі SwDA [4], який складається зі спонтанних телефонних розмов між людьми. Автор експериментує з різною кількістю сусідів. Найкращий результат - близько 72 % точності з трьома сусідами.

Основна ідея навчання на основі трансформацій (TBL) полягає в тому, щоб почати з деякого простого рішення проблеми і застосувати перетворення для отримання кінцевого результату.

TBL можна застосувати до більшості задач класифікації, в тому числі для автоматичного розпізнавання DA. Наприклад, у [18] використовується TBL зі стратегією Монте-Карло на корпусі VERBMOBIL. Отримана точність класифікації DA становить близько 71 %.

Однією з найчастіше використовуваних моделей нейронної мережі в області розпізнавання DA є багатошаровий перцептрон (MLP). Так у [19] описується можливість використання набору бінарних функцій для навчання MLP. Ці функції обчислюються автоматично шляхом поєднання фразового аналізу на основі граматики та методів машинного навчання. Отримана точність розпізнавання DA близько 71 % для англійської та 69 % для німецької мов при використанні датасету NESPOLE.

Приклад використання мережі Кохонена для розпізнавання DA розглянуто в [20]. Автори використовують сім поверхневих ознак висловлювання: мовець, режим речення, наявність чи відсутність слів-маркерів відкритих запитань, наявність чи відсутність знаку питання тощо. Кожне висловлювання представлено шаблоном цих ознак, який кодується у двійковому форматі. Спочатку точна кількість класів DA невідома апріорі. Процес кластеризації переривається після того, як буде знайдено задану кількість кластерів.

Окремо слід розглянути підходи, що базуються на використанні рекурентних нейронних мереж (RNN).

Рекурентні нейронні мережі є дуже важливим варіантом нейронних мереж, які активно використовуються в обробці природної мови. Концептуально їх використання відрізняється від використання стандартної нейронної мережі, оскільки стандартним введенням в RNN є слово, а не вся вибірка, як у випадку стандартної нейронної мережі. Це дає мережі можливість працювати з реченнями різної довжини, чого неможливо досягти в стандартній нейронній мережі через її фіксовану структуру. Це також надає додаткову перевагу спільного використання функцій, засвоєних у різних позиціях тексту, які неможливо отримати в стандартній нейронній мережі.

Незважаючи на усі переваги рекурентних нейронних мереж, вони мають певні недоліки.

По-перше, RNN з класичною архітектурою здатні фіксувати залежності лише в одному напрямку мови. Так у випадку обробки природної мови передбачається, що слово, яке йде після, не впливає на значення слова, яке йде перед. Проте, враховуючи структури мов, можна стверджувати, що це твердження не завжди правильне.

По-друге, RNN також не дуже добре фіксує довгострокові залежності та проблему зникаючого градієнта [21].

Обидва ці обмеження породжують нові типи архітектур RNN, зокрема:

- вентильний рекурентний вузол (GRU) - це модифікація основного рекурентного блоку, яка допомагає фіксувати залежності на великій відстані, а також дуже допомагає у вирішенні проблеми зникаючого градієнта [21];

- двонаправлені рекурентні нейронні мережі (BRNN), здібні враховувати наслідки слова, написаного не тільки перед поточним словом, що дуже важливо при обробці природної мови.

Механізм уваги (Attention Mechanism) пропонується як рішення обмеження моделі кодер-декодер, яка кодує вхідну послідовність до одного вектору фіксованої довжини, з якого декодується вихідний результат в кожний момент часу [22]. Вважається, що ця проблема з'являється під час декодування довгих послідовностей, оскільки нейронній мережі важко справлятися з довгими реченнями, особливо тими, які довші за речення в навчальному корпусі.

Ще одним видом уваги є самоувага. Вона визначає увагу тієї ж послідовності. Замість того, щоб шукати асоціацію/вирівнювання послідовності введення-виведення, відшукуються оцінки між елементами послідовності.

Використання уваги може значно покращити результати моделі завдяки усуненню проблеми зникаючого градієнта, оскільки цей механізм забезпечує прямі зв'язки між станами кодера та декодера. Концептуально він діє подібно до пропуску зв'язків у згорткових нейронних мережах.

Іншою перевагою є зрозумілість. Перевіряючи розподіл ваг уваги, ми можемо отримати уявлення про поведінку моделі, а також зрозуміти її обмеження.

BERT (Bidirectional Encoder Representations from Transformers) - це модель представлення мови, розроблена дослідниками Google AI Language [23]. Ключовою технічною інновацією BERT є застосування популярної моделі двонаправленого навчання Transformer до мовного моделювання. Результати роботи показують, що мовна модель, яка навчається двосторонньо, може мати більш глибоке відчуття мовного контексту та потоку, ніж одно-

спрямовані мовні моделі. У [23] докладно описується нова техніка під назвою Masked LM (MLM), яка дозволяє проводити двонаправлене навчання в моделях, у яких раніше це було неможливо. Існує також ще одна популярна стратегія навчання - прогноз наступного речення (NSP).

Під час навчання моделі BERT, MLM і NSP навчаються разом з метою мінімізації комбінованої функції втрат двох стратегій.

BERT можна використовувати для різноманітних мовних завдань, додаючи лише невеликий шар до основної моделі.

Таким чином, задача класифікації DA в системах обробки природної мови є актуальною, про що також свідчить і велика кількість публікацій на цю тему.

Метою даного дослідження є розробка моделі класифікації DA, що може бути використана для автоматизації процесів в рамках розробки сучасних інтелектуальних діалогових систем.

Для досягнення поставленої мети необхідно вирішити такі основні задачі:

- розробити математичну модель класифікації DA на основі використання апарату рекурентних нейронних мереж;
- провести експериментальну перевірку працездатності та ефективності запропонованої моделі класифікації.

3. Постановка задачі дослідження

Формально, при вирішенні задачі класифікації DA розмову C можна представити як вхідну інформацію, яка являє собою послідовність висловлювань різної довжини $(u_1, u_2, ..., u_L)$. Кожне висловлювання u_i, $i = \overline{1, L}$, у свою чергу, є послідовністю слів різної довжини $(w_i^1, w_i^2, ..., w_i^{N_i})$ і має відповідну цільову мітку y_i, $i = \overline{1, L}$. Таким чином, кожна розмова (послідовність висловлювань) відображається на відповідну послідовність цільових міток $(y_1, y_2, ..., y_L)$, яка представляє DA, пов'язані з відповідними висловлюваннями. Тобто для розпізнання намірів користувача в ході розмови необхідно створити навчений на векторному представленні висловлювання класифікатор, який є одним із важливих складових реалізації процесу автоматичної генерації діалогів.

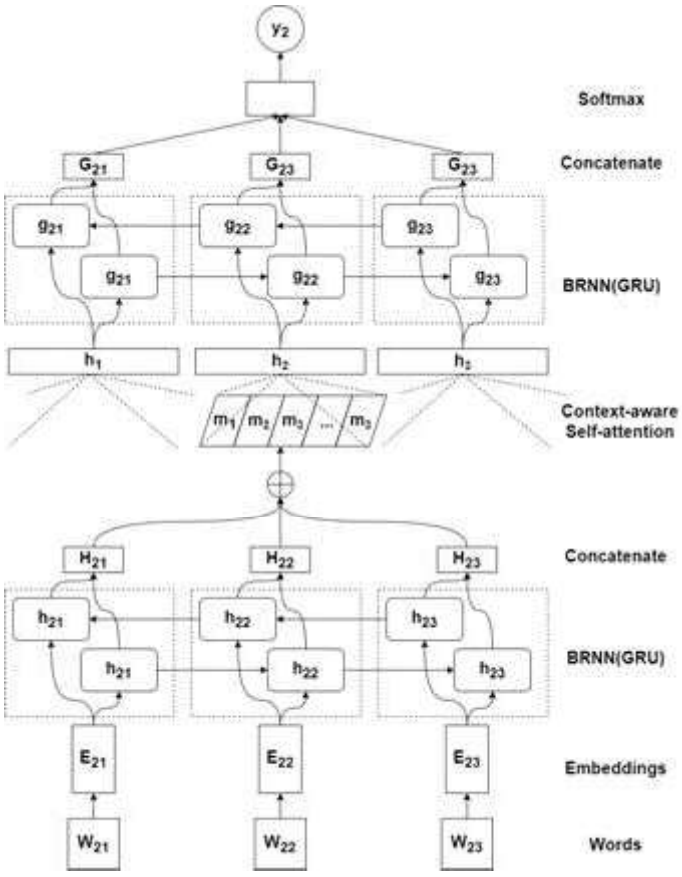

Рис. 1. Архітектура штучної нейронної мережі, що реалізує модель класифікації діалогових актів

4. Модель класифікації діалогових актів

Як показано вище, рекурентні нейроні мережі та їх модифікації дуже добре зарекомендували себе у вирішенні різноманітних задач обробки природної мови. Саме тому буде доцільно обрати саме їх як основу для синтезу моделі. Архітектуру моделі представлено на рис. 1.

Модель складається з трьох основних компонентів: RNN на рівні висловлювання, яка кодує інформацію у висловлюваннях на рівні слів та символів; контекстно-свідомий механізм самоуваги, який об'єднує репрезентації слів у репрезентації висловлювань; RNN рівня розмови - шар

класифікатора для визначення класу DA висловлювання, який об'єднує попередні та координує їх обчислення.

Розглянемо ці компоненти детальніше.

RNN рівня висловлювання. На цьому шарі слова вхідної репліки кодуються в так звані ембедінги - векторні представлення. Для цього використовується модель RoBERTa [24]. Ця модель є вдосконаленням моделі BERT командою Facebook AI Research. До основних відмінностей цієї моделі від BERT можна віднести спрощену систему навчання та значно збільшений корпус даних для навчання. RoBERTa демонструє від 2 до 20 % покращення результату порівняно з BERT в залежності від завдання та застосування.

Таким чином, кожне слово репліки кодується в ембедінг. Потім йде двонаправлений шар GRU (BGRU). Об'єднання прямих і зворотних вихідних даних BGRU генерує ембедінги висловлювання, які служать вхідним сигналом для контекстно-свідомого механізму самоуваги на рівні висловлювання, який вивчає остаточне представлення висловлювання.

Далі йде контекстно-свідомий шар з самоувагою. Самоуважні представлення кодують послідовність змінної довжини у фіксований розмір, використовуючи механізм уваги, який враховує різні позиції в послідовності. Тут використовується попередній прихований стан з RNN рівня висловлювання, який надає контекст розмови на даний момент і потім об'єднує його з прихованими станами всіх складових слів у висловлюванні в самоуважний кодер, який обчислює 2D-представлення кожного вхідного висловлювання.

Репліка u_i, яку можна представити як послідовність слів $(w_i^1, w_i^2, ..., w_i^n)$, за допомогою шару емебдінгів кодується в вектор розмірності d для кожного слова. Потім отримані вектори передаються до шару BGRU, приховані виходи якого об'єднуються для кожного кроку часу. Формально ці кроки можна представити формулами

$$\overrightarrow{h_l^j} = \overrightarrow{GRU}(w_i^j \cdot \overrightarrow{h_l^{j-1}});\ \overleftarrow{h_l^j} = \overleftarrow{GRU}(w_i^j \cdot \overleftarrow{h_l^{j+1}});\ h_i^j = concat(\overrightarrow{h_l^j}, \overleftarrow{h_l^j});\ H_i = (h_i^1, h_i^2, h_i^3, ..., h_i^n), \quad (1)$$

де H_i - n-й вихід BGRU.

Далі контекстні оцінки самоуваги S_i обчислюються за формулою

$$S_i = W_{s2} \cdot (W_{s1} \cdot H_i^T + W_{s3} \cdot \overrightarrow{g_{l-1}} + b), \quad (2)$$

де W_{s1}, W_{s2}, W_{s3} - матриці ваг відповідних розмірів; b - зміщення; $\overrightarrow{g_{l-1}}$ - прихований стан RNN рівня розмови.

Рівняння (2) можна розглядати як двошаровий MLP зі зміщенням b та W_{s1}, W_{s2}, W_{s3} як ваговими параметрами. Оцінки S_i відображаються в матрицю ймовірностей A_i за допомогою функції $A_i = soft\max(S_i)$.

Потім ця матриця ймовірностей A_i використовується для отримання двовимірного представлення M_i вхідного висловлювання шляхом використання прихованих станів GRU H_i відповідно до вагових коефіцієнтів уваги, заданих A_i, за формулою

$$M_i = A_i \cdot H_i. \quad (3)$$

Це двовимірне представлення потім проєктується на одновимірний ембедінг (позначається як h_i), використовуючи повнозв'язаний шар. Потім за допомогою RNN рівня розмови цей ембедінг перетворюється на g_i за формулами

$$\overrightarrow{g_l} = \overrightarrow{GRU}(h_i \cdot \overrightarrow{g_{l-1}});\ \overleftarrow{g_l} = \overleftarrow{GRU}(h_i \cdot \overleftarrow{g_{l+1}});\ g_i = concat(\overrightarrow{g_l}, \overleftarrow{g_l}). \quad (4)$$

Вектор g_i надає контекст розмови, який використовується для вивчення показників уваги та двовимірного представлення M_{i+1} для наступного висловлювання в розмові h_{i+1}.

Наприкінці використовуємо RNN рівня розмови.

Представлення висловлювання з попереднього кроку передається на рівень RNN рівня розмови, який є іншим двонаправленим шаром GRU, що використовується для кодування висловлювань у розмові. Приховані стани об'єднуються, щоб отримати остаточне представлення G_i кожного висловлювання, яке далі поширюється на шар класифікатора. Він, в свою чергу, складається з трьох повноз'єднаних шарів з функцією активації типу "нещільний випрямлений лінійний вузол". Перші два шари необхідні для зменшення розмірності мережі. Обраний датасет має 43 класи, тому вихід нейронної мережі має відповідну розмірність.

Розмірності ключових шарів нейронної мережі наведено в табл. 1.

Таблиця 1

Назва шару	Вхідна розмірність	Вихідна розмірність
RoBERTa Embeddings	256	768
BRNN	768	768*2
ContextAwareAttention	1536	1
BRNN	1	768*2
Linear	1536	256
Linear	256	128
Linear	128	43

5. Реалізація та навчання RNN для класифікації діалогових актів

Як набір даних для навчання та тестування моделі було обрано SwDA (Switchboard Dialogue Act Corpus) [4]. Цей набір складається з близько 2400 телефонних розмов між 543 абонентами (302 чоловіками та 241 жінкою) з усіх регіонів США. Автоматичний комп'ютеризований оператор обробляв дзвінки, надаючи абоненту відповідні записані підказки, вибираючи для участі в розмові іншу особу (абонента) та набираючи її номер, вводячи тему для обговорення та записуючи промову двох суб'єктів на окремі канали. Було представлено близько 70 тем, з яких близько 50 були найбільш розповсюджені. Вибір тем і тих, хто викликає, був обмежений таким чином, щоб, по-перше, два абоненти не говорили разом більше одного разу і, по-друге, ніхто не говорив більше одного разу на певну тему.

SwDA є модифікацією оригінального датасету, кожна розмова була проанотована з використанням набору тегів DAMSL. У кодуванні було використано 220 тегів; 130 з них зустрічалися менше ніж 10 разів кожен, тому 220 тегів було об'єднано у 42 більші класи.

Для реалізації моделі було використано наступний інструментарій: мова програмування Python, фреймворк машинного навчання з відкритим кодом PyTorch, оболонка для високопродуктивних обчислень PyTorch Lightning та бібліотека Pandas для попередньої обробки даних. Навчання моделі відбувалося з використанням платформ Kaggle та Google Colaboratory.

Графіки зміни точності та функції втрати представлено відповідно на рис. 2 та рис. 3. Для відображення загальної тенденції було застосовано згладжування.

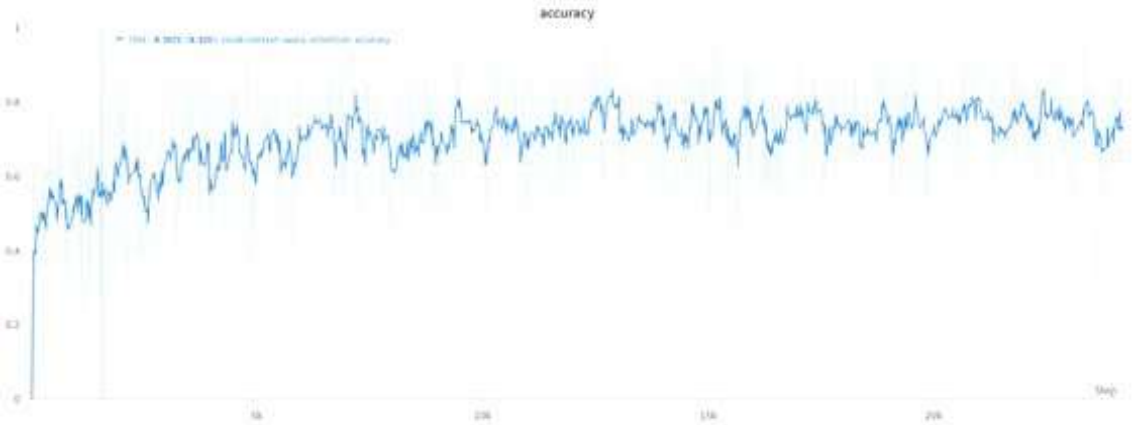

Рис. 2. Графік зміни точності моделі

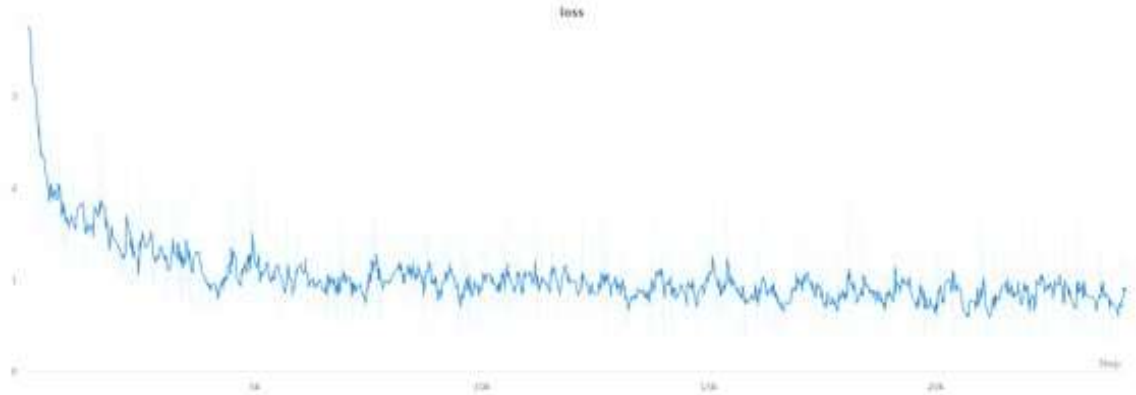

Рис. 3. Графік зміни функції втрат

Запропонована в роботі модель навчалася протягом 10 поколінь. Приблизний час навчання одного покоління з використанням прискорених графічних процесорів склав 2,5 години. В результаті було отримано максимальну точність 75,35 % для тестової вибірки та 76,62 % для валідаційної.

6. Обговорення результатів та висновки

Розпізнавання та класифікація DA дозволяє значно покращити якість генерації відповідей у діалогових системах.

В роботі розроблено математичну модель класифікації DA, яка базується на використанні апарату рекурентних нейронних мереж та механізму уваги. Її можна віднести до семантичних моделей, оскільки вона працює лише з текстовим представленням діалогу та не бере до уваги просодичні ознаки.

Було розроблено архітектуру відповідної RNN та проведено її навчання з використанням одного з найбільш розповсюджених для вирішення цієї задачі набору даних SwDA.

До основних недоліків запропонованої моделі можна віднести достатньо тривалий час навчання через її складність та велику кількість параметрів, а також відсутність можливості опрацювання просодичних ознак, що може вплинути на точність класифікації.

В результаті тестування моделі було отримано достатньо високі показники точності класифікації. Досягнута в роботі точність класифікації перевершує результати, представлені в [25] на 2,2 %, проте все ще менша, ніж в [26], де точність складає 81,3 %.

Таким чином, постає питання про подальше вдосконалення моделі для отримання кращих результатів. Серед перспективних шляхів подальших досліджень можна розглянути застосування моделей умовного випадкового поля та довгої короткочасної пам'яті (LSTM) [27] як класифікатора на останньому шарі запропонованої мережі.

Можна також дослідити інші моделі уваги, використання яких може вплинути на зміну ваги слів у кожному висловлюванні.

В перспективі було б доцільним спробувати об'єднати запропоновану модель з іншими моделями обробки природної мови. Наприклад, використати моделі розпізнавання іменованих сутностей чи аналізу емоційного забарвлення разом з класифікацією.

Література: 1. *Cho K., Merrienboer B., Gulcehre C., Bougares F., Schwenk H., Bengio Y.* Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Proceedings of the conference on empirical methods in natural language processing (EMNLP). 2014. https://doi.org/10.3115/v1/d14-1179. 2. *Воробйов Є. К., Петров К. Е.* Дослідження методів класифікації діалогових актів . Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління: тези доп. дванадцатої міжнар. науково-техн. конф. 2022. С. 14. 3. *Core M., Allen J.* Coding dialogs with the DAMSL annotation scheme. AAAI fall symposium on communicative action in humans and machines. 1997. P. 28-35. 4. *Jurafsky D., Shriberg E., Biasca D.* Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. 1997. 5. *Shriberg E., Dhillon R., Bhagat S., Ang J., Carvey H.* The ICSI meeting recorder dialog act (MRDA) corpus. Proceedings of the Human Language Technology Conference at the North American Chapter of the Association for Computational Linguistics. 2004. https://doi.org/10.21236/ada460980. 6. *Bunt H., Petukhova V., Traum D., Alexandersson J.* Dialogue Act Annotation with the ISO 24617-2 Standard. Multimodal interaction with W3C standards. Cham: Springer International Publishing. 2016. P. 109-135. https://doi.org/10.1007/978-3-319-42816-1_. 7. *Ang J., Liu Yang, Shriberg E.* Automatic dialog act segmentation

and classification in multiparty meetings // IEEE international conference on acoustics, speech, and signal processing (ICASSP '05). 2005. https://doi.org/10.1109/icassp.2005.1415300. 8. *Stolcke A., Ries K., Coccaro N., Shriberg E., Bates R., Jurafsky D., Meteer M.* Dialogue act modeling for automatic tagging and recognition of conversational speech // Computational linguistics. 2000. № 26(3). P. 339-373. https://doi.org/10.1162/089120100561737. 9. *Chen Z., Yang R., Zhao Z., Cai D., He X.* Dialogue Act Recognition via CRF-Attentive Structured Network. The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '18). 2018. P. 225-234. https://doi.org/10.1145/3209978.3209997. 10. *Berger J. O.* Statistical decision theory and Bayesian analysis. 2nd ed. New York: Springer-Verlag. 1985. 617 p. 11. *Grau S., Sanchis E., Castro J., Vilar D.* Dialogue act classification using a bayesian approach / 9th Conference speech and computer. 2004. P. 495-499. 12. *Mast M., Niemann H., Noth E., Schukat-Talamazzini E.G.* Automatic classification of dialog acts with Semantic Classification Trees and Polygrams. Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing. Springer, Berlin, Heidelberg. 1996. P. 217-229. https://doi.org/10.1007/3-540-60925-3_49. 13. *Wright H.* Automatic utterance type detection using suprasegmental features // ICSPL'98. 1998. Vol. 4. P. 1403. 14. *Bard E. G., Sotillo C., Anderson A. H., Thompson H. S., Taylor M. M.* The DCIEM Map Task Corpus: spontaneous dialogue under sleep deprivation and drug treatment. Speech Commun. 1996. Vol. 20. № 1-2. P. 71-84. https://doi.org/10.1016/S0167-6393(96)00045-3. 15. *Ries K.* HMM and neural network based speech act detection // IEEE international conference on acoustics, speech, and signal processing. Proceedings (ICASSP'99). 1999. Vol. 1. P. 497-500. https://doi.org/10.1109/icassp.1999.758171. 16. *Gang Ji, Bilmes J.* Dialog act tagging using graphical models. IEEE international conference on acoustics, speech, and signal processing (ICASSP'05). 2005. Vol. 1. P. 33-36. https://doi.org/10.1109/icassp.2005.1415043. 17. *Rotaru M.* Dialog act tagging using memory-based learning. 2002. P. 255-276. https://doi.org/10.1.1.116.7922. 18. *Samuel K., Carberry S., Vijay-Shanker K.* Dialogue act tagging with Transformation-Based Learning. Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL'98).1998. P. 1050-1056. https://doi.org/10.3115/980432.980757. 19. *Levin L., Langley C., Lavie A., Gates, D., Wallace D., Peterson K.* Domain Specific Speech Acts for Spoken Language Translation // Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue. 2003. P. 208-217. 20. *Andernach T., Poel M., Salomons E.* Finding classes of dialogue utterances with kohonen networks // ECML/MLnet workshop on empirical learning of natural language processing tasks. 1997. P. 85-94. 21. *Cho K., Merrienboer B., Bahdanau D., Bengio Y.* On the properties of neural machine translation: encoder-decoder approaches // Proceedings of SSST-8: Eighth workshop on syntax, semantics and structure in statistical translation. 2014. https://doi.org/10.3115/v1/w14-4012. 22. *Graves A., Wayne G., Reynolds, M. et al.* Hybrid computing using a neural network with dynamic external memory // Nature. 2016. Vol. 538. № 7626. P. 471-476. https://doi.org/10.1038/nature20101. 23. *Devlin J., Chang M., Lee K., Toutanova K.* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019. Vol. 1 P. 4171-4186. https://doi.org/10.18653/v1/N19-1423. 24. *Liu Y., Ott M., Goyal N., Du J.,Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L, Stoyanov V.* RoBERTa: A Robustly Optimized BERT Pretraining Approach. 2019. https://doi.org/10.48550/arXiv.1907.11692. 25. *Lee J. Y., Dernoncourt F.* Sequential short-text classification with recurrent and convolutional neural networks // Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: human language technologies. 2016. https://doi.org/10.18653/v1/n16-1062. 26. *Bothe C., Weber C., Magg S., Wermter S.* A Context-based Approach for Dialogue Act Recognition using Simple Recurrent Neural Networks. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC'2018). 2018. https://doi.org/10.48550/arXiv.1805.06280. 27. *Hochreiter S., Schmidhuber J.* Long Short-Term Memory. Neural computation. 1997. Vol. 9. № 8. P. 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Надійшла до редколегії 22.07.2022

Петров Костянтин Едуардович, доктор технічних наук, професор, завідувач кафедри інформаційних управляючих систем Харківського національного університету радіоелектроніки. Наукові інтереси: методи прийняття рішень, оптимізація організаційних систем. Адреса: пр. Науки, 14, м. Харків, 61166, Україна. E-mail: kostiantyn.petrov@nure.ua, тел.: +38(057)7021451.

Воробйов Євген Костянтинович, магістрант кафедри штучного інтелекту Харківського національного університету радіоелектроніки. Наукові інтереси: обробка природної мови, методи машинного навчання. Адреса: пр. Науки, 14, м. Харків, 61166, Україна. E-mail: yevhen.vorobiov@nure.ua, тел.: +38(095)8567606.

Кобзев Ігор Володимирович, кандидат технічних наук, доцент, доцент кафедри інформатики та комп'ютерної техніки Харківського національного економічного університету імені Семена Кузнеця. Наукові інтереси: методи управління організаційними системами. Адреса: пр. Науки, 9-А, м. Харків, 61166 Україна. E-mail: ikobzev12@gmail.com, тел.: +38(057)5038907.