УДК 621.3

DOI: 10.20535/2617-0965.eae.228077

Алгоритм розпізнавання природного мовного сигналу

Осадчук О. Р., ORCID 0000-0003-4934-2565
Кафедра акустичних та мультемедійних електронних систем ames.kpi.ua
Національний технічний університет України
«Київський політехнічний інститут ім. Ігоря Сікорського», ROR 00syn5v21
Київ, Україна

***Анотація*—В роботі наведено алгоритм розпізнавання і обробки Запитів користувача за допомогою нейронної мережі побудованої на принципі розуміння природної мови та обробки відеоряду для використання в системі підтримки користувачів.**

Ключові слова: намір користувача; алгоритм розпізнавання; обробка природної мови; нейронна мережа; розуміння природної мови; підтримка користувачів.

I. Вступ

На сьогоднішній день сервісні компанії та компанії з власними програмними продуктами мають потребу у короткі часові інтервали обробляти значну кількість запитів від користувачів технологічних продуктів. Одна людина може обробляти лише один запит за одну одиницю часу, в такій ситуації доцільно буде використовувати системи автоматичної обробки звернень користувачів. Враховуючи широку популярність мобільних додатків, пропонується інтеграції розмовного інтерфейсу у так програму. Окремо розглядається інтеграція в веб-інтерфейс, мобільний телефон, бота у соціальних мережах. У такій системі для обробки кожної розмови створюється багато навчальних фраз. Коли вираз для кінцевого користувача нагадує одну з цих фраз алгоритм повинен викликати попередньо визначену подію, яка відповідає наміру.

Метою роботи є формування алгоритму структурування, обробки великої кількості інформації за допомогою розбиття отриманої інформації на наміри користувача суб'єкти, контексти і події [1].

Аналіз намірів перевіряє введення даних користувача та визначає переважаючу суб'єктивну думку, особливо для визначення ставлення користувача до компанії як позитивної, негативної чи нейтральної. Подальше перетворення аудіо ряду з умовним сигналом у текст та його семантичний аналіз дозволяють отримати та автоматизувати обробку великих масивів інформації[2].

II. Поточний стан проблеми

Чи актуальні системи розпізнавання голосу в 2021 році? Технології розпізнавання мови все щільніше входять в наше життя, надаючи зручний засіб управління найрізноманітнішими електронними пристроями - управління голосом. Однією з актуальних проблем, яка вирішується при розробці таких систем керування, є проблема недостатньої точності розпізнавання голосових команд. Вдосконалення ведеться в напрямку підвищення надійності, незалежності від індивідуальних характеристик голосу, зниження негативного впливу фонового шуму на якість розпізнавання.

Для покращення точності розпізнавання голосу в пристрої стали використовувати глибинні нейронні мережі (ГНМ), які в останні роки неодноразово показували суттєві результати в процесах прогнозування, класифікації, розпізнавання образів, рукописного тексту та мовлення. Тому використання ГНМ та їх модифікації у задачах розпізнавання мови є актуальним завданням сьогодення.

III. Точне перетворення мови в текст за допомогою глибинної нейронної мережі

Модель розуміння природної мови, яка розуміє нюанси людської мови і інтерпретує текст або звук кінцевого користувача під час розмови до структурованих даних, які можна програмно зрозуміти і розпізнати складається з сервера-обробки нейронної мережі та інтерфейсу взаємодії з користувачем [3], схему представлено на рис. 1.

Алгоритм перетворення мови в текст може базуватися на одній із декількох моделей машинного навчання для транскрипції аудіофайлу[4]. Для нейронної мережі типу “голос в текст” можна визначити три основні методи для розпізнавання мови:

- Синхронне розпізнавання - надсилає аудіодані серверу-обробщику, виконує розпізнавання цих даних і повертає результати після обробки всього аудіо. Запити на синхронне розпізнавання обмежуються звуковими даними тривалістю менше 1 хвилини.
- Асинхронне розпізнавання надсилає звукові дані серверу-обробщику і ініціює тривалу

DOI: 10.20535/2617-0965.eae.228077

операцію розпізнавання. За допомогою цієї операції ви можете періодично отримувати результати розпізнавання[5].

- Розпізнавання потоку виконує розпізнавання аудіоданих, що надаються в двонаправленому потоці. Запити на трансляцію призначені для розпізнавання в режимі реального часу. Наприклад, для захоплення звуку в реальному часі з мікрофона. Розпізнавання потокового передавання забезпечує проміжні результати під час зйомки звуку, дозволяючи відображати результат, навіть, поки користувач все ще говорить[6]. Цей тип розвізнавання вибрано найбільш доцільним.

Функція перетворення мови в текст може використовувати наступні типи моделей машинного навчання для транскрипції аудіофайлів.

Відео - модель для транскрипції звуку у відеокліпах або включає кілька динаміків. Для найкращих результатів звук повинен бути записаний на частоті дискретизації 16000 Гц або більше.

Телефонний дзвінок - модель для транскрипції звуку з телефонного дзвінка. Зазвичай звук телефону записується з частотою дискретизації 8000 Гц.

Команди користувача в додатках або голосовий пошук - модель для транскрипції коротших відеокліпів. Деякі приклади включають голосові команди або голосовий пошук.

Вступний сигнал повинен бути трансформований і стиснутий для полегшення подальшої обробки. Є різні методи для вилучення корисних характеристик і стиснення вихідних даних в десятки разів без втрати корисної інформації. Найбільш використовувані методи: аналіз Фур'є [8]; лінійне передбачення мови, спектральний аналіз (Рис. 2).

Така Функція повинна містити значення зміщення часу (позначки часу) для початку та кінця кожного вимовленого слова, які розпізнаються у поданому звуці. Значення зміщення в часі являє собою кількість часу, що минув від початку звуку, з кроком у 100 мс [9]. В ідеальному варіанті потрібно забезпечити якомога чистіший звук, використовуючи якісний та добре розміщений мікрофон, якщо запис проводиться за допомогою гарнітури або мобільного пристрою можна вимикати функції шумозаглушення до аудіо перед надсиланням його до системи розпізнавання.

IV. Розпізнавання і обробка намірів користувача

Інтерпретація та обробка природної мови вимагає точного синтаксичного аналізатора. Якісний розмовний інтерфейс для кінцевих користувачів можна забезпечити за допомогою вводу у систему розпізнавання таких параметрів як наміри, контексти та субьекти.

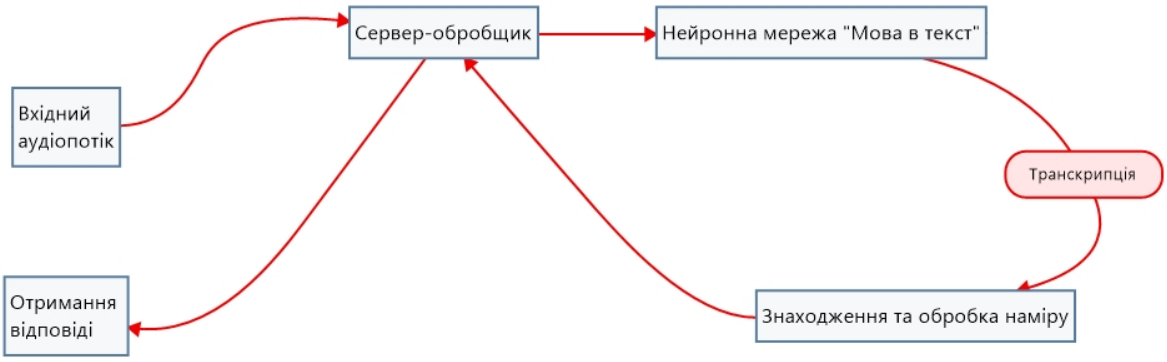
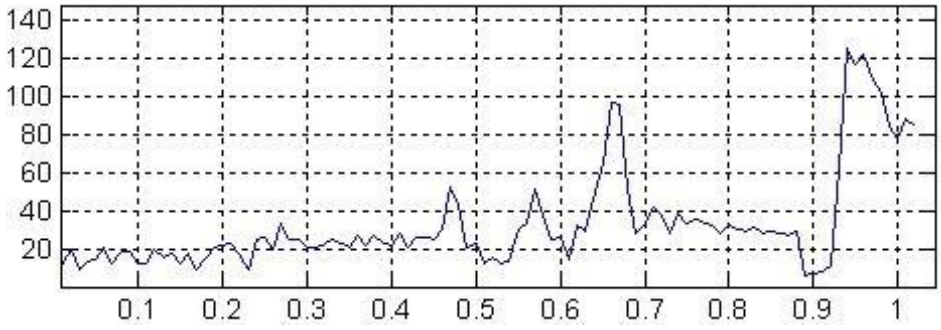


Рис.1 Модель розуміння природної мови



DOI: 10.20535/2617-0965.eae.228077

Намір класифікує намір кінцевого користувача протягом однієї розмови в черзі. Для кожної системи розпізнавання визначається багато намірів, комбіновані наміри можуть обробляти повну розмову. Коли кінцевий користувач пише і говорить щось алгоритм розпізнає тригер, який відповідає найкращому наміру у системі. Отримані значення з виразу кінцевого користувача можна визначити як параметри.

Кожен параметр має отримати тип, який називається типом наміру, який визначає, як саме витягуються дані. На відміну від необробленого введення кінцевим користувачем, параметри - це структуровані дані, які легко можна використовувати для виконання певної логіки або генерування відповідей. Для кожного параметру можна визначити дію. Коли намір розпізнається, його можна використати для активації певних дій, визначених у системі [10].

Для діалогів можна визначити контексти схожі на контекст природної мови. Якщо людина говорить "вони помаранчеві", потрібен контекст, щоб зрозуміти, про що говорять "вони". Подібним чином, щоб алгоритм обробляв вираз для кінцевого користувача, йому потрібно надати контекст, щоб правильно відповідати наміру.

Використовуючи контекст, можна контролювати хід розмови. Можна налаштувати контексти для наміру, встановивши вхідні та вихідні контексти, які визначаються іменами рядків. Коли трапляється збіг наміру, будь-який налаштований вихідний контекст для цього наміру стає активним [3]. Поки будь-який контекст активний, алгоритм повинен відповідати намірам, які налаштовані з вхідними контекстами, що відповідають поточно активним контекстом.

Для порівняння сигналів та отриманих параметрів в нейронну мережу потрібно внести навчальні фрази. Навчальні фрази - це приклади фраз для того, що можуть вводити або говорити кінцеві користувачі [12].

Для кожного наміру створюється багато навчальних фраз. Коли вираз для кінцевого користувача нагадує одну з цих фраз, алгоритм викликає дію, яка відповідає наміру. Наприклад, навчальна фраза "Я хочу піцу" тренує алгоритм розпізнавати вирази для кінцевих користувачів, подібні до цієї фрази, наприклад "Отримати піцу" або "Замовити піцу".

Потрібно створити щонайменше 50-60 (залежно від складності наміру) навчальних фраз, щоб ваша нейронна мережа могла розпізнавати різноманітні вирази для кінцевих користувачів.

Висновки

Традиційні комп’ютерні інтерфейси вимагають структурованого та передбачуваного введення інформації для нормальної роботи, що ускладнює використання цих інтерфейсів неприродньо, а часом і важко, тому інтерфейс може зробити висновок про те, чого хочуть кінцеві користувачі, виходячи з природної мови, якою вони користуються та співставити з заздалегідь розробленими відповідями [11]. Був сформований алгоритм структурування, обробки великої кількості інформації за допомогою розбиття отриманої інформації на наміри користувача суб'єкти, контексти і події.

Таку модель роботи можна забезпечити за допомогою розпізнавання і подальшого розбиття відної інформації на наміри, суб'єкти і контексти які запускають визначені для них подіях. Точність розпізнавання забезпечується за допомогою вводу навчальних фраз.

Надійшла до редакції 30 березня 2021 р.

Перелік посилань

[1]. Tekhnolohiya syntezu movlennya **URL**: https://www.understood.org/en/school-learning/assistive-technology/assistive-technologies-basics/text-to-speech-technology-what-it-is-and-how-it-works

[2]. Dialogflow **URL**: https://cloud.google.com/dialogflow/cx/docs

[3]. 8 Leading Language Models For NLP In 2020 **URL**: https://www.topbots.com/leading-nlp-language-models-2020/

[4]. J.-H. Lee, J.-H. Lee, S.-G. Sohn, J.-H. Ryu, and T.-M. Chung, “Effective Value of Decision Tree with KDD 99 Intrusion Detection Datasets for Intrusion Detection System,” in 2008 10th International Conference on Advanced Communication Technology, 2008, pp. 1170–1175, **DOI:** 10.1109/ICACT.2008.4493974.

[5]. “What is Natural Language Processing (NLP)? - Twilio.” **URL:** https://www.twilio.com/docs/glossary/what-natural-language-processing-nlp#what-about-semantic-analysis.

[6]. S. Peddabachigari, A. Abraham, C. Grosan, and J. Thomas, “Modeling intrusion detection system using hybrid intelligent systems,” J. Netw. Comput. Appl., vol. 30, no. 1, pp. 114–132, Jan. 2007, **DOI:** 10.1016/j.jnca.2005.06.003

[7]. “Speech recognition - Wikipedia.” **URL:** https://en.wikipedia.org/wiki/Speech_recognition

[8]. “Natural Language Understanding - IBM Cloud API Docs.” **URL:** https://cloud.ibm.com/apidocs/natural-language-understanding.

[9]. Benchmarking Natural Language Understanding Systems: Google, Facebook, Microsoft, Amazon **URL:** https://medium.com/snips-ai/benchmarking-natural-language-understanding-systems-google-facebook-microsoft-and-snips-2b8ddcf9fb19

[10]. L. N. Yasnitskiy, Vvedenie v Iskusstvennyiy Intellekt: Uchebnoe Posobie. Akademiya, 2010, **ISBN:** 978-5-7695-7042-1.

[11]. “Les concepts de base de Dialogflow pour écrire votre Chatbot | SoftFluent.” **URL:** https://www.softfluent.fr/blog/concepts-de-base-dialogflow-ecrire-chatbot/

[12]. Statcounter Global Stats 2020 **URL:** https://gs.statcounter.com/

[13]. Primeneniye golosovykh assistentov i chat-botov dlya biznesa **URL:** https://vc.ru/services/60540-primenenie-golosovyh-assistentovi-chat-botov-dlya-biznesa.

[14]. Marr, Bernard. How Artificial Intelligence IS Making Chatbots Better For Business. **URL:** https://www.forbes.com/sites/bernardmarr/2018/05/18/how-artificialintelligence-ismaking-chatbots-better-for-businesses/#69638bae4e72.

[15]. Informatsionnyye potoki: ponyatiye, vidy i sushchnost' **URL:** https://studwood.ru/1987457/informatika/informatsionnye_potoki_ponyatie_vidy_suschnost.

[16]. M. Mutiwokuziva, M. Chanda, P. Kadebu, A. Mukwazvure, T. Gotora “A neural-network based chat bot” 2017 2nd International Conference on Communication and Electronics Systems (ICCES), 19-20 Oct 2017. – P. 212217. **DOI:** 10.1109/CESYS.2017.8321268

UDC 621.3

Natural Speech Signal Recognition Algorithm

O. R. Osadchuk, ORCID 0000-0003-4934-2565
Department of Acoustic and Multimedia Electronic Systems ames.kpi.ua
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", ROR 00syn5v21
Kyiv, Ukraine

DOI: 10.20535/2617-0965.eae.228077

***Abstract*—Speech recognition technologies are becoming more and more part of our lives, providing a convenient way to control a variety of electronic devices - voice control. One of the current problems that is solved in the development of such control systems is the problem of insufficient accuracy of voice command recognition. Improvements are being made to increase reliability, independence from individual voice characteristics, and reduce the negative impact of background noise on recognition quality.**

The paper presents an algorithm for recognizing and processing user intentions using a neural network built on the principle of understanding natural language and processing audio signals for use in the user support system.

Keywords: user intent; recognition algorithm; natural language processing; neural network; understanding of natural language; user support.