

даних та стратегій. Велика кількість різномірних експертів забезпечує більше повне покриття можливих ринкових сценаріїв.

3. Агрегатор рішень - ключова частина запропонованої архітектури, пропонується що LLM буде отримувати на вхід результати роботи експертів, інтерпретувати результати їх роботи та приймати рішення чи потрібно виконувати торговельну операцію. LLM виступає у якості мета-експерта, який оцінює впевненість різних експертів в залежності від ситуації на ринку. Такий агрегатор може враховувати також додаткові фактори, як новини, аналізуючи їх, розпізнавати приховані залежності та ризики, які можуть не бути очевидними для числових моделей. Також важливим фактором є те, що агрегатор може створити пояснення для того чому він прийняв певне рішення, що підвищує прозорість роботи системи.
4. Модуль виконання операції - виконує операцію на біржі після прийняття рішення. Коли агрегатор приймає рішення про відкриття позиції він передає модулю виконання операції інформацію про тип позиції, стартову ціну та очікувану ціну закриття позиції. До цього модулю також входить модуль контролю ризику, за допомогою якого контролюється розмір позиції в залежності від впевненості агрегатора у позиції, стану рахунка та заданого рівня ризику. Таким чином розраховується співвідношення прибутку до ризику, на основі якого розраховується розмір позиції. Після чого модуль відправляє на біржу запит на відкриття позиції.

Запропонована архітектура торговельного бота поєднує сильні сторони спеціалізованих моделей-експертів та великих мовних моделей. Використання МоЕ забезпечує диверсифікацію аналізу за рахунок великої кількості експертів, кожна група з яких навчена на своєму типі даних для підзадачі, що підвищує стійкість системи до різних ринкових умов. Використання LLM як агрегатора дає можливість системі динамічно обирати найбільш релевантних експертів, обґрунтовувати прийняте рішення. Таким чином архітектура системи є більш адаптивною порівняно з традиційними ботами заснованими на алгоритмічних стратегіях. Очікується, що така система матиме кращі результати в дохідності та пояснюваності рішень. В майбутньому планується провести експериментальну оцінку запропонованого підходу на історичних даних ф'ючерного ринку та порівняти їх з результатами існуючих моделей. Така архітектура поєднує методи машинного навчання з елементами міркування природною мовою, що робить торгові рішення більш надійними та пояснювальними.

Список джерел:

1. Liu K. M., Lo M. C. LLM-Based Routing in Mixture of Experts: A Novel Framework for Trading //arXiv preprint arXiv:2501.09636. – 2025.
2. Chaudhary R. Advanced Stock Market Prediction Using Long Short-Term Memory Networks: A Comprehensive Deep Learning Framework //arXiv preprint arXiv:2505.05325. – 2025.
3. Bhuiyan M. S. M., Al Rafi M. D., Rodrigues G. N., Hossain Mir M. N., Ishraq A., Mridha M. F., Shin J. Deep learning for algorithmic trading: A systematic review of predictive models and optimization strategies // Array. – 2025. – № 100390. – ISSN 2590-0056. – DOI: 10.1016/j.array.2025.100390
4. Bai Y., Zhang X., Li Y., Wang J. A review of reinforcement learning in financial applications // Annual Review of Statistics and Its Application. – 2025. – Т. 12, № 1. – С. 209–232
5. Yu Z. et al. MIGA: Mixture-of-Experts with Group Aggregation for Stock Market Prediction //arXiv preprint arXiv:2410.02241. – 2024.

КОНТРОЛЬ ІНТЕНСИВНОСТІ ЕМОЦІЙ В УКРАЇНСЬКОМУ СИНТЕЗІ МОВЛЕННЯ ЧЕРЕЗ КРОС-МОВНЕ ПЕРЕНЕСЕННЯ ЗНАНЬ ТА АДАПТЕРИ LORA / EMOTION INTENSITY CONTROL IN UKRAINIAN TTS THROUGH CROSS-LINGUAL TRANSFER KNOWLEDGE AND LORA ADAPTERS

Іващенко Д.С. / Ivashchenko D.S.

Національний університет «Києво-Могилянська академія» /

National University of "Kyiv-Mohyla Academy" (NaUKMA)

04655, м. Київ, вулиця Григорія Сковороди, 2, НаУКМА, Факультет інформатики

d.ivashchenko@ukma.edu.ua

This paper presents a novel approach to controllable emotional text-to-speech synthesis for the Ukrainian language using cross-lingual emotion transfer and Low-Rank Adaptation (LoRA) adapters.

The proposed method leverages pre-trained multilingual TTS models and enables fine-grained control over emotional intensity through interpolation of latent representations. The paper investigates the effectiveness of LoRA adapters for emotion-specific fine-tuning, which provides a parameter-efficient way to inject emotional expressiveness while maintaining the naturalness of synthesized speech. The experimental setup includes adaptation of XTTS v2 model to Ukrainian using publicly available speech datasets, followed by emotion transfer from high-resource languages. This work addresses the critical gap in emotional speech synthesis in controllable emotional TTS for Ukrainian.

Емоційна виразність є критичною характеристикою природного мовлення, що робить синтез мовлення з контрольованими емоціями важливою задачею в галузі обробки мовлення. Для української мови, яка є малоресурсною в контексті технологій синтезу мовлення, розробка таких емоційних систем залишається актуальною. Сучасні системи синтезу українського мовлення [1] зосереджені переважно на нейтральному мовленні, не забезпечуючи контролю над емоційною складовою. Крім того, відсутність великих емоційно анотованих датасетів українського мовлення ускладнює тренування моделей, здатних синтезувати емоційно виразне мовлення.

Сучасні підходи до емоційного TTS базуються на глибоких нейронних мережах, зокрема на архітектурах типу VITS [2], YourTTS [3] та XTTS, які демонструють високу якість синтезу. Однак більшість цих моделей навчені на високоресурсних мовах (англійська, мандарин) і потребують адаптації для роботи з українською мовою. Метою даного дослідження є розробка методу контрольованого емоційного синтезу мовлення для української мови з використанням крос-мовного переносу знань та легких адаптерів LoRA.

Підхід заснований на трьох ключових компонентах: адаптації базової моделі до української мови, імплементації механізму контролю емоцій та використання LoRA адаптерів для ефективного донавчання моделей.

В якості базової моделі обрано XTTS v2 (Coqui TTS), яка є оновленою версією XTTS, багатомовною моделлю з підтримкою клонування голосу та перенесення стилю з аудіозразків. Модель базується на архітектурі GPT для автоматичного модуля обробки тексту, об'єднаний з генеративно-змагальним вокодером HiFi-GAN для синтезу аудіо з проміжних представлень. XTTS v2 навчена на 17 мовах, серед яких українська не представлена.

Адаптація здійснюється через донавчання моделі на українських мовленнєвих наборах даних, що містять текст, відповідний зразок мовлення та інформацію про мовця, таку як стать чи вік. Використовуються наступні відкриті ресурси: Ukrainian TTS dataset (голоси Lada, Mykyta) [4], набір даних M-AILABS та Mozilla Common Voice Ukrainian. Загальний обсяг тренувальних даних становить близько 30 годин високоякісного мовлення. Процес адаптації включає корегування текстового кодера для правильної обробки української фонетики та декодера для генерації природного мовлення.

Low-Rank Adaptation (LoRA) [5] використовується для ефективного навчання емоційно-специфічних адаптерів. Замість повного донавчання всієї моделі, LoRA додає низькорангові матриці до ваг шару уваги (attention layer). Для кожної емоції e тренується окремий LoRA адаптер A_e з рангом $r = 16$, що складає менше 1% від загальної кількості параметрів моделі. Це дозволяє ефективно змінювати емоційні представлення та комбінувати їх через зважену суму виходів шарів адаптерів.

Контроль емоційної інтенсивності реалізується через інтерполяцію латентних представлень. Для кожної цільової емоції e (радість, смуток, гнів) створюється референсне аудіо. Модель генерує умовні латентні вектори $z_{neutral}$ для нейтрального мовлення та z_e для емоційного. Контроль інтенсивності здійснюється через лінійну інтерполяцію:

$$z_{mixed} = \alpha \cdot z_e + (1 - \alpha) \cdot z_{neutral},$$

де $\alpha \in [0, 1]$ контролює інтенсивність емоції. Значення $\alpha = 0$ відповідає нейтральному мовленню, $\alpha = 1$ – максимальній емоційній інтенсивності.

Оскільки існує мало емоційно анотованих українських мовленнєвих датасетів, застосовується метод перенесення знань (transfer learning) з високоресурсних мов. Використовуються емоційні датасети для англійської мови (зокрема, ESD, RAVDESS) для попереднього навчання емоційних LoRA адаптерів. Завдяки багатомовності моделі XTTS,

латентні представлення емоцій є частково мовно-інваріантними, що забезпечує ефективне перенесення знань.

Оцінка якості синтезованого мовлення визначається за об'єктивними метриками (мел-кепстральне спотворення (MCD), частка помилково синтезованих слів (WER), класифікація емоційного мовлення моделями глибокого навчання) та суб'єктивними (усерелнена суб'єктивна якісна оцінка (MOS) з 5 оцінювачами, суб'єктивний тест на розпізнавання емоцій).

Попередні експерименти показують, що адаптована модель здатна генерувати природне українське мовлення з $MOS > 3.5$. Очікується, що механізм контролю інтенсивності забезпечить плавний перехід між нейтральним та емоційними станами з точністю розпізнавання емоцій $> 75\%$. LoRA адаптери демонструють ефективність у використанні пам'яті при збереженні якості емоційної виразності, порівнянної з повним донавчанням.

Запропонований підхід надає практичне рішення для розробки емоційно виразних систем синтезу мовлення для малоресурсних мов. Комбінація крос-мовного перенесення та LoRA шарів дозволяє ефективно адаптувати актуальні багатомовні моделі до української мови з контрольованою емоційною інтенсивністю. Подальші дослідження будуть зосереджені на розширенні набору емоцій, покращенні якості крос-мовного перенесення знань у моделях машинного навчання та удосконаленні підходу до моделювання емоційної інтенсивності.

Список використаних джерел

1. Ukrainian TTS using ESPnet framework // GitHub repository. 2021. URL: <https://github.com/robinhad/ukrainian-tts>.
2. Kim J., Kong J., Son J. Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech // Proceedings of the 38th International Conference on Machine Learning (ICML). 2021. P. 5530-5540.
3. Casanova E., Weber J., Shulby C.D. et al. YourTTS: Towards Zero-Shot Multi-Speaker TTS and Zero-Shot Voice Conversion for Everyone // Proceedings of the 39th International Conference on Machine Learning (ICML). 2022. P. 2709-2720.
4. Casanova E. et al. XTTS: a Massively Multilingual Zero-Shot Text-to-Speech Model // arXiv preprint arXiv:2406.04904. 2024.
5. Hu E.J., Shen Y., Wallis P. et al. LoRA: Low-Rank Adaptation of Large Language Models // International Conference on Learning Representations (ICLR). 2022.

ВДОСКОНАЛЕННЯ ВІЗУАЛІЗАЦІЇ МОДЕЛЕЙ АРХІТЕКТУР ПРОГРАМ ДЛЯ ЗМЕНШЕННЯ КОГНІТИВНОГО НАВАНТАЖЕННЯ КОРИСТУВАЧА / IMPROVING VISUALIZATION OF SOFTWARE ARCHITECTURE MODELS TO REDUCE USER COGNITIVE LOAD

Франків О.О. / Frankiv O.O.

Національний університет «Києво-Могилянська академія» / National University of Kyiv-Mohyla Academy

04070, м. Київ, вул. Г. Сковороди 2, каф. інформатики, тел. (044) 425 60 64

E-mail: o.frankiv@ukma.edu.ua, тел. (044) 425 60 64

The main priority in developing automated visualization tools is to provide a clear and easily interpretable experience for the user. When visualizing complex and abstract entities such as software architecture, reducing the cognitive effort required to perceive information is essential, allowing the user to focus on the actual domain. In this paper, we examine practical