

А.Л. Литвинов¹, П.Д. Андрейчева¹, В.М. Бредіхін¹, В.І. Вербицька²

¹Харківський національний університет міського господарства імені О.М. Бекетова, Україна

²Харківський національний автомобільно-дорожній університет, Україна

ТЕНДЕНЦІЇ РОЗВИТКУ МОДЕЛЕЙ ГЕНЕРАЦІЇ УКРАЇНСЬКОЇ МОВИ

Стаття досліджує розвиток технологій генерації мови від моделей машинного навчання до рідких нейронних мереж для генерації текстів. Розглянуто проблеми навчання та шляхи їх вирішення як для моделей машинного навчання, так і нейронних мереж. Як результат, визначено етапи розробки нейронної мережі для генерації текстів українською мовою.

Ключові слова: нейронна мережа, українська мова, машинне навчання, синтез мови, озвучування текстів.

Постановка проблеми

Мовні технології Text-to-Speech (TTS) використовують абсолютно в різних галузях: для озвучення контенту; для підключення автовідповідача; для автоматизації роботи операторів; для озвучення субтитрів; для роботи віртуальних помічників та для допомоги автомобілістам.

Ключове призначення сервісу генерації (синтезу) української мови на основі NNS-технологій – надавати користувачу можливість перекладу повідомлення, записаного символами українського алфавіту, у відповідну звукову хвилю, що імітує мову живої людини, без втрати змістовної частини цього повідомлення [1].

Англійська мова є однією з найпоширеніших у світі, вона є офіційною мовою більш ніж в 60 державах. Відповідно, інвестиції, що вкладаються в розвиток технологій для розпізнавання цієї мови, досить швидко приносять дохід. Тим часом, українською мовою спілкуються близько 50 млн людей, і вона є другою слов'янською мовою за кількістю користувачів. На розвиток мовних технологій, крім економічних проблем, передусім впливають особливості української мови та складності, що виникають у процесі її обробки. Як основні можна виділити: відсутність строгих граматичних конструкцій побудови речень, а також численні правила словотворення, представлення слів і розстановки наголосів, причому практично до всіх правил існують винятки [2].

Проблеми, що виникають зі створенням систем автоматичного синтезу мови і систем автоматичного розпізнавання мови, пов'язані з надзвичайною мінливістю основних характеристик мовного сигналу. Відсутність закінченої моделі мовного сигналу не дозволяє грамотно сформулювати вимоги до первинних пристроїв-синтезаторів та аналізаторів. У результаті всі рішення, що є на сьогодні, супроводжуються великою кількістю помилок. Особливо це стосується систем автоматичного розпізнавання мови [3].

Аналіз останніх досліджень і публікацій

Різним теоретичним та практичним аспектам генерації мови, і української зокрема, присвячені наукові та практичні роботи як вітчизняних, так і закордонних дослідників. Завдяки появі великих обчислювальних можливостей у середині 2010-х років широкого поширення набув End-to-End метод генерування мови (рис. 1).

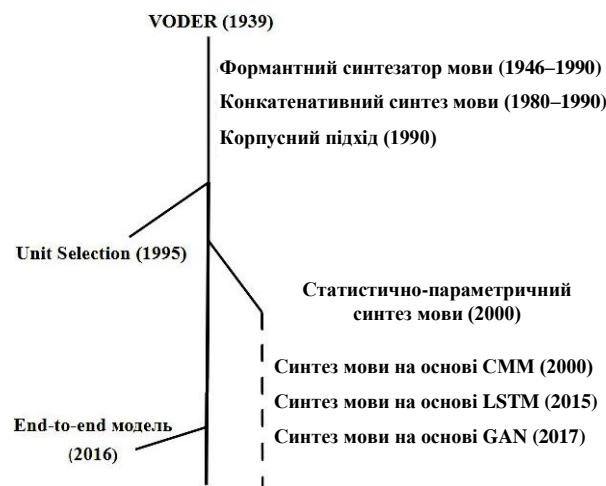


Рис. 1. Генеалогічне дерево розвитку технологій синтезу мови [4]

Цей метод передбачає використання однієї нейронної мережі для генерування сигналу мови з лінгвістичних параметрів, з'єднуючи у такий спосіб акустичну модель з вокодером в одну нейронну мережу. Незважаючи на те, що метод здатний видавати високоякісний мовний сигнал, він погано застосовується для роботи в реальному режимі, оскільки потребує дуже великих обчислювальних ресурсів.

Якщо звернути увагу на огляд сучасних моделей на момент середини 2021 року, то можна помітити, що за останні 4,5 роки системи на базі архітектури штуч-

них нейронних мереж почали демонструвати метрику MOS на англійських наборах даних максимально близьку до значень, що отримуються на голосі живої

людини. Однак за весь цей час так і не перевершили її. Крім того, якість кращих моделей синтезу мови практично перестала відрізнятися одна від одної (рис. 2).

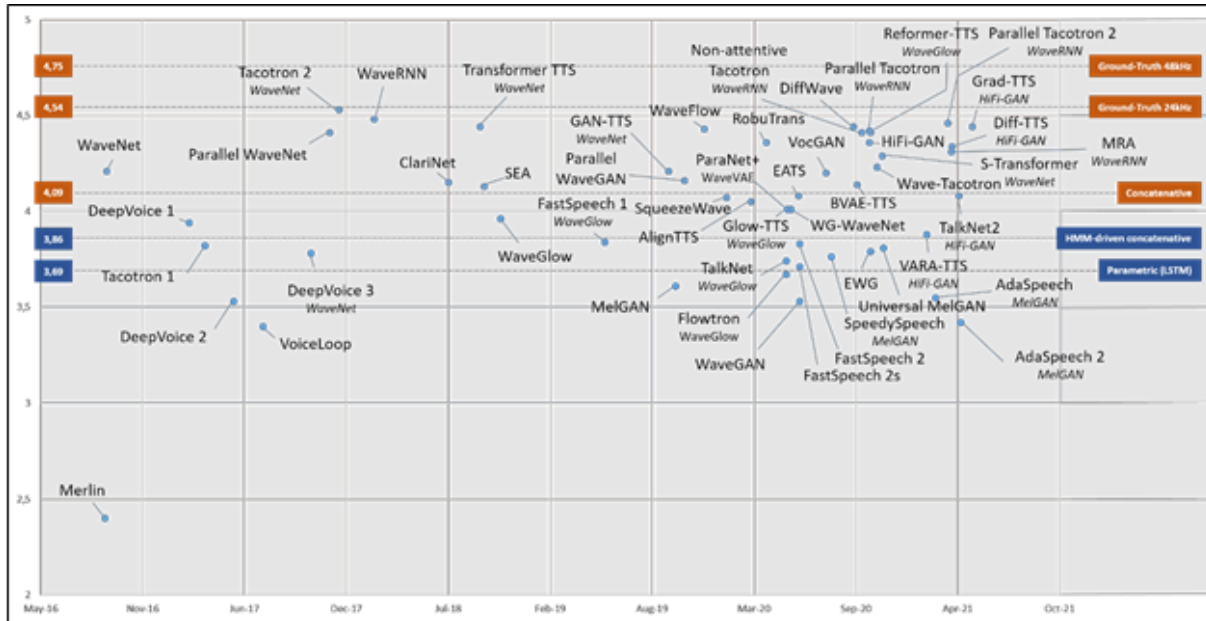


Рис. 2. Порівняльний аналіз нейромережових архітектур для синтезу мови [5]

Системи синтезу мовлення можна класифікувати за способами створення мовного сигналу і вирізняють три основні напрями [6]:

- артикуляторний синтез – створення штучних мовленнєвих сигналів на основі моделювання мовного апарату людини;
- формантний синтез – створення штучних мовленнєвих сигналів на основі акустичних моделей і динамічної зміни таких параметрів, як значення частот основного тону формант та зашумленості;
- конкатенативний синтез – створення вихідного акустичного сигналу на основі конкатенації (по-слідовного додавання) необхідних елементів синтезу.

Мета статті

Мета статті полягає у виборі архітектури нейронної мережі для побудовання застосунку генерації української мови і визначення етапів її створення.

Виклад основного матеріалу

Розробка нейронних мереж для генерації природної мови, зокрема української, є важливим напрямком у сучасній комп'ютерній науці, особливо під час війни для збереження автентичності мови та ведення інформаційної війни. Для виконання цього завдання використовують рекурентні або трансформерні архітектури нейронних мереж, як-от LSTM, GRU або BERT [7].

Визначимо основні етапи розробки нейронної мережі для генерації української мови:

1. Збір та підготовка даних.

На цьому етапі потрібно зібрати корпус текстів українською мовою, провести попередню обробку даних, серед яких токенізація, видалення непотрібних символів та інші операції очистки.

Для цього треба ідентифікувати джерела текстових даних українською мовою, які охоплюють:

- українські новинні портали та медіа-ресурси;
- літературні твори, книги та тексти з української літератури;
- україномовні веб-сайти, блоги, форуми та соціальні мережі;
- публічні документи, наукові роботи та інші види текстів.

Тексти з вибраних джерел треба зберігати у вигляді текстового корпусу.

Після проведення попередньої обробки текстового корпусу його слід розділити на тренувальний та тестовий набори, після чого виявити відмінності у стилі та темах, забезпечивши різноманітність текстів у корпусі, враховуючи різні стилі, теми та жанри. Це допоможе моделі навчатися різноманітності української мови.

Попередньо оброблені дані потрібно зберегти у вигляді файлу або бази даних для зручності подальшого використання.

Результатом виконання цього етапу буде підготовлений корпус текстів, який можна використовувати для тренування моделі генерації тексту українською мовою.

2. Вибір архітектури нейронної мережі.

Далі слід визначитись з типом архітектури нейронної мережі (рекурентна, трансформерна тощо).

Існує кілька різних трансформерних архітектур, які можна використовувати для генерації природної мови:

- GPT (Generative Pre-trained Transformer). Це серія моделей GPT, розроблена компанією OpenAI. Вона використовує трансформерну архітектуру та тренується на великому обсязі текстових даних. Як приклад конкретної моделі – це GPT-3 або 4;

- BERT (Bidirectional Encoder Representations from Transformers). Це модель від Google, яка використовує двонаправлене кодування контексту для кращого розуміння мови. Ця архітектура спеціалізується на розумінні контексту та використовується в різних завданнях NLP. Як приклад конкретної моделі – це BERT-Base або BERT-Large;

- XLNet. Ця модель поєднує ідеї з GPT та трансформерів, додаючи механізми вивчення контексту та взаємодії між словами в оберненому порядку. Як приклад конкретної моделі – це XLNet-Base або XLNet-Large [8];

- T5 (Text-to-Text Transfer Transformer). Ця модель розглядає всі завдання тексту як задачі перетворення тексту та універсально застосовує трансформер для різних мовних завдань. Як приклад конкретної моделі – це T5-Base або T5-Large [9];

- RoBERTa (Robustly optimized BERT approach). Вона є оптимізованою версією BERT з деякими покращеннями, серед яких виключення маскування слів та тренування на довших послідовностях. Як приклад конкретної моделі – це RoBERTa-Base або RoBERTa-Large [10].

Ці архітектури мають свої особливості та застосовуються в різних варіантах завдань генерації тексту. Вибір конкретної архітектури залежить від конкретних вимог та доступності моделей для української мови.

3. Навчання моделі.

На цьому етапі модель буде тренуватися на підготовленому раніше корпусі українських даних, розділеному для тренування та тестування, з використанням втратного функціоналу, який включений внутрішньо в модель та оптимізатор AdamW.

Важливо налаштувати параметри тренування (наприклад, кількість епох, розмір пакета тощо), які залежать від обсягу корпусу даних та обчислювальних ресурсів.

4. Адаптація до української мови.

На цьому етапі слід забезпечити, щоб модель розуміла особливості української мови: граматичні конструкції, словниковий запас та інші мовні особливості. Використати техніки, які враховують специфіку української мови під час навчання.

5. Оптимізація та тюнінг.

На цьому етапі треба провести оптимізацію гіперпараметрів для поліпшення якості генерації текстів та врахувати різноманітні методи тюнінгу, наприклад, Dropout, Batch Normalization та інші, для зменшення перенавчання.

б. Оцінка якості.

Як і при навчанні інших моделей машинного навчання, тут використовують валідаційний та тестовий набори для оцінки якості генерації, застосовують метрики, як-от BLEU, ROUGE та інші, для оцінки подібності та якості результатів. Розглянемо їх більш розширено [11]:

- метрики автоматичної оцінки:

- 1) BLEU (Bilingual Evaluation Understudy) – визначає схожість між модельним та відповідним референсним текстом. Використовується для автоматичної оцінки машинного перекладу, але також може бути застосована до генерації текстів;

- 2) ROUGE (Recall-Oriented Understudy for Gisting Evaluation) – застосовується для вимірювання схожості між генерованим текстом та референсними текстами;

- метрики для оцінки різноманітності перевіряють, наскільки різноманітні та різні за змістом є генеровані тексти. Можна використовувати власні метрики для вимірювання різноманітності, зокрема унікальні слова чи різноманітність фраз;

- метрики людської оцінки перевіряють якість роботи моделі за допомогою людських експертів або користувачів. Для цього експертам пропонують оцінити генерований текст за шкалою від 1 до 5, де 5 – це висока якість, а 1 – низька якість;

- метрики візуальної перевірки використовують, щоб переконатись, що згенеровані тексти логічні, граматично правильні та відповідають контексту, чи модель враховує специфічні мовні особливості української мови;

- метрика врахування взаємодії з користувачем використовується в конкретному застосуванні, для оцінки, наскільки добре модель взаємодіє з користувачем та враховує його потреби.

Важливо зауважити, що оцінка якості моделі мовної генерації є відкритим завданням, і немає єдиного стандарту для її проведення. Комбінування різних метрик та підходів може надати більш повний образ про якість генерації текстів.

Але як і для інших моделей машинного навчання, при генерації текстів українською мовою теж є проблеми, які необхідно вирішувати.

По-перше, це недостатність тренувальних даних. Для вирішення цієї проблеми треба збільшити обсяг тренувальних даних за допомогою аугментації або використання схожих мовних корпусів.

Проблема перенавчання. Для її вирішення треба використовувати регуляризаційні техніки, як-от Dropout та Weight Decay, для зменшення

ефекту перенавчання.

При низькій різноманітності генерації треба додати різноманіття в генеровані тексти шляхом використання температури при виборі слів під час генерації.

Проблема поганої обробки мовної інформації вирішується завдяки застосуванню попередньо навчених векторних представлень слів (Word Embeddings), для поліпшення обробки мовної інформації.

Проблема врахування контексту вирішується завдяки використанню архітектур нейронних мереж, які здатні враховувати ширший контекст, наприклад, трансформерні архітектури.

Незважаючи на те, що нейронні мережі є потужним інструментом штучного інтелекту, вони

мають певні обмеження:

- вони вимагають значної кількості помічених навчальних даних;
- вони обробляють дані непослідовно, що робить їх неефективними під час обробки даних у часі.

Тому з'явилась нова архітектура для генерації текстів – рідкі нейронні мережі (LNN) – тип нейронної мережі, яка навчається під час своєї роботи, а не лише на етапі навчання [12]. Вона є безперервною у часі рекурентною нейронною мережею (RNN), яка послідовно обробляє дані, зберігає пам'ять про минулі вхідні дані, регулює свою поведінку на основі нових вхідних даних і може обробляти вхідні дані змінної довжини для розширення можливостей розуміння завдань (рис. 3).

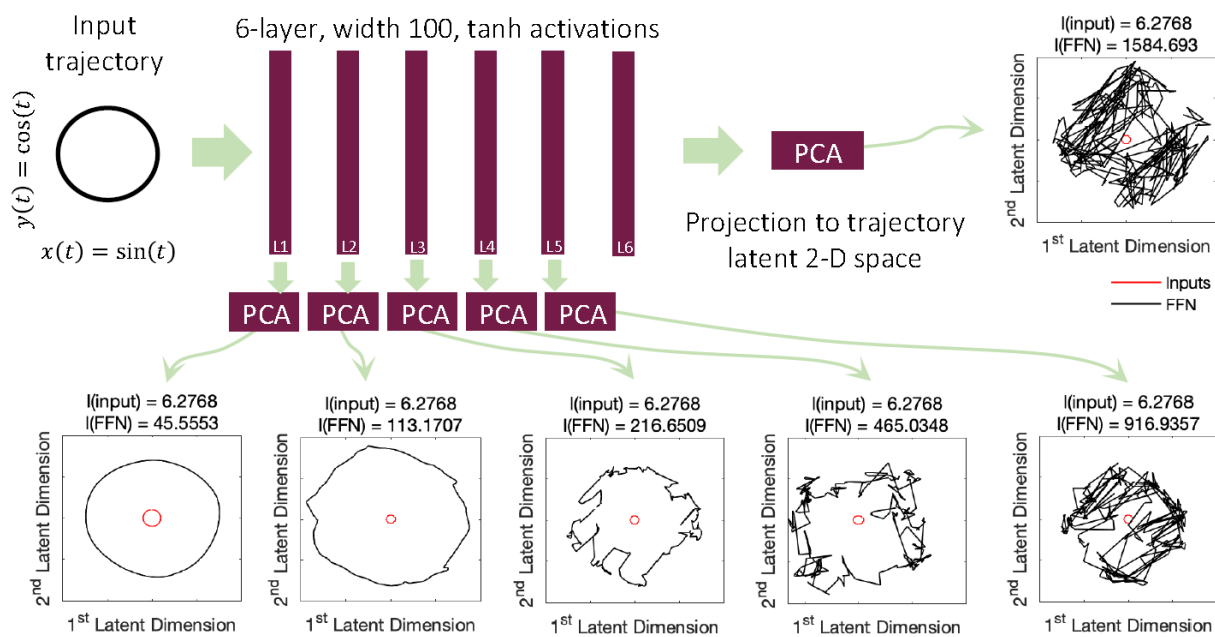


Рис. 3. Для LNN-мереж прихований простір траєкторії стає складнішим, оскільки вхідні дані проходять через приховані шари [13]

Архітектура LNN відрізняється від традиційних нейронних мереж своєю здатністю ефективно обробляти безперервні дані або дані часових рядів. Якщо доступні нові дані, LNN можуть змінити кількість нейронів та з'єднань на шарі.

Мережа виражає стан системи у часі, як традиційний підхід у нейронних мережах (NN), але LNN має дві ключові особливості:

- динамічна архітектура інакша, де нейрони виразніші, ніж нейрони звичайної нейронної мережі, що робить LNN більш інтерпретованими. Вони можуть ефективно обробляти послідовні дані у реальному часі;

- безперервне навчання та адаптивність LNN-мереж дає їм змогу адаптуватись до змінних даних навіть після навчання, більш точно імітувати мозок живих організмів, ніж традиційні NN, які перестають вивчати нову інформацію після етапу навчання моделі.

Отже, LNN не вимагають великих обсягів помічених навчальних даних для отримання точних результатів.

Оскільки нейрони LLM пропонують багаті зв'язки, які можуть передавати більше інформації, вони менші за розміром, ніж звичайні NN, тому дослідникам стає легше пояснити, як LNN ухвалила рішення. Крім того, менший розмір моделі та менша кількість обчислень можуть зробити їх масштабованими на рівні використання. До того ж ці мережі більш стійкі до шуму та перешкод у вхідному сигналі, як порівняти з NN.

Завдяки своїй адаптивності, можливостям навчання в реальному часі та динамічній топології LNN дуже добре розуміють довгі текстові послідовності природною мовою.

Здатність LNN вчитися на даних у реальному часі допомагає їм аналізувати діалект, що розвива-

ється, і нові фрази, що дозволяє проводити більш точний аналіз тональності. Подібні можливості можуть бути корисними і при машинному перекладі.

Проте вони також мають деякі обмеження та проблеми:

– проблема зниклого градієнта. У глибоких нейронних мережах проблема зниклого градієнта виникає, коли градієнти для оновлення ваг нейронних мереж стають надзвичайно малими. Ця проблема не дозволяє нейронним мережам досягати оптимальних ваг. Це може обмежити їхню здатність ефективно вивчати довгострокові залежності;

– проблема налаштування параметрів. LNN мають декілька параметрів, включно з вибором ЗДР (звичайні диференціальні рівняння) для вирішувача, параметрами регуляризації та мережевою архітектурою, які необхідно налаштувати для досягнення найкращої продуктивності.

Пошук відповідних параметрів часто потребує ітеративного процесу, на який потрібен час. Якщо налаштування параметрів неефективне або виконане неправильно, це може призвести до неоптимальної реакції мережі та зниження продуктивності. Однак дослідники намагаються вирішити цю проблему, з'ясовуючи, наскільки менше нейронів потрібно для виконання того чи іншого завдання.

Висновки

Як бачимо, нейронні мережі для генерації текстів українською мовою перетворилися з MLP (багатошаровий перцептрон) на LNN (рідкі нейронні мережі), що є більш динамічними, адаптивними, ефективними та надійними, ніж традиційні нейронні мережі, та мають безліч потенційних варіантів використання.

Література

1. DOU Моделі text-to-speech для синтезу української мови: як працюють, хто їх створює та де знайти [Electronic resource] – Regime of access: <https://dou.ua/forums/topic/40194/>, free (date of the application: 10.02.2024).
2. D. Suendermann, H. Høge, and A. Black, “Challenges in speech synthesis” *Speech Technology*, Boston: Springer – pp. 19–32, – 2010. DOI:10.1007/978-0-387-73819-2_2
3. A. Iida, and N. Campbell, “Speech database design for a concatenative text-to-speech synthesis system for individuals with communication disorders,” *International Journal of Speech Technology* – vol. 6, – no. 4, – pp. 379–392 – 2023; DOI:10.1023/A:1025761017833.
4. A. Van Den Oord et al. “WaveNet: A generative model for raw audio,” *SSW* – 2016 – vol. 125 DOI:10.48550/arXiv.1609.03499
5. State Of The Art of Speech Synthesis at the End of May 2021 [Electronic resource] – Regime of access <https://towardsdatascience.com/state-of-the-art-of-speech-synthesis-at-the-end-of-may-2021-6ace4fd512f2>, free (date of the application: 13.02.2024)
6. М.М. Шатковський Моделі комп'ютерного подання українського мовлення для проблеми конкатенативного

сегментивного синтезу Інженерія програмного забезпечення № 3 (7) – 2011. С. 58-67.

7. ResearchGate H.i Murfi, T. Gowandi, G. Ardaneswari BERT-Based Combination of Convolutional and Recurrent Neural Network for Indonesian Sentiment Analysis 2022, DOI:10.48550/arXiv.2211.05273.

8. XLNet Embeddings (Large Cased) [Electronic resource] – Regime of access <https://sparknlp.org/2021/07/07/xlnet-large-cased-en.html>, free (date of the application: 15.02.2024)

9. T5-large [Electronic resource] – Regime of access <https://huggingface.co/google-t5/t5-large>, free (date of the application: 17.02.2024)

10. P. Banerjee, T. Gokhale, C. Baral Self-Supervised Test-Time Learning for Reading Comprehension – 2021, DOI:10.18653/v1/2021.naacl-main.95

11. Hakan Tekgul, *ML Solutions Engineer Published March 16, 2023 Monitoring Text-Based Generative AI Models Using Metrics Like Bleu Score* [Electronic resource] – Regime of access <https://arize.com/blog-course/generative-ai-metrics-bleu-score/>, free (date of the application: 18.02.2024)

12. O. Okozelsk Reservoir Computing for .NET (RCNet) [Electronic resource] – Regime of access: <https://github.com/okozelsk/NET?tab=readme-ov-file#reservoir-computing-for-net-rcnet>, free (date of the application: 18.02.2024).

13. R. Hasani, M.s Lechner, A. Amini, D. Rus, R. Grosu Liquid Time-constant Networks – 2020 – arXiv:2006.04439v4

References

1. DOU Моделі text-to-speech для синтезу української мови: як працюють, хто їх створює та де знайти [Electronic resource] – Regime of access: <https://dou.ua/forums/topic/40194/>, free (date of the application: 10.02.2024).
2. D. Suendermann, H. Høge, and A. Black, “Challenges in speech synthesis” *Speech Technology*, Boston: Springer – pp. 19–32, – 2010. DOI: 10.1007/978-0-387-73819-2_2
3. A. Iida, and N. Campbell, “Speech database design for a concatenative text-to-speech synthesis system for individuals with communication disorders,” *International Journal of Speech Technology* – vol. 6, – no. 4, – pp. 379–392 – 2023; DOI:10.1023/A:1025761017833.
4. A. Van Den Oord et al. “WaveNet: A generative model for raw audio,” *SSW* – 2016 – vol. 125 DOI:10.48550/arXiv.1609.03499
5. State Of The Art of Speech Synthesis at the End of May 2021 [Electronic resource] – Regime of access <https://towardsdatascience.com/state-of-the-art-of-speech-synthesis-at-the-end-of-may-2021-6ace4fd512f2>, free (date of the application: 13.02.2024)
6. М.М. Shatkovskiy Modeli kompiuternoho podання ukrainskoho movlennia dla problemy konkatentativnoho sehmentyvnoho syntezu Inzheneriia prohrannoho zabezpechennia № 3 (7) – 2011. С. 58-67.
7. ResearchGate H.i Murfi, T. Gowandi, G. Ardaneswari BERT-Based Combination of Convolutional and Recurrent Neural Network for Indonesian Sentiment Analysis 2022, DOI:10.48550/arXiv.2211.05273.
8. XLNet Embeddings (Large Cased) [Electronic resource] – Regime of access <https://sparknlp.org/2021/07/07/xlnet-large-cased-en.html>, free (date of the application: 15.02.2024)
9. T5-large [Electronic resource] – Regime of access <https://huggingface.co/google-t5/t5-large>, free (date of the application: 17.02.2024)
10. P. Banerjee, T. Gokhale, C. Baral Self-Supervised Test-Time Learning for Reading Comprehension – 2021,

DOI:10.18653/v1/2021.naacl-main.95

11. Hakan Tekgul, ML Solutions Engineer Published March 16, 2023 Monitoring Text-Based Generative AI Models Using Metrics Like Bleu Score [Electronic resource] – Regime of access <https://arize.com/blog-course/generative-ai-metrics-bleu-score/>, free (date of the application: 18.02.2024)

12. O. Okozelsk Reservoir Computing for .NET (RCNet) [Electronic resource] – Regime of access: <https://github.com/okozelsk/NET?tab=readme-ov-file#reservoir-computing-for-net-rcnet>, free (date of the application: 18.02.2024).

13. R. Hasani, M.s Lechner, A. Amini, D. Rus, R. Grosu Liquid Time-constant Networks – 2020 – arXiv:2006.04439v4

Рецензент: д-р фіз.-мат. наук, проф. Н.Д. Сізова, Харківський національний університет міського господарства імені О.М. Бекетова, Україна.

Автор: ЛИТВИНОВ Анатолій Леонідович доктор технічних наук, професор, професор кафедри комп'ютерних наук та інформаційних технологій Харківський національний університет міського господарства імені О.М. Бекетова

E-mail – anatoliy.litvinov@kname.edu.ua

ID ORCID: <https://orcid.org/0000-0003-2440-6358>

Автор: АНДРЕЙЧЕВА Поліна Дмитрівна здобувач вищої освіти 4-го курсу бакалавратури навчально-наукового інституту енергетичної, інформаційної та транспортної інфраструктури Харківський національний університет міського господарства імені О.М. Бекетова

E-mail – polina.andreicheva@kname.edu.ua

Автор: БРЕДІХІН Володимир Михайлович кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук та інформаційних технологій Харківський національний університет міського господарства імені О.М. Бекетова

E-mail – bredixinv@gmail.com

ID ORCID: <https://orcid.org/0000-0002-6063-5046>

Автор: ВЕРБИЦЬКА Вікторія Іванівна кандидат економічних наук, доцент, доцент кафедри обліку і оподаткування

Харківський національний автомобільно-дорожній університет

E-mail – verbytska67@gmail.com

ID ORCID: <https://orcid.org/0000-0001-7103-6738>

DEVELOPMENT TENDENCIES OF GENERATION MODELS OF THE UKRAINIAN LANGUAGE

A. Lytvynov¹, P. Andreicheva¹, V. Bredikhin¹, V. Verbytska²

¹O.M. Beketov National University of Urban Economy in Kharkiv, Ukraine

²Kharkiv National Automobile and Highway University, Ukraine

The article explores the development of language generation technologies, from machine learning models to fluid neural networks for text generation. English is one of the most widespread languages in the world: it is the official language of more than 60 countries. The events of recent years have led to the development of the popularity of the Ukrainian language not only in the country but also abroad. The article analyses scientific sources on this topic, the results of which formed a base for creating a genealogical tree of the development of language synthesis technology. The study pays particular attention to the problem of automating the generation of texts in the Ukrainian language, which includes the absence of strict grammatical constructions of sentences, numerous rules of word formation, presentation of words, placement of accents, and exceptions to almost all rules. The article presents several different transformative architectures that one can use to generate the Ukrainian language. We considered the problems of training such networks and ways to solve them for machine learning models and neural networks. As a result, we determined the stages of development of a neural network for the generation of texts in the Ukrainian language. Even though neural networks are a powerful tool of artificial intelligence, they have certain limitations that the new architecture – liquid neural networks (LNN) – is devoid of. The article also discusses their advantages and disadvantages. The LNN architecture differs from traditional neural networks in its ability to efficiently process continuous or time series data. If new data is available, LNNs can change the number of neurons and connections in a layer. Due to these two main features of LNN, the network expresses the system's state over time as opposed to the traditional neural network (NN) approach, so it does not require large amounts of observed training data to obtain accurate results. In conclusion, we should note that, despite their shortcomings, LNN networks are more dynamic, adaptive, efficient, and reliable than traditional neural networks and have many potential uses.

Keywords: neural network, Ukrainian language, machine learning, speech synthesis, text-to-speech.