

МАШИННЕ НАВЧАННЯ У СПЕЦІАЛІЗОВАНИХ ГОЛОСОВИХ СИСТЕМАХ

Єфімова В.В.

e-mail: valerii.iefimova@nure.ua

Науковий керівник — д-р техн. наук, проф. Машталір С.В.
Харківський національний університет радіоелектроніки, каф. ІНФ,
м. Харків, Україна

Machine learning has become a cornerstone in the development of modern voice assistants, enabling them to move beyond simple command execution toward complex, context-aware interactions. While mainstream systems such as Alexa, Siri, and Google Assistant rely on large-scale datasets and general-purpose models, specialized voice systems require tailored approaches. Supervised learning allows classification of user queries, reinforcement learning optimizes responses through feedback, and unsupervised learning uncovers hidden patterns in organizational workflows. These methods, combined with explainable AI and low-data learning techniques, make specialized assistants adaptable to diverse communities.

За останні роки голосові помічники стали невід'ємною частиною цифрової екосистеми: вони інтегровані в операційні системи, мобільні застосунки, браузерери та розумні пристрої. Їх використання охоплює широкий спектр завдань — від пошуку інформації та організації робочих процесів до автоматизації побуту й підтримки здоров'я.

Голосові системи — це складні програмно-апаратні комплекси, які поєднують технології обробки природної мови (NLP), розпізнавання мовлення (ASR) та генерації мовлення (TTS). Їхня ефективність значною мірою залежить від машинного навчання, яке дозволяє системам адаптуватися до нових даних, покращувати точність та забезпечувати персоналізовану взаємодію.

Існують такі методи машинного навчання у голосових системах, як:

Навчання з учителем (Supervised Learning). Використовується для класифікації запитів та розпізнавання мовлення. Модель навчається на парі «вхідні дані — правильна відповідь» [1]. Формально задача класифікації описується як мінімізація функції втрат:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N l(f_{\theta}(x_i), y_i),$$

де x_i — вхідні дані, y_i — правильна мітка, f_{θ} — модель з параметрами θ , l — функція втрат (наприклад, крос-ентропія).

Навчання без учителя (Unsupervised Learning). Використовується для виявлення прихованих закономірностей у мовних даних, наприклад кластеризації акцентів чи стилів мовлення [1].

Алгоритми: k-means, PCA, автоенкодері. Мета — мінімізувати відстань між точками та центрами кластерів:

$$J = \sum_{i=1}^N \|x_i - \mu_{c_i}\|^2.$$

Навчання з підкріпленням (Reinforcement Learning). Застосовується для оптимізації діалогових стратегій. Система отримує винагорода за правильні відповіді та штраф за помилки. Основне рівняння:

$$Q(s, a) = r + \gamma \max_{a'} Q(s', a'),$$

де s — стан, a — дія, r — винагорода, γ — коефіцієнт дисконтування.

Архітектура нейронних мереж є ключовим чинником у створенні голосових систем, адже саме вона визначає, як модель обробляє мовні дані та будує контекст. Історично першими застосовувалися рекурентні нейронні мережі (RNN), які добре працюють із послідовними даними, такими як мовлення чи текст. Проте класичні RNN мають проблему «згасання градієнтів», що ускладнює роботу з довгими залежностями. Для її вирішення були запропоновані LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit), які завдяки спеціальним «воротам» здатні утримувати інформацію протягом тривалих інтервалів [3].

Сучасним стандартом стали трансформерні моделі (Transformer), які замінили рекурентність механізмом уваги (attention). Основна ідея полягає у тому, що кожне слово чи елемент послідовності може «звертати увагу» на інші елементи, незалежно від їхньої позиції. Формула механізму уваги виглядає так:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

де Q — матриця запитів, K — матриця ключів, V — матриця значень, а d_k — розмірність простору ключів.

Цей підхід дозволяє моделі враховувати глобальний контекст, що критично важливо для розпізнавання мовлення та генерації природних відповідей. На основі трансформерів створено моделі BERT, GPT, Whisper, які сьогодні є базою для більшості голосових систем.

Навчання нейронних мереж у голосових системах відбувається шляхом мінімізації функції втрат. Найпоширеніший метод — градієнтний спуск:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} L(\theta_t),$$

де θ — параметри моделі, η — швидкість навчання, L — функція втрат.

Для задач класифікації мовних сигналів часто використовується крос-ентропія:

$$L = - \sum_{i=1}^N y_i \log(p_i),$$

де y_i — фактична мітка (0 або 1), p_i — передбачувана ймовірність для цієї мітки [2].

Щоб уникнути перенавчання, застосовуються методи регуляризації: Dropout (випадкове вимикання нейронів під час навчання), L2-регуляризація (штраф за великі ваги). Для прискорення навчання використовуються оптимізатори, такі як Adam, які комбінують адаптивну швидкість навчання та моментум.

У випадку діалогових систем важливим є також застосування навчання з підкріпленням (Reinforcement Learning), де модель оптимізує стратегію відповіді на основі винагороди [4]. Це дозволяє системі поступово вдосконалювати діалог, враховуючи реакцію користувачів.

У багатьох сферах неможливо зібрати великі корпуси мовних даних. Для цього застосовуються методи few-shot та zero-shot learning, які дозволяють моделі виконувати нові завдання з мінімальною кількістю прикладів або навіть без них, використовуючи знання з попереднього навчання [5].

Архітектури нейронних мереж, методи оптимізації та сучасні підходи до навчання моделей формують основу спеціалізованих голосових систем. Використання трансформерів, градієнтного спуску, регуляризації та навчання з підкріпленням дозволяє створювати системи, здатні до високоточного розпізнавання та адаптації.

Список використаних джерел:

1. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. Cambridge, MA: MIT Press, 2016. 775 p.
2. Markus Kasanmascheff. Why Entropy is the Secret Ingredient for Smarter AI. 2024. URL: <https://winbuzz.com/2024/12/15/why-entropy-is-the-secret-ingredient-for-smarter-ai-xcxwbn/>
3. Hochreiter, S., Schmidhuber, J. Long Short-Term Memory // Neural Computation. 1997. Vol. 9(8). P. 1735–1780.
4. Sutton, R. S., Barto, A. G. Reinforcement Learning: An Introduction. 2nd ed. Cambridge, MA: MIT Press, 2018. 552 p.
5. Dave Breegman. What is zero-shot learning? 2023. URL: <https://www.ibm.com/think/topics/zero-shot-learning>