РОЗПОЗНАННЯ МОВИ: ЕТАПИ РОЗВИТКУ СУЧАСНОЇ ТЕХНОЛОГІЇ ТА ПЕРСПЕКТИВИ

Кирилов Д.І., студент МК51-20
Науковий керівник – ***Плєхова Г.А.,*** доц., к.т.н.
Харківський національний автомобільно-дорожній університет

Розпізнавання і синтез мовлення – це інноваційні технології, що дозволяють комп'ютерам розуміти та відтворювати людське мовлення. За останні кілька десятиліть відбувся значний прорив у цій сфері завдяки поєднанню передових алгоритмів та збільшення обчислювальних потужностей.

Перші спроби створення систем розпізнавання мовлення датуються 1950-1970-ми роками, коли використовувалися словники та правила для розпізнавання окремих слів. Однак точність таких систем залишалася низькою, досягнувши лише 70-80%. Справжній прорив стався у 1990-х роках з появою технологій на основі прихованих марковських моделей (HMM). Вони значно підвищили точність розпізнавання мовлення, що дозволило почати використовувати цю технологію в кол-центрах та інших сферах.

У 2000-х роках розвиток глибоких нейронних мереж дав новий поштовх в галузі розпізнавання та синтезу мовлення. Однією з ключових інновацій стали рекурентні нейронні мережі, особливо LSTM, які ідеально підходять для роботи з послідовними даними, такими як мовлення.

Сучасні системи, такі як DeepSpeech та BERT, засновані на глибоких нейронних мережах і демонструють високі показники точності: від 95-98% для коротких фраз і до 85-90% для довгих висловлювань.

Крім розпізнавання, нейронні мережі дозволили досягти природного звучання в системах синтезу мовлення. Тут лідерами є WaveNet від DeepMind, а також моделі, розроблені компаніями Baidu та NVIDIA.

Останнім часом розвиваються енд-ту-енд системи, які можуть одночасно розпізнавати та синтезувати мовлення, не використовуючи окремі компоненти. Це спрощує розробку додатків для роботи з мовленням.

Загалом технології розпізнавання та синтезу мовлення за останні роки досягли значного прогресу. Проте все ще залишаються певні проблеми:

- Складність розпізнавання мовлення в умовах шуму.
- Важкість розпізнавання мовлення з сильним акцентом чи діалектом.
- Неприродне або монотонне звучання синтезованого мовлення.
- Високі вимоги до обчислювальних потужностей для роботи нейронмереж.

Подальший розвиток цих технологій пов'язаний зі створенням більш ефективних моделей, здатних працювати в режимі реального часу на мобільних пристроях. Також велика увага приділяється тому, щоб нейронні мережі могли краще розуміти контекст і підтекст мовлення, що наблизить нас до створення справді інтелектуальних діалогових систем.

Перспективні напрямки розвитку

1. **Мультимодальні системи**: Поєднують акустичні дані з візуальною інформацією, що дозволяє поліпшити точність розпізнавання мовлення в умовах шуму. Наприклад, зчитування мови з губ допомагає розпізнавати фрази навіть при сильному фоні.

2. **Генеративні суперницькі мережі (GAN)**: Моделі, що дозволяють генерувати високоякісне синтезоване мовлення, важко відрізнити від справжнього людського. GAN допомагають досягти природного звучання голосу за мінімальних зусиль.

3. **Самонавчання і підкріплене навчання**: Ці методи дозволяють моделям самостійно поліпшувати свої можливості, вивчаючи мовлення користувачів і адаптуючись до їхніх особливостей.

4. **Розподілене навчання**: Дозволяє тренувати моделі безпосередньо на пристроях користувачів, забезпечуючи конфіденційність даних і зберігаючи приватні мовні записи на самому пристрої.

Український контекст

Незважаючи на прогрес, більшість рішень у сфері розпізнавання та синтезу мовлення орієнтовані на англійську мову, тоді як для української мови таких рішень обмежено. Створення високоякісних україномовних моделей вимагає великих обсягів транскрибованих даних і врахування особливостей української вимови й граматики.

Соціальна значущість технологій

Крім комерційних можливостей, технології розпізнавання та синтезу мовлення мають значний потенціал у соціальній сфері. Вони можуть поліпшити доступ до інформації для людей з вадами слуху та мовлення, забезпечуючи автоматичне створення субтитрів та переклад жестової мови.

Надійність і безпека

Забезпечення надійності й безпеки технологій є ключовими для їх подальшого розвитку. Системи обробки мовлення мають уникати помилок і збоїв, особливо в критично важливих сферах. Важливо також гарантувати конфіденційність даних, використовуючи сучасні методи машинного навчання та кібербезпеки.

Таким чином, незважаючи на досягнутий прогрес, технології розпізнавання та синтезу мовлення мають ще великі перспективи для вдосконалення, особливо для української мови. Їх подальший розвиток матиме вагомий вплив як на технологічний сектор, так і на суспільство загалом.