

Микола Биков, В'ячеслав Ковтун, Євген Пустовіт (Вінниця)

УДОСКОНАЛЕННЯ ПРОЦЕСУ НАВЧАННЯ АВТОМАТИЗОВАНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ МОВЦЯ КРИТИЧНОГО ЗАСТОСУВАННЯ ІЗ НММ-DNN СКЛАДОВОЮ

Розширення сфери застосування інформаційних технологій призводить до зростання актуальності забезпечення надійної та зручної біометричної ідентифікації особи. У ряді критичних систем, наприклад, транспортному засобі, який рухається, виникає потреба у забезпеченні можливості розпізнавання особи, але використання традиційних способів ідентифікації за індивідуальними характеристиками ока, пальця чи руки виявляється неможливим без відволікання оператора. В таких обставинах процедуру розпізнавання особи пропонується робити за індивідуальними ознаками голосу використовуючи автоматизовану систему розпізнавання мовця критичного застосування (АСРМКЗ) [1], основною вимогою до якої є забезпечення прогнозованості результатів розпізнавання із урахуванням умов оточуючого мовця акустичного середовища. Виконання цієї умови ускладнюється стохастичним характером мовного сигналу, на який впливає множина керованих, некерованих і невизначених обставин. Отже, логічним виявляється пошук у мовному сигналі відносно усталених елементів, наприклад, фонем або похідних від них конструкцій, індивідуальні ознаки яких можна використати для розпізнавання мовця, описавши останні засобами математичного апарату прихованих Марковських мереж (НММ). Втім, практичні дослідження виявили, що класичний спосіб моделювання тривалості станів у НММ неадекватно відтворює щільність імовірності тривалості перебування у стані (щільність тривалості стану) для фізичних та, зокрема, мовних сигналів, що змушує явно моделювати тривалість стану обмежуючи діапазон застосувань системи розпізнавання. Компенсувати цей недолік пропонується доповненням НММ глибокою нейромережею (DNN), що при безпосередній реалізації призводить до суттєвого ускладнення процесу навчання системи розпізнавання. Для спрощення процесу навчання АСРМКЗ із DNN-НММ складовою пропонується адаптувати крос-ентропійний критерій та критерій максимуму взаємної інформації для факторного аналізу інформативних DNN-ознак та створити метод інтеграції оптимізованих ознак у НММ-компоненту АСРМКЗ. Такий підхід до опису процесу навчання НММ-DNN-моделі мовотворення, на відміну від існуючих, забезпечить оцінювання різниці між апостеріорним розподілом імовірності всіх станів НММ та фактичним апостеріорним розподілом імовірності, обчисленим DNN, із урахуванням ступеня збурень у вхідному мовному сигналі, та забезпечить використання у процесі розпізнавання особи мовця семантичної інформації, яка описується ознаками, що спостерігаються на послідовності фреймів, на які розбито вхідну фонограму, що дозволить підвищити ефективність текстозалежного розпізнавання особи мовця при експлуатації АСРМКЗ у шумному акустичному середовищі.

Практична наслідком отриманих математичних результатів є формулювання заходів для структурної адаптації та опису практичних аспектів застосування НММ-DNN моделей у задачі розпізнавання мовця, зокрема, при обґрунтуванні вибору виду та методу нормалізації векторів інформативних ознак на рівні фреймів, при визначенні кількості станів НММ та параметрів GMM в залежності від параметрів обраної моделі мовотворення, а також при інтерпретації результатів розпізнавання. Проведені дослідження довели ефективність удосконаленого крос-ентропійного критерію навчання НММ-DNN, у якому застосовано концепцію формування множини кодів стану (МКС), яка задає тенденцію щодо ступеня інформативності окремих множин інформативних ознак. Емпірично доведено, що оптимальна кількість елементів у МКС для АСРМКЗ знаходиться у діапазоні 10-15% від кількості нейронів на прихованих шарах DNN. Також виявилось, що удосконалений критерій максимуму взаємної інформації показує співставні із удосконаленим крос-ентропійним критерієм результати розпізнавання мовців лише при тривалості паролльної фрази понад 2 хв., що робить процедуру розпізнавання некомфортною.

Література

1. Mykola M. Bykov, Viacheslav V. Kovtun and others, «Research of neural network classifier in speaker recognition module for automated system of critical use» Proc. SPIE Vol. 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017; 1044521 (2017); doi: 10.1117/12.2280930.