

ДОСЛІДЖЕННЯ МЕТОДІВ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

Вінницький національний технічний університет

Анотація

Обґрунтовано актуальність задачі розпізнавання емоцій, проведено огляд основних методів розпізнавання емоцій та вказано основні переваги і недоліки кожного з них, здійснено порівняння та вибір нейронної мережі для розв'язання даної задачі.

Ключові слова: розпізнавання емоцій, емоція, обличчя людини, нейронна мережа, глибока нейронна мережа, згортоква нейронна мережа.

Abstract

The urgency of the problem of recognition of emotions is proved, the review of the basic methods of recognition of emotions, the main advantages and disadvantages of each method, the comparison and selection of the neural network for the solution of this problem.

Keywords: emotion recognition, emotion, human face, neural network, deep neural network, convolutional neural network.

З тих пір, як були розроблені комп'ютери, вчені та інженери думали про штучні інтелектуальні системи, які психічно та/або фізично рівноцінні людям. Протягом останніх десятиліть збільшення загальнодоступної обчислювальної потужності допомагало розвивати машини швидкого навчання, тоді як Інтернет надав величезну кількість даних для навчання. Ці дві розробки сприяли дослідженню інтелектуальних систем самонавчання, нейронні мережі – один з найбільш перспективних напрямів. Розпізнавання обличчя може використовуватися для спостереження з боку правоохоронців, а також у натовпі. Інші сучасні програми передбачають автоматичне розмивання обличчя на кадрах Google Streetview та автоматичне розпізнавання друзів Facebook за фотографіями. Крім того, такі методи використовуються в цифрових камерах для автоматичного фотографування, коли користувач посміхається. Однак найбільш перспективні розробки стосуються гуманізації штучних інтелектуальних систем. Якщо комп'ютери здатні відслідковувати психічний стан користувача, роботи можуть реагувати на це і вести себе належним чином. Отже, безперечно, розпізнавання емоцій відіграє ключову роль у поліпшенні взаємодії людини з машиною.

Метою роботи є порівняння існуючих методів розпізнавання емоцій людини та вибір найдоцільнішого з них, а також аналіз нейронних мереж, та обґрунтування вибору нейронної мережі, яка буде відповідати найважливішим, для системи розпізнавання, критеріям.

Зростання доступної обчислювальної потужності на користувацьких комп'ютерах на початку двадцять першого століття сприяв розвитку алгоритмів, що використовуються для інтерпретації зображень. В області класифікації зображень можна виділити два основних підходи. З одного боку, заздалегідь запрограмовані методи виділення ознак можуть використовуватися для аналітичного розбиття рисунку, щоб класифікувати зображений об'єкт. Аналогом такого підходу є нейронні мережі, що навчаються та само навчаються. Розроблена таким чином програмна система розробляє правила класифікації об'єктів шляхом виділення подібних ознак при навчанні на тестовій вибірці даних. Зважаючи на сьогоденну наявність навчальних даних та обчислювальну потужність, можна очікувати, що продуктивність моделей, що базуються на нейронних мережах, може бути суттєво покращена. Нижче перераховані деякі останні розробки, виконані в вище окресленому напрямку:

П.1. Проривна публікація про автоматичну класифікацію зображень дана Крижевським та Гінтоном [1]. Ця робота показує глибоку нейронну мережу, яка нагадує функціональність зорової кори людини. Використовуючи самостійно розроблену розмічену колекцію з 60000 зображень у 10 класах, що називається набором даних CIFAR-10, була розроблена модель для категоризації об'єктів із зображень. Ще одним важливим результатом дослідження є ві-

зуалізація ознак для категоризації у нейронній мережі, щоб можна було оцінити, як модель розбиває картинки.

П.2. В іншій роботі, яка використовує набір даних CIFAR-10 [2], розроблена дуже широка і глибока мережева архітектура, поєднана з підтримкою GPU, щоб зменшити час навчання. На популярних наборах даних, таких як рукописні цифри MNIST, китайські символи та зображення CIFAR-10, досягається майже людська продуктивність. Вкрай низькі показники помилок значно перевищили попередні результати досліджень. Однак слід зазначити, що мережа, що використовується для набору даних CIFAR10, складається з 4-х згорткових шарів і має складну структуру. Як результат, навіть із використанням GPU, час навчання становить кілька днів.

П.3. У роботі Крижевського та ін., робота над покращенням архітектури [3] була продовжена, і було розроблено мережу з 5 згортковими шарами, навченої за допомогою 1,2 мільйонів зображень високої роздільної здатності з конкурсу ImageNet LSVRC-2010. Мережа показала аналогічний [1] результат при спрощеній архітектурі. Крім того, до сьогодні проводяться експерименти зі зменшенням розміру мережі, констатуючи, що кількість шарів можна значно зменшити, а продуктивність лише трохи знизиться.

П.4. Що стосується розпізнавання виразів обличчя, то це специфічна мережа глибоких переконань. Найпомітнішою особливістю мережі є концепція ієрархічного розбору обличчя [4], тобто зображення передається через мережу кілька разів, щоб вперше виявити обличчя, після цього очі, ніс і рот, і, нарешті, належну емоцію.

П.5. Інша робота в напрямку розпізнавання емоцій [5] використовує фільтрацію Gabor для обробки зображень та підтримку векторної машини (SVM) для класифікації. Фільтр Габора особливо підходить для розпізнавання образів у зображеннях і, як стверджується, імітує функцію зорової системи людини. Точність розпізнавання емоцій висока, коливаючись від 88% на гнів до 100% на здивованість. Великим недоліком підходу є те, що потрібна дуже точна попередня обробка даних, така що кожне зображення відповідає строгому формату, перш ніж подавати його в класифікатор.

П. 6. Одне з останніх досліджень з розпізнавання емоцій описує нейронну мережу, здатну розпізнавати расу, вік, стать та емоції за зображеннями облич [6]. Набір даних, що використовується для останньої категорії, походить з виклику розпізнавання виразів обличчя (FERC-2013). Чітко організована глибока мережа, що складається з трьох згорткових шарів, одного повністю з'єднаного шару та декількох невеликих шарів між ними дозволила досягти середню точність близько 67% за класифікацією емоцій, що дорівнює попереднім сучасним публікаціям на тому ж наборі даних. Крім того, ця теза закладає цінний аналіз ефекту коригування розміру мережі, об'єднання та випадання.

Як вказано у ряді інших джерел [7,8], найбільш перспективною концепцією аналізу виразів обличчя є використання глибоких згорткових нейронних мереж. Однак мережа з пункту П.2 вищенаведеного списку вважається занадто важкою для нашого обмеженого обсягу доступних ресурсів обробки. Оригінальна мережа з П.3 також велика, але більш дрібні версії вважаються достатньо придатними для роботи. Крім того, через надзвичайно високі вимоги до формату вхідних зображень ми не будемо оцінювати П.4 та П.5, так як існує необхідність роботи при різній якості вхідних даних, особливо якщо це стосується розпізнавання емоцій у режимі реального часу. Отже у дослідженні буде розглянуто три глибокі архітектури для вирішення проблеми класифікації емоцій. Ці архітектури походять із мереж, наведених у пунктах П.1, П.3 та П.6.

Проведений аналіз показав, що мережа П.6 показує дещо повільнішу криву навчання, але остаточна точність набору валідацій аналогічна вимогам мережі П.3. Вимоги до обробки знаходяться між вимогами інших мереж, тому виходячи з цього факту, застосування мережі П.1 здається найбільш перспективним підходом до нашого завдання розпізнавання емоцій. Однак продуктивність мережі П.6 на додатковому тестовому наборі фотографій облич RaFD суттєво краща (60%), ніж у мережі П.1 (50%). Це свідчить про кращі узагальнюючі можливості, що дуже важливо для розробки на її основі майбутнього програмного додатку.

Отже, остання мережа з П.6 виявилася найбільш перспективною для практичних застосувань.

Висновки

Під час дослідження було проведено аналіз основних підходів щодо розпізнавання емоцій людини. Доведено доцільність використання глибокої згорткової нейронної мережі для вирішення задачі розпізнавання людських емоцій, проаналізовано основні варіанти архітектур нейронних мереж. Проведено дослідження роботи нейронних мереж на тестовій вибірці. В результаті дослідження для найбільш ефективного розпізнавання людських емоцій було вирішено використати глибоку нейронну мережу, що складається з 3-х згорткових шарів, одного повністю з'єднаного шару та декількох невеликих шарів між ними.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
2. D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 3642–3649. IEEE, 2012
3. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
4. Y. Lv, Z. Feng, and C. Xu. Facial expression recognition via deep learning. In Smart Computing (SMARTCOMP), 2014 International Conference on, pages 303–308. IEEE, 2014.
5. T. Ahsan, T. Jabid, and U.-P. Chong. Facial expression recognition using local transitional pattern on gabor filtered facial images. IETE Technical Review, 30(1):47–52, 2013.
6. A. Gudi. Recognizing semantic features in faces using deep learning. arXiv preprint arXiv:1512.00743, 2015.
7. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 248–255. IEEE, 2009.
8. Яровий А. А. Розпізнавання мімічних мікровиразів обличчя людини на основі Time Delay Neural Network / Яровий А. А., Кашубін С. Г., Кулик О. О., Липкань І. М. // Вісник Хмельницького національного університету. Технічні науки. – 2015. – № 1. – С. 122 – 126.
9. Інтелектуальна система нейромережевого розпізнавання мімічних мікровиразів обличчя людини / Кашубін С., Яровий А.: Збірник праць ІХ Міжнародної науково-практичної конференції [Інтернет — Освіта — Наука (ІОН-2014)], (Вінниця, 14 –17 жовтня 2014 р.) – Вінниця, ВНТУ, 2014. – с. 60 –62.

Зінов'єв Євгеній Вікторович – студент групи ІКН-19М, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: zinovieff55@gmail.com

Арсенюк Ігор Ростиславович – к. т. н., доцент, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Yevgeny Zinoviev V. – Student of Department of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: zinovieff55@gmail.com

Ihor Arseniuk R. – Cand. Sc. (Eng), Assistant Professor of the Chair of Computer Science, Vinnytsia National Technical University, Vinnytsia.