

Р.В. Кириленко, Н.А. Гук, О.В. Верба

Дніпровський національний університет імені Олеся Гончара

ПЕРЕДБАЧЕННЯ ПСИХОЛОГІЧНОЇ РЕАКЦІЇ КОРИСТУВАЧА ПРИ ПЕРЕГЛЯДІ ЦИФРОВИХ ЗОБРАЖЕНЬ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ

В статті розглядається комплекс алгоритмів для розв'язання задачі прогнозування психологічної реакції користувача на зображення при перегляді цифрових зображень. Для виділення облич на зображеннях використовуються діаграми орієнтованих градієнтів. Прогнозування психологічної реакції виконується з використанням нейронної мережі.

Ключові слова: штучна нейронна мережа, метод зворотного розповсюдження помилки, антропологічні точки, гістограма орієнтованих градієнтів, згорткова нейронна мережа, прогнозування.

В статье рассматривается комплекс алгоритмов для решения задачи прогнозирования психологической реакции пользователя на изображения при просмотре цифровых изображений. Для выделения лиц на изображениях используются диаграммы ориентированных градиентов. Прогнозирование психологической реакции выполняется с использованием нейронной сети.

Ключевые слова: искусственная нейронная сеть, метод обратного распространения ошибки, антропологические точки, гистограмма ориентированных градиентов, сверточная нейронная сеть, прогнозирование.

In the article, a problem of prediction a user's psychological response to the presented image is considered. A complex algorithm that solves proposed problem with respect to fuzzy input data which operates in real time is proposed.

A classification problem for each particular user based on previously generated content by the user using a set of algorithms including machine learning algorithm is considered. Due to a huge amount of variety in possible input data and it's complexity algorithms that reduce fuzziness are considered. In order for a prediction system to understand and learn from data it has been provided, it has to be prepared in such a way that the algorithm could more easily find patterns and inferences. For that purpose, all incoming data passes two additional steps which also allows the system to neglect fuzziness.

At the first step, the software has to definitely determine a presence of the desired object (face, in the article) which must be done fast enough to be made in a time of presenting the picture to a user and it must be stable to data fuzziness. In order to solve that problem histogram of oriented graphs is considered. Because of its nature, the algorithm solves a problem in comparatively fast and robust way. At the second step, the face landmark estimation algorithm is considered. It allows the system to reduce the number of patterns that have to be learned in order to predict the behavior reducing the number of such patterns by narrowing down the two-dimensional transform of the object.

The last step, prediction, is implemented with an artificial neural network that embraces incoming object and determines user's resolution based on its experience. Such algorithm

allows the system to keep learning throughout its life cycle which leads to constant improvements in its results. The proposed algorithms complex was implemented. High speed, fuzziness resistance, independence of external conditions can be considered as its advantages. Slow learning cycle can be considered as its main disadvantage. Tests' results provided in the article.

Key words: artificial neural network, method of reverse error propagation, anthropological points, histogram of oriented gradients, convolutional neural network, forecasting.

Вступ. Автоматизація процесів у наш час здійснюється у багатьох сферах діяльності людини. Потреба в автоматизації виникла ще в глибоку давнину і постійно зростала в процесі розвитку людських потреб і засобів виробництва. Задача, що розглядається у даній роботі, тісно пов'язана із автоматизацією процесів мозку людини, а саме процесу розпізнавання обличчя на зображенні та прогнозування психологічної реакції людини при перегляді цього зображення. Прогнозування відбувається на основі паттернів поведінки людини та існуючої навчальної вибірки, отриманої при попередньому перегляді зображень. Результати прогнозування можливе використовувати при відборі облич для рекламних зображень, при формуванні контекстних вибірок зображень на сайтах знайомств.

Аналіз публікацій за темою дослідження. В останні роки спостерігається розвиток підходів щодо розпізнавання зображень та оцінювання привабливості облич, заснований на системах машинного зору та класифікаторах, для яких розробляються алгоритми навчання. Із використанням запропонованих методів та алгоритмів можливо класифікувати обличчя. У роботі [4] у якості ознак, за якими визначається привабливість, пропонується використовувати пропорції рис обличчя по ключовим точкам. Недоліком запропонованого підходу є необхідність визначати необхідні точки на обличчі власноруч. В [6] для виділення характерних ознак, за якими класифікується зображення, окрім пропорцій було використано також метод головних компонент. В [8] для розв'язання задачі використовувався нейромережевий підхід.

Постановка задачі. Задача полягає у розробці комплексного підходу, що дозволяє передбачити психологічну реакцію користувача на зображення, що демонструється, базуючись на попередньому досвіді системи – емоційних реакціях людини при перегляді зображень, що демонструвались йому раніше. В якості вхідних даних виступає фотографічне зображення обличчя людини в момент перегляду зображення. На виході програма повинна надати оцінку емоційної реакції користувача на переглянуте зображення.

Модель задачі прогнозування. Сформульовану задачу будемо розглядати як задачу класифікації зображень для кожного окремо взятого користувача. Оскільки рішення про емоційну оцінку людини, що приймає система, повинно ґрунтуватись на попередньому досвіді, будемо використовувати моделі та методи машинного навчання, як альтернативу створенню математичної моделі. Слід зазначити, що математичні алгоритми машинного навчання працюють з великими обсягами даних і здатні знаходити приховані

закономірності поведінки об'єкту спостереження. Головним при використанні таких моделей є наявність інформації про об'єкт. Для розв'язання сформульованої задачі будемо використовувати вибірку, яка складається з ситуаційних прикладів та відомих реакцій людини, виявлених під час перегляду цих прикладів. Розроблений алгоритм повинен узагальнити досвід попередніх спостережень за реакціями людини для обробки нових випадків, які не зустрічалися раніше, та результат яких заздалегідь не є відомим. Така система повинна вміти налагоджувати свій алгоритм прийняття рішень автоматично, отримуючи навчальні приклади при опитуванні користувача в реальному режимі часу.

В рамках галузі аналізу даних машинне навчання є методом, який використовується для ідентифікації складних моделей. Методи машинного навчання дозволяють побудувати процес навчання на існуючих прикладах і виконувати передбачуваний аналіз на них.

Метод розв'язання задачі. Для розв'язання сформульованої задачі прогнозування емоційної реакції людини на переглянуте зображення необхідно зробити попередню обробку фотографії людини під час перегляду зображення, відобразити обличчя у вигляді масиву числових даних, розробити систему підтримки прийняття рішень, що оцінює емоційну реакцію людини стосовно переглянутого зображення, як позитивну або негативну.

Першим кроком загального алгоритму розв'язання задачі є визначення обличчя на фотографії. Метод повинен відпрацьовувати досить швидко, оскільки система повинна працювати в режимі реального часу при перегляді зображень користувачем. Двома іншими важливими критеріями є стійкість до наявності шуму у вхідних даних та освітлення фотографії. Додатковим кроком у попередньої обробці зображення є вирівнювання обличчя (позиціонування). Цей крок дозволяє системі отримати стійкі до поворотів голови у тривимірному просторі вхідні дані. Наступний крок – кодування обличчя. Головним результатом роботи алгоритму кодування обличчя є представлення вхідного зображення у вигляді масиву числових даних.

Фінальний крок – це прийняття рішення стосовно емоційної оцінки переглянутого зображення. На цьому етапі система повинна побудувати розв'язок на базі алгоритму машинного навчання. Навчання системи відбувається у режимі реального часу з використанням вхідного масиву даних, похибка навчання визначається як різниця між реальною відповіддю користувача та відповіддю системи, отриманої у результаті прогнозування. Для визначення обличчя на фотографії необхідно застосувати метод, який є стійким до геометричних, фотометричних похибок та яскравості обличчя.

У роботі для однозначного визначення обличчя на фотографії було реалізовано дескриптор ознак [3], відомий як гістограма спрямованих градієнтів. В основі цього методу лежить припущення, що вигляд розподілу градієнтів інтенсивності зображення дозволяє досить точно визначити наявність і форму розташованих на ньому об'єктів.

Для опису зображення здійснюється його дискретизація на кілька невеликих ділянок, комірок прямокутної форми. В комірках обчислюються гістограми h_i спрямованих градієнтів внутрішніх точок. Потім гістограми внутрішніх точок об'єднуються в єдину гістограму $h = f(h_1, \dots, h_k)$, яка нормалізується за яскравістю із використанням $L1$ або $L2$ – норми:

$$h_{\sqrt{L_1}} = \sqrt{L_1}, \quad h_{L_2} = \frac{h}{\sqrt{|h|_2^2 + \varepsilon}}, \quad h_{L_1} = \frac{h}{\sqrt{|h|_1 + \varepsilon}},$$

де ε – деяка мала константа.

Таким чином, даний дескриптор містить просторову інформацію про зображення і є інваріантним до висвітлення зображення.

При обчисленні градієнтів проводиться згортка зображення з ядрами $[-1, 0, 1]$ и $[-1, 0, 1]^T$ в результаті чого утворюються дві матриці D_x та D_y похідних уздовж осей x та y відповідно. Означені матриці використовуються для обчислення кутів і величин (модулів) градієнтів в кожній точці зображення. Нехай множина кутів $(-\pi, \pi]$ розбивається на n рівних інтервалів, що мають вигляд

$$\left(-\frac{k-1}{n}\pi, \frac{k}{n}\pi \right],$$

де $k = \overline{1, n}$.

Кожному інтервалу ставиться у відповідність промінь гістограми. Тоді гістограма комірки заповнюється у такий спосіб: величина градієнта в кожній внутрішній точці додається до величини променя, що відповідає інтервалу, який містить кут даного градієнта.

Для обчислення інтегральних градієнтів вихідного зображення $X \in R^{m \times n}$ уздовж осей x та y , дві різницьові матриці D_x та D_y обчислюються у такий спосіб:

$$D_x(x, y) = X(x, y) - X(x+1, y), \quad x \in \{0, \dots, m-1\}, \quad y \in \{0, \dots, n\},$$

$$D_y(x, y) = X(x, y) - X(x, y+1), \quad x \in \{0, \dots, m\}, \quad y \in \{0, \dots, n-1\}.$$

З їх допомогою обчислюються матриці кутів A і модулів V градієнтів:

$$A(x, y) = \arctan \frac{D_x}{D_y}, \quad V(x, y) = \sqrt{D_x(x, y)^2 + D_y(x, y)^2}$$

де $x \in \{0, \dots, m-1\}$, $y \in \{0, \dots, n-1\}$

Для побудови гістограми з n променів матриці A та V зводяться до дискретного вигляду:

$$A_n(x, y) = \left[A(x, y) \frac{n}{2\pi} \right],$$

$$V_i(x, y) = \begin{cases} V(x, y), & \text{якщо } A_n(x, y) = i; \\ 0, & \text{якщо } A_n(x, y) \neq i, \end{cases}$$

де n — кількість променів, $i \in \{0, \dots, n-1\}$

Множина матриць $I(V_i)$ буде називатись променями інтегральних градієнтів зображення.

Експериментально було виявлено [10], що оптимальна якість розпізнавання досягається при кількості променів порядку восьми. Далі до променів інтегральних градієнтів застосовується операція проріджування.

Обчислення дескрипторів здійснюється в процесі сканування зображення на різних масштабах з деяким ковзаючим вікном U . Нехай $U = (U_x, U_y)$ — розміри вікна, $dU = (dU_x, dU_y)$ — приріст кроку, $V = (V_x, V_y)$ — розмір комірок, з яких складається вікно U , а $dV = (dV_x, dV_y)$ — відстань між комірками.

Введемо також поняття одиниці довжини $E = (E_x, E_y)$, як вектор максимальної довжини, до якого кратні вектори U , dU , V , и dV . При $E \neq (1, 1)$, не впливаючи на остаточний результат, інтегральне зображення може бути проріджене так, щоб нова одиниця довжини стала дорівнювати одиничному вектору. Проріджена матриця $T(Y)$ інтегрального зображення $Y = (IX)$ обчислюється наступним чином:

$$\forall Y \in R^{(m+1) \times (n+1)} \exists Z \in R^{\left(1 + \left\lceil \frac{m}{E_x} \right\rceil\right) \times \left(1 + \left\lceil \frac{n}{E_y} \right\rceil\right)}, Z = T(Y): Z(u, v) = Y(u * E_x, v * E_y)$$

При проріджуванні всі параметри ковзання зберігають свої значення при всіх масштабах ковзного вікна, змінюється тільки одиниця довжини E . При цьому відношення dU_x/U_x та dU_y/U_y мають постійні значення, тому якість розпізнавання є інваріантною до розміру ковзного вікна. Отримавши дескриптор зображення, система поетапно порівнює його з шаблоном обличчя, однозначно визначає положення обличчя на зображенні та трансформує вхідне зображення таким чином, щоб на ньому залишилось лише обличчя.

Для розв'язання задачі позиціонування було реалізовано алгоритм, відомий під назвою оцінка орієнтирів обличчя (face landmark estimation). Основна ідея полягає в тому, що на обличчі людини визначаються 68 особливих точок (званих орієнтирами), які існують на кожному обличчі характерними серед них є: верхня частина підборіддя, зовнішня точка кожного ока, внутрішня точка кожної брови та інші. Положення особливих точок на будь-якому обличчі визначається за допомогою алгоритму машинного навчання.

В роботі [9] показано, що вирівнювання обличчя можна зробити за допомогою каскаду функцій регресії. В даному випадку кожна функція регресії в каскаді здатна ефективно оцінювати форму від початкової оцінки і інтенсивності розрідженого набору пікселів, проіндексованих щодо цієї початкової оцінки. Форма обличчя $S = [x_1, y_1, \dots, x_{N_{fp}}, y_{N_{fp}}]^T$ складається з N_{fp} орієнтирів обличчя. Мета вирівнювання обличчя полягає в тому, щоб оцінити форму S , яка максимально наближена до істинної форми обличчя S_t , тобто мінімізувати вираз $\|S - S_t\|$. У роботі було використано посилену регресію, щоб об'єднати T слабких регресивів (R^1, \dots, R^T) адитивним чином. Використовуючи зображення обличчя I і початкову форму обличчя S_0 , кожен з регресорів обчислює приріст форми δS , а потім оновлює форму обличчя каскадним чином:

$$S^t = S^{t-1} + R^t(I, S^{t-1}), \quad t = 1, \dots, T$$

де t -ий слабкий регресор R^t оновлює попередню форму S^{t-1} до нового виду S^t .

Маючи N навчальних прикладів, регресори (R^1, \dots, R^T) послідовно навчаються до тих пір, поки значення помилки навчання не стабілізується. Кожен з регресорів R^t навчається шляхом явної мінімізації суми помилок вирівнювання:

$$R^t = \arg \min_R \sum_{i=1}^N \left\| \hat{S}_i - \left(S_i^{t-1} + R(I_i, S_i^{t-1}) \right) \right\|$$

де S_i^{t-1} – форма, яка обчислюється на попередньому кроці.

Розв'язання задачі кодування обличчя здійснюється з використанням згорткової нейронної мережі глибокого навчання. Метою навчання мережі є визначення характерних точок для кожного обличчя. Під час навчання мережі водночас здійснюється аналіз трьох зображень: #1 – зображення відомою людиною, обране для навчання мережі; #2 – інше зображення тієї ж людини; #3 – зображення зовсім іншої людини. Навчання нейронної мережі відбувається до тих пір, поки вимірювання, отримані для зображень #1 і #2, будуть більш схожі, а вимірювання для зображень #2 і #3 – менш схожі. Процес навчання згорткової нейронної мережі для отримання карти осіб вимагає великої кількості даних і потужного комп'ютера. У роботі використовується існуюча у відкритому доступі нейронна мережа [7].

На останньому кроці відбувається прийняття рішення реакції людини на переглянуте зображення. Для цього використовується штучна нейронна мережа типу персептрон. Топологія мережі складається з чотирьох шарів: вхідного, двох прихованих та вихідного. На вхідному шарі було розташовано 128 нейронів; приховані шари мають 64 та 32 нейрона відповідно; вихідний

шар складається з двох нейронів. Для навчання такої мережі було реалізовано алгоритм зворотного поширення помилки, який належить до класу алгоритмів навчання з вчителем. Оскільки мережа навчається у режимі реального часу у процесі роботи з користувачем, у системі було реалізовано алгоритм безперервного навчання. Редагування ваг зав'язків мережі, відбувається на кожному кроці навчання, що дозволяє покращувати результат передбачення з кожним наступним кроком. Запропонований підхід було реалізовано у вигляді пакету прикладних програм.

Результати. При розробці кожного з етапів роботи програми (алгоритмів) проводились тестування та аналіз якості роботи. Аналіз якості визначення обличчя людини на зображенні виконано у обчислювальному експерименті. Для експерименту було відібрано 2103 довільних зображень без обличчя та 1698 довільних зображень з обличчям з відкритої бази зображень Open Images Dataset V4. У якості входу програмі надавалось зображення, на вихід програма давала відповідь у вигляді булевої змінної – true, або false, що виражає присутність обличчя на поданому зображенні. У табл. 1 подано порівняння результатів розпізнавання облич на зображеннях з обличчями та без них. Слід зазначити, що відносна похибка ідентифікації менша у випадку використання зображень без облич, оскільки такі зображення можуть містити велику кількість зайвої інформації, що миттєво відкидається. Як видно з таблиці, з використанням запропонованого підходу відносна похибка не перевищує 1% при відносно великій швидкості роботи алгоритму.

Таблиця 1.

Результат тестування дескриптору ознак для ідентифікації облич на зображенні

Типи зображень	Результат ідентифікації	Дійсне значення	Відносна похибка, %
Зображень з обличчям	1711	1698	0.7
Зображень без обличчя	2109	2103	0.2

У таблиці 2 наведено результати прогнозування позитивної реакції людини на переглянуте зображення на ітераціях алгоритму.

У обчислювальному експерименті користувачеві демонструвалось зображення з обличчям, під час перегляду відбувався аналіз стану обличчя глядача із використанням контрольних точок обличчя. Розроблена система надавала прогноз у вигляді відсотку імовірності позитивної реакції глядача на зображення, що демонструвалось. Під ітерацією мається на увазі порядковий номер зображення, що демонструється; під прогнозом мається на увазі імовірність позитивної відповіді користувача, визначена на інтервалі $[0;1]$, що отримано із використанням побудованої системи; під дійсною відповіддю розуміється обраний користувачем варіант реакції (1 – позитивна, 0 – негативна). Під час експерименту результат прогнозу не демонструвався користувачеві аби не впливати на подальші його відповіді.

Таблиця 2

Результат тестування моделі прийняття рішень

Ітерація	Прогноз (позитивна реакція)	Дійсна відповідь
1	0.781	1
10	0.677	0
25	0.602	0
50	0.364	0
75	0.79	1
100	0.891	1
150	0.155	0
200	0.923	1
250	0.931	1
300	0.939	1
350	0.921	1
400	0.95	1
450	0.045	0
500	0.943	1

Висновки. Запропонований у даній роботі комплекс алгоритмів успішно реалізовано та перевірено у практичних умовах. Даний підхід є стійким до вхідних даних за наявності шуму та здатен до самовдосконалення з плином часу.

Бібліографічні посилання

1. **Галушкін, А. І.** Теорія нейронних мереж. Кн. 1: Посібник для вузів [Текст] / А. І. Галушкін. – М., 2000. – 416 с.
2. **Головко, В. А.** Нейронні мережі: навчання, організації та використання [Текст] / Головко В. А., – М., 2001. – 256 с.
3. **Южаков, Г.Б.** Алгоритм быстрого построения дескрипторов изображения, основанных на технике гистограмм ориентированных градиентов // Информатика, математическое моделирование, экономика. - Москва: ТРУДЫ МФТИ, 2013. – Том 5, № 3, с. 84-91.
4. **Aarabi, P.** The automatic measurement of facial beauty / P. Aarabi, D. Hughes, K. Mohajer, M. Emami // IEEE International Conference on Systems, Man, and Cybernetics, 710 October 2001, Tucson, USA. V. 4. P. 2644-2647.
5. **Dalal, N.** Histograms of Oriented Gradients for Human Detection. [Текст] / Dalal N., Triggs B. // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 20-25 June 2005 San Diego, USA.
6. **Eisenthal Y., Dror G., Ruppin E.** Facial Attractiveness: Beauty and the Machine / Y. Eisenthal, G. Dror, E. Ruppin // Neural Computation. – 2006. – V. 18, № 1. – P. 119-142.
7. "Face recognition with deep neural networks." – [Електронний ресурс]. – Режим доступу: <https://github.com/cmusatyalab/openface>
8. **Gan J., Li L., Zhai Y.** Deep self-taught learning for facial beauty prediction / J. Gan, L. Li, Zhai Y. Deep // Neurocomputing. DOI: 10.1016/j.neucom.2014.05.028

9. **Salton G.** Introduction to Modern Information Retrieval. [Текст] / Gerard Salton – New York: McGraw-Hill, 1983 – 400 с.
10. **Vahid Kazemi** One Millisecond Face Alignment with an Ensemble of Regression Trees. [Текст] / Vahid Kazemi, Josephine Sullivan // IEEE Conference on Computer Vision and Pattern Recognition. – 2014.
11. **Xavier P.** Robust face landmark estimation under occlusion [Текст] / Xavier P. Burgos-Artizzu, Pietro Perona, Piotr Dollar // 2013 IEEE International Conference on Computer Vision. 1-8 Dec. 2013. Sydney, NSW, Australia.

Надійшла до редколегії 17.09. 2018.