УДК 0048:681.3 DOI: 10.26661/2413-6549-2018-1-05

ЗАСТОСУВАННЯ АНСАМБЛЕВОГО НАВЧАННЯ В ЗАДАЧАХ КЛАСИФІКАЦІЇ АКУСТИЧНИХ ДАНИХ

Кривохата А. Г., Кудін О. В., к. ф.-м. н., Давидовський М. В., Лісняк А. О., к. ф.-м. н.

Запорізький національний університет,
вул. Жуковського, 66, м. Запоріжжя, 69600, Україна

avk256@gmail.com

Сьогодні розробка машин з чуттєвими можливостями, такими як зір та слух, є однією з визнаних складних проблем у техніці та інформатиці. Системи, які мають можливість визначати сенс з аудіовізуальної інформації, усе частіше використовуються як в науковій галузі, так і в промисловості. Отже, безумовно, є необхідність у ефективних підходах для автоматичного розпізнавання звукових, графічних та відеоданих. На думку авторів, методи машинного навчання мають бути висвітлені насамперед в цьому контексті як найбільш популярні та перспективні засоби розробки подібних проблем. У запропонованій статті розглядаються моделі та методи машинного навчання, що використовуються для вирішення проблеми класифікації акустичних даних різного походження, таких як мова, музика, звуки природи тощо. Одним з практично важливих напрямів у рамках даного сімейства проблем є розробка систем машинного слуху. Іншим, особливо важливим напрямом є розробка автоматизованих систем призначення міток звукозаписам (оцінка схожості треків, системи рекомендацій музичних записів тощо), де під «міткою» розуміється коротке ім'я, яке певним чином характеризує звуковий файл. Важливо зазначити, що для вирішення вищезгаданих проблем не існує єдиного підходу. Отже, необхідно проводити більш детальний аналіз різних методів машинного навчання. В основному процес автоматизованого класифікування звуку можна розділити на чотири етапи: обробку аудіоданих, вилучення характеристик, застосування алгоритмів машинного навчання та оцінку точності. На етапі аудіопредставлення вихідний акустичний сигнал піддається сегментації на короткі фрагменти за допомогою деякої віконної функції. Загальний підхід тут полягає в тому, щоб перетворити акустичний сигнал на кадри певної довжини. Отримання компактного зображення акустичних характеристик сигналу є метою стадії вилучення характеристик. На цьому етапі використовуються спеціальні коефіцієнти, такі як швидкість нульового переходу, форма спектра, короткочасні перетворення Фур'є, мел-частотні кепстральні коефіцієнти тощо. Аудіокласифікація традиційно включає такі методи машинного навчання, як метод K-середніх, SVM, KNN, дерева рішень та інші. Протягом останніх двох десятиліть методика глибокого навчання також отримала популярність для задач класифікації. У цьому контексті слід звернути увагу на методи, засновані на згорткових та рекурентних нейронних мережах. Глибокі нейронні мережі можуть мати достатню точність при роботі як з вихідним акустичним сигналом, так і з набором вилучених характеристик. Своєю чергою етап оцінки точності використовує методи оцінки якості побудованої моделі. У роботі пропонується короткий огляд сучасних методів машинного навчання та методів, що використовуються для автоматичної класифікації акустичних даних. Наведено математичні основи методів машинного навчання та проаналізовано їх сильні та слабкі сторони. Розроблено модель та її програмну реалізацію для класифікації акустичних даних на основі згорткових нейронних мереж та побудови ансамблю нейронних мереж.

Ключові слова: акустичні дані, класифікація, машинний слух, машинне навчання.

APPLYING OF ENSEMBLE METHODS IN ACOUSTIC DATA CLASSIFICATION

Kryvokhata A. G., Kudin O. V., Davidovsky M. V., Lisnyak A. O.

Zaporizhzhya National University,
Zhykovsky str., 66, Zaporizhzhya, 69600, Ukraine

avk256@gmail.com

Today, developing machines with sensing capabilities, such as vision and hearing is one of the acknowledged challenging problems in engineering and computer science. The systems which have the ability to extract meaning from audiovisual information are increasingly used both in academia and industry. Thus, more effective approaches for automatic recognition of sound, image, and video data are definitely needed. To the authors' opinion, the machine learning methods should primarily be

highlighted in this context as the most popular and promising means for solving this sort of problems. The proposed article discusses the models and methods of machine learning used to solve the problem of classifying acoustic data of various origins, such as speech, music, sounds of nature, etc. One of the practically important directions within the scope of this family of problems is the development of machine hearing systems. Another particularly important direction is the development of automated audio tagging systems (estimating song similarity, music recommendation systems, etc.), where "tag" is understood as a short name for a label applied to some audio by an automatic tagging algorithm. It is important to notice that there is no "one-fits-all" approach to address the abovementioned sort of problems. Thus, a finer-grained look at various machine learning techniques is needed. Basically, automated audio tagging systems can be roughly decomposed into four parts: audio representation, features extraction, machine learning algorithm, and accuracy estimation. Audio representation stage implies that a raw signal is subject to segmentation into shorter signal chunks by some windowing process. A common approach here is to convert the original acoustic signal to the frames of a certain length. Receiving a compact representation of the acoustic characteristics of a signal is the aim of the feature extraction stage. This stage exploits special coefficients such as Zero-crossing rate, Spectrum shape, Short-Time Fourier Transform and Mel-frequency cepstral coefficients. Audio classification traditionally involves such machine learning methods as K-means, SVM, KNN, decision trees to name a few. During the last two decades, the deep learning based methods have also gained popularity for audio tagging. The methods based on convolutional neural networks or recurrent neural networks should be referenced in this context. Deep neural networks can benefit from operating on both raw acoustic signal and features extracted from it. In turn, accuracy estimation stage deploys quality assessment methods. In this paper, we propose a brief survey of the state-of-the-art machine learning approaches and methods used for automated classification of acoustic data. We study the mathematical foundations of the overviewed methods and analyze their strengths and weaknesses. Further we build a proof-of-concept system for the classification of acoustic data on the basis of convolutional neural networks and construct a neural network ensemble. In addition, we outline a direction for further development of machine hearing systems based on our analysis and experimental model. The approach which follows this direction can use different types of ensemble learning methods with classifiers based on feature extraction and deep neural networks.

Key words: acoustic data, classification, machine hearing, machine learning.

ВСТУП

Сучасний розвиток засобів телекомунікації та поширеність інструментів для редагування вмісту інтернет-сайтів призводить до того, що в глобальній мережі Інтернет поряд з текстовою інформацією великого поширення набувають мультимедійні дані різного вмісту, зокрема акустичні дані. Прикладом акустичних даних можуть бути музичні записи, записи лекцій, доповідей, записи звуків різного походження тощо. Для можливості пошуку серед таких даних зазвичай використовуються метадані, які описують у текстовому вигляді вміст відповідного аудіофайлу. Формування таких метаданих виконується вручну, що не завжди зручно при обробці великих об'ємів даних. Тому актуальною задачею є розробка автоматизованих систем класифікації акустичних даних.

Прикладом автоматизованих систем обробки мультимедійних даних є рекомендаційні системи, які пропонують користувачам певний контент залежно від даних, указаних у профілі користувача та історії попередніх запитів. Також актуальним напрямом в останні роки є машинний слух [17]. Однією із задач цього напряму є розробка ефективних методів класифікації звуків різного походження, наприклад мови, музики, природних звуків тощо. При цьому, найбільш досліджуваними є саме задачі аналізу музики та мови [4]. Іншою задачею яка досить часто розглядається авторами, є виявлення звукових подій. Ця задача спрямована на обробку неперервного акустичного сигналу та перетворення його в символічні описи відповідних звукових подій, присутніх на слуховій сцені [17].

У загальному вигляді процес аналізу цифрових акустичних даних зазвичай складається з декількох етапів. На початковому етапі виконується попередня обробка неперервного акустичного сигналу з метою представлення його у дискретному цифровому вигляді. При цьому зазвичай використовується ряд стандартних підходів [8, 20]. Далі виділяються ознаки акустичного сигналу, суттєві для розв'язання поставленої задачі аналізу. Серед найбільш широко вживаних ознак використовуються коефіцієнти перетворення Фур'є та автокореляції, мел-кепстральні коефіцієнти, хромограми. Після цього отримані ознаки

використовуються як вхідні параметри математичної моделі (наприклад алгоритму класифікації, кластеризації або нейронної мережі). На заключному етапі виконується верифікація отриманих результатів та впровадження розробленої системи акустичного аналізу.

Серед великої кількості оглядових робіт, присвячених тематиці розробки систем машинного слуху, можна виділити декілька, які є найбільш загальними. Так, в оглядових статтях [4, 6, 23] наводиться опис компонент системи автоматичної класифікації звуків, яка містить модулі попередньої обробки, екстракції ознак, алгоритм навчання та модуль обчислень.

У [4] детально розглянуто підходи до виділення ознак сигналу. Наводяться критерії, за якими можна класифікувати мову, музику та природні звуки. Виділяються методи, засновані на фізичних властивостях сигналів та особливостях людського сприйняття звуків. Частіше за все використовуються методи виділення ознак, які представляють акустичний сигнал у таких областях: часовій, частотній, кепстральній та вейвлет.

Огляди [6, 24] містять аналіз загальних підходів та публікацій з автоматичної класифікації музичних записів за жанрами. Пропонується множина найбільш інформативних міток, які можуть використовуватись як класи при навчанні класифікаторів. Розглядаються найбільш вживані джерела розмічених акустичних даних, які можуть використовуватися в системах навчання з учителем. Зазвичай це відкриті музичні бази в мережі Інтернет, розмічені користувачами записи, наприклад у соціальних мережах, та дані, які згенеровано спеціально для розв’язання задач машинного слуху. У роботі [24] окремо розглядається питання оцінки ефективності систем класифікації музичних файлів за жанрами.

Для класифікації даних за певними ознаками можуть використовуватися як статистичні методи (класифікатор Баєса [10, 20], дискримінантний аналіз [7, 13], EM алгоритм тощо), так і методи, які ґрунтуються на мірах схожості та відмінності (метод k-середніх [8, 15, 21], метод опорних векторів [8, 13, 19], метод k найближчих сусідів [8, 9] тощо).

В останні роки все більше робіт присвячено використанню нейронних мереж як при вилученні ознак з даних, так і безпосередньо при класифікації [5, 7, 9, 16, 17, 19, 21, 23, 25-27].

У роботі [12] визначається онтологія, що формалізує набір аудіоданих, який може використовуватися для навчання систем машинного слуху. Онтологія, яка вводиться в статті, визначає систему можливих категорій звуків для розпізнавання. Пов’язаний із цією роботою ресурс research.google.com/audioset містить набір розмічених вручну звукових роликів з YouTube (понад 2 мільйони файлів). Кожен з 10-ти секундних сегментів може відноситись до одного чи декількох класів онтології. Також відомими платформами з даними для систем машинного слуху є freesound.org, DCASE (dcase.community) та певною мірою kaggle.com [10, 11].

З аналізу літературних джерел можна зробити висновок, що задача класифікації акустичних даних і загалом розробки систем машинного слуху є досить актуальною. Опубліковані на сьогодні наукові статті можна умовно розділити на три категорії.

До першої відносять роботи, у яких виконується попередня обробка сигналу з метою сегментації та вилучення ознак, далі навчається класифікатор, на вхід якого подаються вектори ознак. У цих роботах зазвичай застосовується перетворення Фур’є, обчислення мел-частотних кепстральних коефіцієнтів та інших частотних або спектральних характеристик сигналу. З класифікаторів частіше застосовуються метод опорних векторів, k-найближчих сусідів, дерева прийняття рішень, метод k-середніх, нейронні мережі та інші. Може також застосовуватися ансамбль декількох класифікаторів, у такому випадку клас-переможець визначається шляхом голосування.

До другої категорії можна віднести публікації, в яких автори намагаються автоматизувати процес побудови оптимального набору ознак для застосування класифікаторів. Серед підходів, які застосовуються для такої автоматизації, можна виділити генетичні алгоритми та нейронні мережі. Класифікатори використовуються ті ж, що і в публікаціях попередньої категорії.

У публікаціях третьої категорії застосовуються підходи глибинних нейронних мереж. Часто це згорткові нейронні мережі, на вхід яких можуть подаватися як дані без попередньої обробки, так і набори ознак акустичних даних. Ефективність такого підходу пояснюється багатошаровою архітектурою згорткових нейронних мереж. Передбачається наявність декількох типів шарів: шари згортки, у яких виділяються певного виду ознаки, агрегувальні шари, у яких відбувається зменшення розмірності та декілька повністю зв'язних шарів, у яких виконується класифікація. До недоліків такого підходу можна віднести складність налаштування нейронних мереж зі складною архітектурою та вимогливість до обчислювальних ресурсів. Реалізація глибинних нейронних мереж зазвичай потребує системи паралельних та розподілених обчислень, залучення графічних процесорів GPU.

Метою цієї роботи є побудова системи класифікації акустичних даних різного походження на основі згорткових нейронних мереж та застосування підходу [14] для побудови ансамблю нейронних мереж. Перевагою такого підходу є його висока ефективність та адаптивність з точки зору вимогливості до обчислювальних ресурсів, оскільки, за необхідності, можна коригувати кількість класифікаторів, які входять до ансамблю.

ПОПЕРЕДНЯ ОБРОБКА ТА ОБЧИСЛЕННЯ ОЗНАК СИГНАЛУ

На практиці безпосередній аналіз звукового сигналу в часовій області (залежність амплітуда-час) майже не застосовується, оскільки не є досить ефективним та вимагає додаткових часових та просторових ресурсів для збереження і обробки даних. Для найбільш раціонального представлення акустичного сигналу використовують класичні методи цифрової обробки сигналів. Серед них можна виділити перетворення, які розкладають сигнал за ортогональними базисними функціями: перетворення Фур'є, Хартлі, Мелліна, вейвлет тощо, а також різноманітні ознаки сигналу, які обчислюються на базі цих перетворень, наприклад мел-кепстральні коефіцієнти, центроїди, енергія сигналу тощо [1]. Часто використовуються одиниці виміру, які пов'язані з психофізичними особливостями людського сприйняття частоти та сили звуку: мел, барк, фон тощо. Наприклад, мел – це одиниця суб'єктивної частоти звуку, яка сприймається людиною. Вона пов'язана з частотою сигналу (f , Гц) таким співвідношенням [1]:

$$m = 1127,01048 \ln\left(1 + \frac{f}{700}\right).$$

Крім представлення сигналу в часовій області, зазвичай використовується також частотна область (амплітуда-частота, спектр) та кепстральна.

Нехай $x(m)$ – вихідний цифровий сигнал, тобто сигнал, який отримано з аналогового шляхом дискретизації за часом та квантуванням за рівнем. Тоді дискретне перетворення Фур'є задається таким рівнянням [1]:

$$S(n) = \sum_{m=0}^{N-1} x(m) e^{-i\frac{2\pi}{N}mn}, \quad n = \overline{0, N-1},$$

а зворотне дискретне перетворення Фур'є [1]:

$$x(m) = \frac{1}{N} \sum_{n=0}^{N-1} S(n) e^{i\frac{2\pi}{N}mn}, \quad m = \overline{0, N-1}.$$

Послідовність $S[n]$ називається спектром сигналу $x[m]$. Для того, щоб обчислити кепстр потужності сигналу, необхідно обчислити зворотне перетворення Фур'є від логарифма модуля спектра [4]:

$$C(m)=\frac{1}{N}\sum_{m=0}^{N-1}\ln\left|S(n)\right|^2 e^{i\frac{2\pi}{N}mn},\quad m=\overline{0,N-1},$$

де

$$\left|S(n)\right|^2=\left|\mathrm{Re}\,S(n)\right|^2+\left|\mathrm{Im}\,S(n)\right|^2.$$

Перетворення сигналу з часової області у спектр або кепстр дозволяє отримати більш компактне та наглядне представлення інформації. Наприклад, одна гармоніка, яка в часовій області може мати безліч точок, перетворюється на одну точку в спектрі.

Зазвичай на практиці використовується швидке перетворення Фур'є в комбінації з віконною функцією, яка дозволяє обробляти звуковий сигнал по частинах довжиною в декілька мілісекунд. Для отриманих значень спектра кожного вікна t у задачах класифікації звуку додатково обчислюються такі ознаки сигналу.

Перетин нульової позначки. Обчислюється як кількість змін знаку послідовних значень цифрового сигналу [1, 4, 7, 8, 18]:

$$Z_t=\frac{1}{2}\sum_{m=1}^{N}\left|sign\left(x(m)\right)-sign\left(x(m-1)\right)\right|,$$

де функція $sign(\ )$ приймає значення 1 або 0 відповідно для позитивних та негативних чисел.

Спектральний центроїд. Визначається як центроїд, або центр ваги, значень спектра сигналу [1, 4, 7, 8, 18, 22]:

$$C_t=\frac{\sum_{n=0}^{N-1}n\left|S(n)\right|}{\sum_{n=0}^{N-1}\left|S(n)\right|}.$$

Спектральна пропускна здатність p-го порядку. Визначається через поняття центроїду [1, 4, 7, 8, 18]:

$$B_t^p=\frac{\sum_{n=0}^{N-1}(n-C_t)^p\left|S(n)\right|^p}{\sum_{n=0}^{N-1}\left|S(n)\right|^p}.$$

Енергія сигналу. Є важливим параметром при обробці сигналів з перемінною силою звуку, наприклад голосових сигналів [1, 4, 7, 8, 18]:

$$E_t=\sum_{m=0}^{N-1}\left(x(m)\right)^2.$$

Спектральна частота згортання. Визначає частоту R_t, нижче якої знаходиться k-та частина значень спектра [1, 4, 7, 8, 18]:

$$\sum_{m=0}^{R_t}S(n)=k\sum_{n=0}^{N-1}S(n),$$

де k змінюється від 0 до 1; на практиці зазвичай використовують значення 0.8–0.9.

Мел-частотні кепстральні коефіцієнти. Для отримання коефіцієнтів вихідний цифровий сигнал підлягає перетворенню Фур'є, потім виконується перехід у мел-шкалу, накладання цифрових фільтрів та обчислення кепстру. Ці коефіцієнти широко застосовуються в задачах класифікації звуку, розпізнавання мовлення тощо.

Окрім указаних, практично універсальних ознак, можуть використовуватися й інші, більш спеціалізовані. Наприклад, для аналізу музичних даних використовуються так звані хромограми. Хромограма – це проекція спектра звуку в 12 класів, які представляють собою 12 півтонів музичної октави. Очевидно, що таке представлення ефективне для класифікації або кластеризації музичних композицій.

Обчислення наведених та інших ознак звукових даних, які часто застосовуються на практиці, може відбуватися за допомогою таких спеціалізованих бібліотек, як наприклад Python-бібліотека Librosa.

Після виконання попередньої обробки звуку та обчислення векторів характерних ознак можуть застосовуватися класифікатори, такі як метод опорних векторів, метод k-найближчих сусідів, дерева прийняття рішень.

ВИКОРИСТАННЯ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Останніми роками для розпізнавання зображень та звуку широко застосовуються підходи, засновані на використанні глибинних нейронних мереж. Часто це згорткові нейронні мережі, на вхід яких можуть подаватися як дані без попередньої обробки, так і набори виділених ознак даних. При цьому точність отриманої моделі для задач розпізнавання та класифікації часто перевищує класичні підходи машинного навчання. Ефективність такого підходу пояснюється багатошаровою архітектурою згорткових нейронних мереж з використанням декількох типів шарів, які чергуються один з одним, утворюючи складну структуру.

У шарах згортки виконується афінне перетворення вхідного шару, який називається мапою ознак, у вихідний, який є, відповідно, входом для наступного шару нейронів. Нехай в l-му шарі згортки виконується лінійне перетворення двомірних вхідних даних, які представлені матрицею $X^l_{M\times N}$ у вихідну матрицю $Y^l_{M'\times N'}$ $(M' \le M, N' \le N)$ за допомогою вагової матриці $W_{d\times d}$ $(d < M, N)$:

$$y^l_{i',j'} = \sum_{a,b=0}^{d-1} W_{a,b} x^l_{i+a,j+b} .$$

Матриця $W_{d\times d}$ називається ядром згортки, а її значення визначаються алгоритмом навчання нейронної мережі. Зазвичай це варіанти методу градієнтного спуску, такі як пакетний або стохастичний градієнтний спуск. Фактично ядро згортки – це вікно перетворень, яке переміщується мапою ознак, обчислюючи результуючі значення. Коефіцієнти ядра згортки можуть застосовуватись декілька разів до різних областей мапи ознак. Також операція згортки призводить до того, що кожне значення на виході згорткового шару залежить тільки від декількох вхідних значень, тоді як у повнозв'язній нейронній мережі кожне значення залежить від усіх входів. Отже, операцію згортки можна інтерпретувати як виділення певних локальних ознак вхідних даних. Чим більше в нейронній мережі шарів згортки, тим більше локальних ознак може бути виділено [3, 13, 28].

У класичній архітектурі згорткової мережі, після обчислення згортки, до мапи ознак застосовується нелінійна функція активації. Це може бути одна із функцій, які використовуються в нейронних мережах: логістична функція, гіперболічний тангенс, ReLU тощо [3, 13, 28].

Шари субдискретизації виконують зменшення розмірності мапи ознак, яка поступає на вхід з попереднього шару згортки. Зазвичай це перетворення виконується шляхом проходження по мапі ознак вікном, як це відбувається на етапі згортки, однак при субдискретизації обирається максимальне (або середнє) значення серед тих, що потрапили у вікно. Цей процес можна інтерпретувати як узагальнення ознак та, як наслідок, зменшення мапи ознак [3].

Останні шари згорткових нейронних мереж є багатошаровим перцептроном. Після декількох ітерацій послідовного виконання згортки та субдискретизації на вхід перцептрона поступає набір досить абстрактних ознак, які використовуються для навчання при класифікації [3, 13].

Завдяки своїй структурі глибинні нейронні мережі, при вдалому налаштуванні, можуть знаходити закономірності у вихідних даних та використовувати їх для розв'язання задач класифікації. Сучасні архітектури згорткових нейронних мереж можуть використовувати десятки шарів. Так, архітектура VGG (англ. Visual Geometry Group) використовує 19 шарів «згортка-нелінійність-субдискретизація», а GoogLeNet та DenseNet (англ. Dense Convolutional Network) відповідно 22 та 250 шарів [3]. Кількість шарів сучасних ResNet (англ. Residual Convolutional Network) вже близько тисячі. Також розробляються спеціальні архітектури для пристроїв з обмеженими ресурсами, наприклад MobileNets, SqueezeNet [3].

Реалізація глибинних нейронних мереж зазвичай потребує системи паралельних та розподілених обчислень, часто із залученням графічних процесорів (NVIDIA GPU), та може займати тижні [3, 13]. На цей момент існує декілька бібліотек (TensorFlow, Keras, Theano, Caffe, PyTorch, CUDA), у яких реалізовано основні архітектури глибинних нейронних мереж, зокрема згорткові мережі.

ВИКОРИСТАННЯ МЕТОДІВ АНСАМБЛЕВОГО НАВЧАННЯ

Ансамблеве навчання передбачає об'єднання декількох моделей, наприклад класифікаторів, в одну модель з наступним узгодженням результатів усіх моделей за деяким алгоритмом. Дослідження свідчать, що ефективність ансамблю зазвичай вище ефективності окремих моделей [13]. При цьому висувається вимога відсутності кореляції між моделями, що входять до ансамблю.

Одним з підходів до побудови ансамблів є використання різних типів класифікаторів (наприклад дерева прийняття рішень, SVM, класифікатор Баєса тощо) та формування спільного прогнозу шляхом простого голосування, обчислення середнього або спеціального алгоритму машинного навчання. Альтернативою може бути навчання однакових типів класифікаторів на різних підмножинах тренувальних даних з подальшим осередненням результату прогнозу. Такий підхід іноді називається беггінгом. Ще один клас методів ансамблевого навчання – методи підсилення (бустінгу). Методи цього класу (AdaBoost, градієнтний бустінг) послідовно навчають класифікатори таким чином, щоб наступний підсилював результат попереднього [13].

Останнім часом саме використання ансамблів глибинних нейронних мереж призводить до розвитку в практичному застосуванні машинного навчання [3, 13, 15]. Але, незважаючи на високу точність, ансамблеве навчання нейронних мереж застосовується не так широко, як ансамблі більш класичних методів машинного навчання. Це пояснюється високою вимогливістю до часових та просторових ресурсів.

Більшість робіт із застосування ансамблю нейронних мереж направлена на дослідження методів генерації спільного результату з отриманих результатів навчених класифікаторів.

У роботі [14] пропонуються замість навчання M нейронних мереж навчати одну мережу. Основна ідея полягає в тому, що при застосуванні методу стохастичного градієнтного спуску запам'ятовуються значення вагової матриці при потраплянні у M точок локального мінімуму. Після цього для кожної з M вагових матриць генерується відповідна нейронна

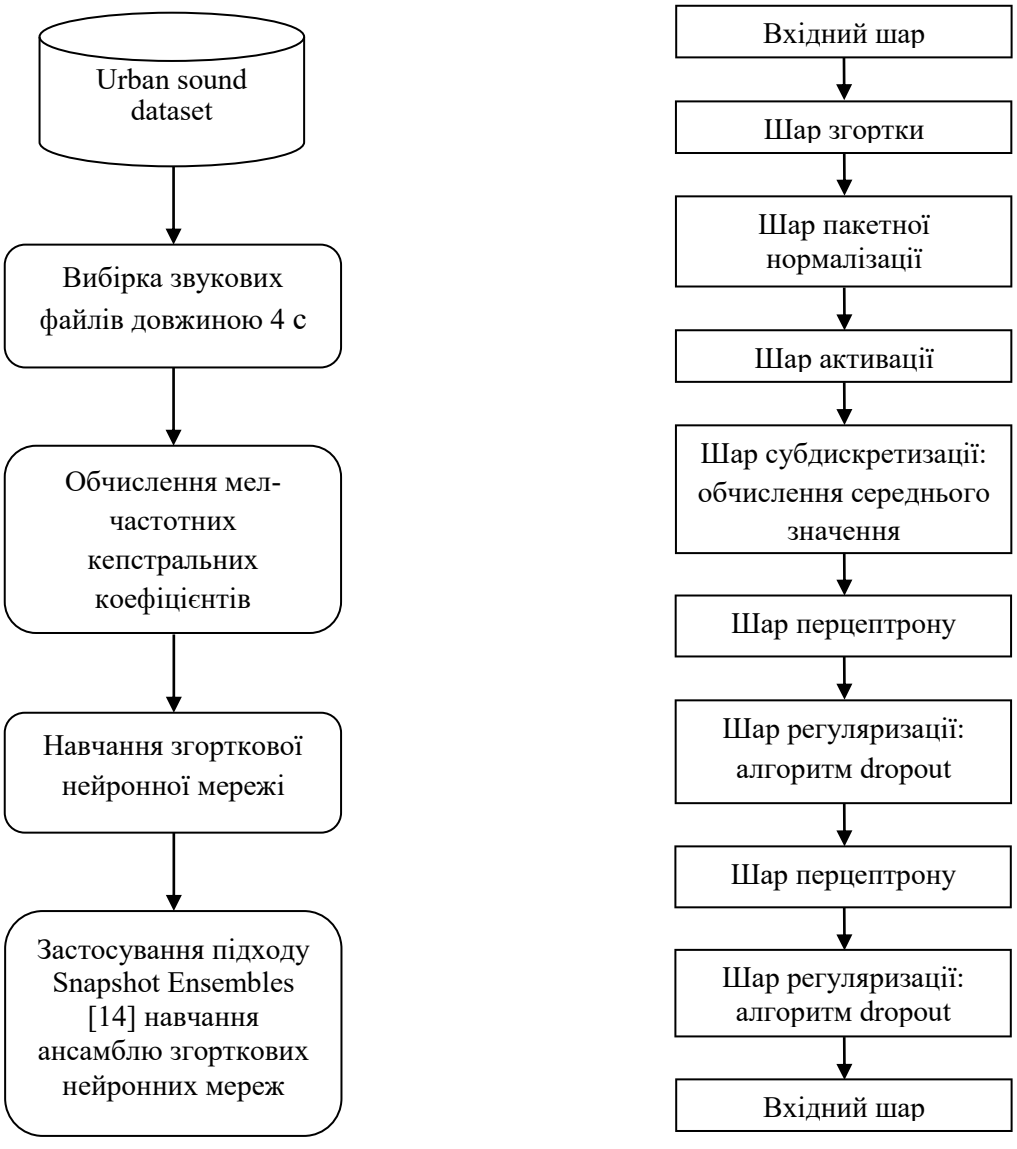
мережа. Отже, час навчання ансамблю майже не відрізняється від часу навчання однієї нейронної мережі.

ОПИС ТЕСТОВОГО НАБОРУ АКУСТИЧНИХ ДАНИХ

Для тестування запропонованих підходіву в роботі використовуються дані з ресурсу www.kaggle.com, а саме – набір даних Urban Sound Classification [11, 23]. Набір даних містить 3449 звукових файлів у форматі .wav для навчання та тестування системи. Вибірка для навчання містить звукові файли з 9 категорій. Здебільшого це вуличні міські звуки: шум транспорту, автомобільні сирени, дитячий сміх тощо. Мінімальна кількість файлів в одній категорії – 94, максимальна – 300. Тривалість звукових файлів переважно становить 4 с.

ОПИС МОДЕЛІ

Розв'язання задачі класифікації акустичних даних виконується в декілька етапів (рис. 1, а), основне призначення яких – підготовка даних та подальше навчання класифікаторів.



а) б)

Рис. 1. Етапи розв'язання задачі та структура нейронної мережі

При цьому згорткова нейронна мережа, яка використовується безпосередньо для класифікації, має структуру, зображену на рис. 1, б. Одними з основних складностей, які виникають при навчанні нейронних мереж, є визначення способу початкової ініціалізації ваг

мережі та проблема перенавчання. Параметри нейронної мережі та деталі реалізації засобами Python пакету Keras наведені на рис. 2.

Повна реалізація наведена за посиланням https://www.kaggle.com/avk256/fork-of-urban-sound-class-using-cnn-snapshot-ensem?scriptVersionId=5625604.

```
kernel_initializer='lecun_uniform'
bias_initializer='zeros'
kernel_regularizer=None
activation = "selu"
############ Визначення шарів згорткової нейронної мережі
model = models.Sequential()
model.add(Conv2D(128, 32, 32, border_mode="same",
                 input_shape = input_shape,kernel_initializer=kernel_initializer,
                 bias_initializer=bias_initializer, kernel_regularizer=None))
model.add(BatchNormalization())
model.add(Activation(activation))
model.add(AveragePooling2D())
############ Додавання повнозв'язного шару
model.add(Flatten())
model.add(Dense(1024,kernel_initializer=kernel_initializer,bias_initializer=bias_initializer))
model.add(Activation("relu"))
model.add(Dropout(0.6))
model.add(Dense(1024,kernel_initializer=kernel_initializer,bias_initializer=bias_initializer))
model.add(Activation("relu"))
model.add(Dropout(0.8))
model.add(Dense(9,kernel_initializer=kernel_initializer,bias_initializer=bias_initializer))
model.add(Activation('softmax'))
############ Компіляція моделі
############ Головні параметри ансамблю
M = 150 # Кількість станів, які зберігаються
nb_epoch = T = 200 # Кількість епох навчання
alpha_zero = 0.0001 # Коефіцієнт швидкості навчання
model_prefix = 'Model_'
snapshot = SnapshotCallbackBuilder(T, M, alpha_zero)
optimizer = optimizers.Nadam(lr=alpha_zero, beta_1=0.9, beta_2=0.999,
                             epsilon=None, schedule_decay=0.004)
model.compile(loss = "categorical_crossentropy", optimizer = optimizer,
              metrics = ["accuracy"])
history = model.fit(X_mfcc_train, y_train, batch_size = batch_size,
                    epochs = nb_epoch, verbose=2, validation_data = (X_mfcc_test, y_test),
                    callbacks=snapshot.get_callbacks(model_prefix=model_prefix))
```

Рис. 2. Програмна реалізація згорткової нейронної мережі

Точність побудованої моделі та функція похибки залежно від епох навчання наведені на рис. 3. Наводяться значення для навчальної та тестової вибірки.

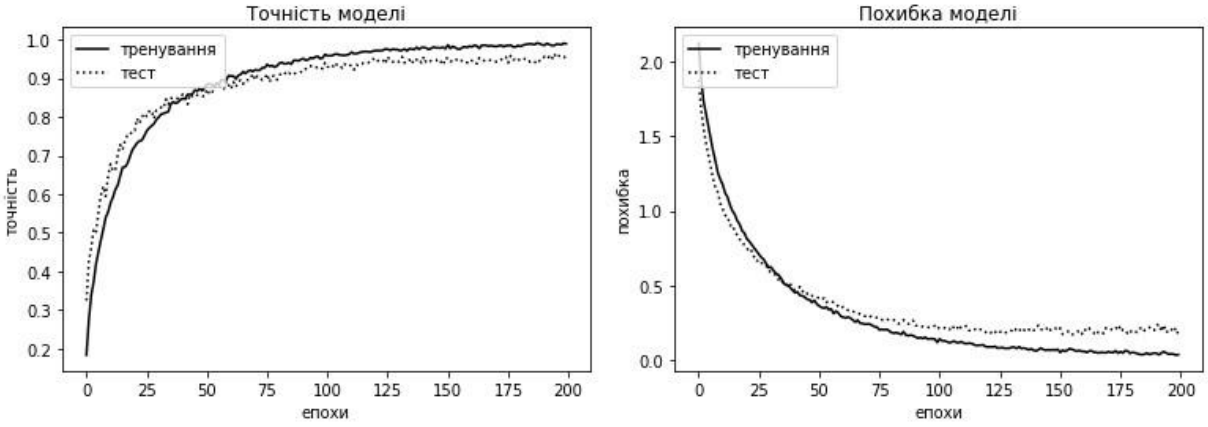


Рис. 3. Точність та похибка розробленої моделі

Ефективність прогнозування розробленої моделі за кожним класом можна розглянути за допомогою матриці розбіжностей. Індекси за строками позначають фактичні дев’ять класів, а за стовпцями – класи, спрогнозовані за допомогою розробленої моделі. Отже, матриця

розбіжностей ідеального класифікатора є діагональною. Для тестування розробленої моделі обрана стратифікована вибірка з 382 файлів вибірки для навчання.

$$Confusion\_matrix = \begin{bmatrix} 47 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 48 & 0 & 1 & 3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 41 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 35 & 3 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 47 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 42 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 48 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 43 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 15 \end{bmatrix}.$$

Значення матриці розбіжностей свідчать про задовільну точність класифікації. Так, спрогнозовані значення класів з шостого по восьмий повністю збігаються з фактичними класами, розміченими для даних звукових файлів.

ВИСНОВКИ

Побудований у роботі ансамбль згорткових нейронних мереж дозволяє досить ефективно розв'язувати задачу класифікації акустичних даних. Точність моделі на тестових даних становить 96,5%. Слід зазначити, що для розв'язку поставленої задачі достатньою виявилася структура нейронної мережі всього з одним пакетом шарів згортки-активації-субдискретизації. Перспективи подальших досліджень пов'язані з поширенням розглянутого підходу на акустичні дані складнішої структури: відмінність звукових файлів за довжиною, якістю запису тощо.

ЛІТЕРАТУРА

1. Бондарев В. Н., Трёстер Г., Чернега В. С. Цифровая обработка сигналов: методы и средства. Харьков: Конус, 2001. 398 с.
2. Кривохата А. Г., Кудін О. В., Лісняк А. О. Огляд методів машинного навчання для класифікації акустичних даних. *Вісник Херсонського національного технічного університету.* 2018. №3(66), Т.1. С. 327–331.
3. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. Санкт-Петербург: Питер, 2018. 480 с.
4. Alías F., Socoró J. C., Sevillano X. A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Applied Sciences*. 2016. № 6(5):143.
5. Bach J.-H., Meyer A.-F., McElfresh D., Anemüller J. Automatic classification of audio data using nonlinear neural response models. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan*. 2012. P. 357–360.
6. Bertin-Mahieux T., Eck D., Mandel M. Automatic tagging of audio: the state-of-the-art. *Machine audition: principles, algorithms and systems. IGI Global*. 2011. P. 334–352.
7. Burges C. J. S., Platt J. C., Jana S. Extracting noise-robust features from audio data. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, USA, 13–17 May 2002*. 2002. P. 1021–1024.
8. Camastra F., Vinciarelli A. Machime learning for Audio, Image and Video analysis. London: Springer-Verlag, 2015. 561 p.
9. Costa C. H. L., Valle Jr. J. D., Koerich A. L. Automatic classification of audio data. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics.* 2004. P. 562–567.
10. Dongge Li, Ishwar K. Sethi, Nevenka Dimitrova, Tom McGee. Classification of general audio data for content-based retrieval. *Pattern Recognition Letters*. 2001. № 22(5). P. 533–544.

11. Free sound General-Purpose Audio Tagging Challenge. URL: https://www.kaggle.com/c/freesound-audio-tagging/data (Дата звернення 06.06.2018).

12. Gemmeke J. F., Ellis D. P. W., Freedman D., Jansen A., Lawrence W., Moore R. C., Plakal M., Ritter M. Audio set: an ontology and human-labeled dataset for audio events. *Proceedings of the Acoustics, Speech and Signal Processing International Conference*. 2017. URL: https://research.google.com/pubs/archive/45857.pdf (Дата звернення 06.06.2018).

13. Geron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Sebastopol: O`Reilly. 2017. 861 p.

14. Howel J., Rooth M., Wagner M. Acoustic classification of focus: on the web and in the lab. *Laboratory Phonology: Journal of the Association for Laboratory Phonology.* 2017. № 8(1):16. P. 1-41.

15. Huang G. et al. Snapshot Ensembles: Train 1, Get M for Free. arXiv, 2017. URL: http:// arxiv.org/abs/1704.00109 (Дата звернення 06.06.2018).

16. Ibrahim Z. Al A., Ferrane I., Joly P. Audio Data Analysis Using Parametric Representation of Temporal Relations. *IEEE International Conference on Information and Communication Technologies: from Theory to Applications (ICTTA).* 2006.

17. Kong Q., Xu Y., Wang W., Plumbley M. D. Convolutional gated recurrent neural network incorporating spatial features for audio tagging. *The 2017 International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, Alaska*. 2017.

18. Kong Q., Xu Y., Wang W., Plumbley M. D. A joint separation-classification model for sound event detection of weakly labelled data. *ICASSP 2018 - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 15-20 April 2018, Calgary, Canada. 2018.

19. Lyon R. F. Machine Hearing: An Emerging Field. *IEEE Signal Process. Mag.* 2010, Vol. 27. P. 131–139.

20. Mierswa I., Morik K. Learning feature extraction for learning from audio data. *Technische Universität Dortmund. Technical Reports*. 2004. No. 55.

21. Oppenheim A. V., Schafer R. W. Discrete-Time Signal Processing. Third edition. Pearson Education Limited. 2014. 1055 p.

22. Rizzi A., Buccino M., Panella M., Uncini A. Optimal short-time features for music/speech classification of compressed audio data. *International Conference on Intelligent Agents. 28 November-1 December 2006. Sydney, NSW, Australia.*

23. Salamon J., Jacoby C., Bello J. P. A dataset and taxonomy for urban sound research. DOI: http://dx.doi.org/10.1145/2647868.2655045, 2017. P. 1–4.

24. Stastný J., Skorpil V., Fejfar J. Audio Data Classification by Means of New Algorithms. *36th International conference on Telecommunications and Signal Processing,* Rome, Italy. 2013. P. 507–511.

25. Sturm B. L. A Survey of Evaluation in Music Genre Recognition. *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation. AMR 2012. Lecture Notes in Computer Science*. 2014. Vol 8382. P. 29–66.

26. Wichern G., Yamada M., Thornburg H., Sugiyama M., Spanias A. Automatic audio tagging using covariate shift adaptation. *IEEE international conference Acoustics speech and signal processing (ICASSP), 14–19 March 2010*.

27. Xu Y., Huang Q., Wang W., Foster P., Sigtia S., Jackson P. J. B., Plumbley M. D. Unsupervised Feature Learning Based on Deep Models for Environmental Audio Tagging. *IEEE/ACM transactions on audio, speech and language processing*. 2017. Vol 25(6). P. 1230–1241.

28. Zaccone G., Karim Md. R. Deep learning with TensorFlow. Packt Publishing. 2018. 767 p.

REFERENCES

1. Bondarev, V.N., Tryoster, G. & Chernega, V.S. (2001). Digital signal processing: methods and tools. Harkov: Konus (In Russian).

2. Kryvokhata, A.G., Kudin, O.V. & Lisnyak, A.O. (2018). A Survey of Machine Learning Methods for Acoustic Data Classification. Visnyk of Kherson National Technical University, No. 3(66), pp. 327–331.

3. Nikolenko, S., Kadurin, A. & Arhangelskaya, E. (2018). Deep learning. Saint Petersburg: Piter (In Russian).
4. Alías, F., Socoró, J.C. & Sevillano X. (2016). A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. Applied Sciences, No. 6(5):143.
5. Bach, J.-H., Meyer, A.-F., McElfresh, D. & Anemüller, J. (2012). Automatic classification of audio data using nonlinear neural response models. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (pp. 357–360), Kyoto, Japan.
6. Bertin-Mahieux, T., Eck, D. & Mandel, M. (2011). Automatic tagging of audio: the state-of-the-art. Machine audition: principles, algorithms and systems. IGI Global, pp. 334–352.
7. Burges, C.J.S., Platt, J.C. & Jana, S. (2002). Extracting noise-robust features from audio data. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), (pp. 1021–1024), Orlando, FL, USA, 13–17 May 2002.
8. Camastra, F. & Vinciarelli, A. (2015). Machime learning for Audio, Image and Video analysis. London: Springer-Verlag.
9. Costa, C.H.L., Valle, Jr. J.D. & Koerich, A.L. (2004). Automatic classification of audio data. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp. 562–567.
10. Dongge, Li, Ishwar, K. Sethi, Nevenka, Dimitrova & Tom, McGee (2001). Classification of general audio data for content-based retrieval. Pattern Recognition Letters, No. 22(5), pp. 533–544.
11. Free sound General-Purpose Audio Tagging Challenge. Retrieved from https://www.kaggle.com/c/freesound-audio-tagging/data.
12. Gemmeke, J.F., Ellis, D.P.W., Freedman, D., Jansen, A., Lawrence, W., Moore, R.C., Plakal, M. & Ritter, M. (2017). Audio set: an ontology and human-labeled dataset for audio events. Proceedings of the Acoustics, Speech and Signal Processing International Conference, 2017. Retrieved from https://research.google.com/pubs/archive/45857.pdf.
13. Geron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow. Sebastopol: O`Reilly.
14. Howel, J., Rooth, M. & Wagner, M. (2017). Acoustic classification of focus: on the web and in the lab. Laboratory Phonology: Journal of the Association for Laboratory Phonology, No. 8(1):16, pp. 1-41.
15. Huang, G. et al. (2017). Snapshot Ensembles: Train 1, Get M for Free. Retrieved from http:// arxiv.org/abs/1704.00109.
16. Ibrahim, Z. Al A., Ferrane, I., & Joly, P. (2006). Audio Data Analysis Using Parametric Representation of Temporal Relations. IEEE International Conference on Information and Communication Technologies: from Theory to Applications (ICTTA), 2006.
17. Kong, Q., Xu, Y., Wang, W. & Plumbley, M.D. (2017). Convolutional gated recurrent neural network incorporating spatial features for audio tagging. The 2017 International Joint Conference on Neural Networks (IJCNN 2017), Anchorage, Alaska.
18. Kong, Q., Xu, Y., Wang, W. & Plumbley, M.D. (2018). A joint separation-classification model for sound event detection of weakly labelled data. ICASSP 2018 - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 15-20 Apr 2018, Calgary, Canada.
19. Lyon, R.F. (2010). Machine Hearing: An Emerging Field. IEEE Signal Process. Mag., Vol. 27, pp. 131-139.
20. Mierswa, I. & Morik, K. (2004). Learning feature extraction for learning from audio data. Technische Universität Dortmund. Technical Reports, No. 55.
21. Oppenheim, A.V. & Schafer, R.W. (2014). Discrete-Time Signal Processing. Third edition. Pearson Education Limited.
22. Rizzi, A., Buccino, M., Panella, M. & Uncini, A. (2006). Optimal short-time features for music/speech classification of compressed audio data. International Conference on Intelligent Agents. 28 Nov.-1 Dec. 2006. Sydney, NSW, Australia.
23. Salamon, J., Jacoby, C. & Bello, J.P. (2017). A dataset and taxonomy for urban sound research, pp. 1-4. doi: http://dx.doi.org/10.1145/2647868.2655045.

24. Stastný, J., Skorpil, V. & Fejfar, J. (2013). Audio Data Classification by Means of New Algorithms. 36th International conference on Telecommunications and Signal Processing 2013, (pp. 507-511), Rome, Italy.

25. Sturm, B.L. (2014). A Survey of Evaluation in Music Genre Recognition. Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation. AMR 2012. Lecture Notes in Computer Science, Vol 8382, pp. 29-66.

26. Wichern, G., Yamada, M., Thornburg, H., Sugiyama, M. & Spanias, A. (2010). Automatic audio tagging using covariate shift adaptation. IEEE international conference Acoustics speech and signal processing (ICASSP), 14-19 Mar 2010.

27. Xu, Y., Huang, Q., Wang, W., Foster, P., Sigtia, S., Jackson, P.J.B. & Plumbley, M.D. (2017). Unsupervised Feature Learning Based on Deep Models for Environmental Audio Tagging. IEEE/ACM transactions on audio, speech and language processing, Vol 25(6), pp. 1230-1241.

28. Zaccone, G. & Karim, Md. R. (2018). Deep learning with TensorFlow. Packt Publishing.

УДК 519.172 DOI: 10.26661/2413-6549-2018-1-06

АЛГЕБРАИЧЕСКИЕ МЕТОДЫ ПОСТРОЕНИЯ ТОПОЛОГИЧЕСКОГО РИСУНКА НЕПЛАНАРНОГО ГРАФА

Курапов С. В., к. ф.-м. н., доцент, Сгадов С. А.

Запорожский национальный университет,
ул. Жуковского, 66, г. Запорожье, 69600, Украина

lilili5050@rambler.ru, sgadovsa@bk.ru

В данной работе рассматриваются алгебраические методы построения топологического рисунка несепарабельного непланарного графа. Базовым понятием для построения топологического рисунка является теория вращений. Приведен метод выбора оптимального маршрута для проведения соединения на дуальном графе циклов. Рассмотрена операция включения ранее не проведенных ребер в плоский топологический рисунок.

Ключевые слова: граф, вращение вершины, топологический рисунок.

АЛГЕБРАЇЧНІ МЕТОДИ ПОБУДОВИ ТОПОЛОГІЧНОГО РИСУНКА НЕПЛАНАРНОГО ГРАФА

Курапов С. В., к. ф.-м. н., доцент, Сгадов С. А.

Запорізький національний університет,
вул. Жуковського, 66, м. Запоріжжя, 69600, Україна

lilili5050@rambler.ru, sgadovsa@bk.ru

У цій роботі розглядаються алгебраїчні методи побудови топологічного рисунка несепарабельного непланарного графа. Доведено, що граф довільного виду можна розбити на блоки, кожен із яких являє собою максимальний нероздільний підграф. Показана структурна перебудова сепарабельного графа в несепарабельну частину графа. Застосування методів теорії обертання вершин дозволяє описувати топологічний рисунок плоскої частини графа. Так само обертання вершин індукує прості орієнтовані цикли графа. І це в підсумку дозволяє будувати топологічний рисунок графа алгебраїчними методами, не проводячи ніяких геометричних побудов на площині. Плоска частина графа будується алгоритмом виділення підмножини ізометричних циклів з потужністю, рівною цикломатичному числу графа, що характеризується мінімальним значенням функціоналу Маклейна з подальшим видаленням мінімальної кількості ребер. Подальша побудова для проведення з'єднань, віддалених у процесі планаризації, здійснюється пошуком маршрутів для пересічних з'єднань. Наведено метод вибору оптимального маршруту для проведення з'єднання на дуальному графі циклів. Розглянуто операцію включення раніше не проведених ребер у плоский топологічний рисунок.

Ключові слова: граф, обертання вершини, топологічний рисунок.