

УДК 004.934

Донченко В.С.¹, д.ф.-м.н., проф. В.С.,
Голік А.О.², аспірант А.О.

Застосування еліпсоїдальної та ортогональної відстаней відповідності у задачах розпізнавання мови та жестів.

Проведені паралелі між задачею розпізнавання мови одного диктора на обмеженій множині слів та розпізнаванням дактильної мови жестів. Наведений варіант формування векторів ознак у матричному вигляді для обох задач. Запропоновано використати еліпсоїдальну та ортогональну відстані.

Ключові слова: розпізнавання мови та жестів, ортогональні проектори, еліпсоїдальна відстань, псевдообернення, SVD – розклад.

¹ Київський національний університет імені Тараса Шевченка, 03680, м. Київ, пр-т Глушкова 4д, e-mail: voldon@bigmir.net

² Київський національний університет імені Тараса Шевченка, 03680, м. Київ, пр-т Глушкова 4д, e-mail: analytical@ukr.net

Статтю представив д.т.н., проф. Акіменко В.В.

Серед багатьох задач, які розглядаються в рамках такого напрямку досліджень, як Computer Vision, чільне місце посідають задачі розпізнавання мови та жестів. В пропонованій статті проведені паралелі між цими задачами з подальшим виділенням під проблем, які мають схожий характер та можуть бути розв'язані з використанням однакових інструментів та підходів. Для реалізації та тестування були обрані окремі випадки вище згаданих задач: розпізнавання мови одного диктора на обмеженій множині слів та розпізнавання дактильної мови жестів.

В статті пропонується розгляд матриць, як «природніх» векторів ознак: представників об'єкта, що піддається аналізу. Такий підхід є природнім, коли аналізуються, зокрема, мовні сигнали, які, як відомо, вичерпно представляються спектрограмою, або жести, які на перших етапах обробки вхідних даних представлені у вигляді зображень (або

V.S. Donchenko¹, doctor of ph.-math. s., prof. V.S.
A.O. Golik², post-graduate student A.O.

Usage of ellipsoidal and orthogonal distances in speech and gesture recognition.

Parallels between speech recognition of one speaker on a limited set of words and recognition of a tactile sign language are drawn. Variant of formation of feature vectors in a matrix form is given for both problems. Ellipsoidal and orthogonal compliance distances are suggested.

Key Words: speech and gesture recognition, orthogonal projectors, ellipsoidal distance, pseudoinverse, SVD – decomposition.

¹ Taras Shevchenko National University of Kyiv, 03680, Kyiv, Glushkova av., 4d, e-mail: voldon@bigmir.net

² Taras Shevchenko National University of Kyiv, 03680, Kyiv, Glushkova av., 4d, e-mail: analytical@ukr.net

послідовності зображень), що захоплені з веб-камери чи іншого записуючого пристрою.

Запропоновані алгоритми розпізнавання, які використовують як стандартні уявлення про відстані відповідності, так і так звані «еліпсоїдальні відстані». Еліпсоїдальні відстані будуються на основі певного «мінімального еліпсу», що «накриває» навчальну вибірку того чи іншого класу. Крім того, в статті пропонується ще одна альтернатива – використання ортогональних відстаней, що будуються на основі кортежних групуючи операторів та ортогональних проекторів.

Кластеризація з встановленням відстаней відповідності на базі псевдо обернення та SVD-розкладу може бути успішно застосована до числових векторів. Проте актуальним є випадок, коли навчальну вибірку складають не вектори, а матриці. Однією із головних цілей наведеного дослідження є перенесення властивостей псевдо обернення та SVD-розкладу на множини матричних векторів ознак.

Розглянутий алгоритм розпізнавання «порівняння з еталоном». Наведені результати роботи реалізованих у середовищі Java + CUDA та C# + EmguCV програм розпізнавання, що обґрунтовують введення відповідних підходів, і, зокрема, відстані відповідності на базі ортогональних проекторів.

Результати свідчать про ефективність запропонованих відстаней відповідності та перспективність подальшої роботи над їх вдосконаленням.

Спектрограма як «представник» мовного сигналу в математичній моделі сигналу

Стандартним в аналізі мовних сигналів є гармонійний аналіз: представлення аудіо сигналу сумішшю гармонік. Стандартним математичним апаратом, який використовується для реалізації згаданого підходу за цифрової обробки є дискретне перетворення Фур'є у формі, яка позначається терміном швидке перетворення Фур'є (ШПФ).

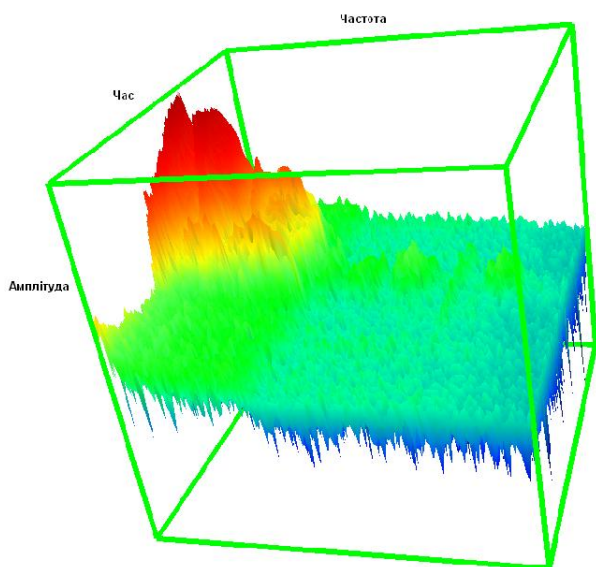


Рис 1. Тривимірний варіант представлення спектрограми мовного сигналу.

За такого підходу числовому вектору, що представляє весь дискретизований аудіо сигнал чи його частину ставиться у відповідність вектор тої самої довжини, але з комплексними компонентами – ШПФ сигналу. Комплексний вектор ШПФ сигналу стандартним чином описується через вектор модулів відповідних компонент чи вектор квадратів модулів згаданих компонент та вектором фаз відповідних компонент. Власне вектор квадратів модулів

компонент ШПФ сигналу, прив'язаний до дискретного набору частот, визначеного ШПФ, називається енергетичним спектром аналізованого сигналу. В практиці використання гармонійного аналізу замість розгляду ШПФ всього сигналу, що відповідає, наприклад вимовленому слову, використовується набір ШПФ частин сигналу у рухомому віконці фіксованої довжини, що зсувається з певним кроком. Відповідно, з досліджуваним аудіо сигналом пов'язується матриця ШПФ ділянок сигналу у кожному з віконць. Відповідно, матриця складена із векторів енергетичних спектрів сигналу для кожного з віконць називають спектрограмою сигналу, що піддається обробці.

Наявна матриця спектрограми, яка описує сигнал з точністю до фазових характеристик, власне, є природним представником аудіо сигналу в математичному описі такого сигналу. Тому питання про побудову ефективних алгоритмів розпізнавання аудіо сигналів є питанням про те, яким чином її доцільно використати для ідентифікації слів.

Контур як представник жести

Перший етап розв'язання задачі розпізнавання жестів полягає у захопленні зображення з веб-камери чи іншого записуючого пристрою з подальшим пошуком та виділенням на отриманому зображенні долоні та її контуру. Саме контур дає досить повну інформацію, що може бути використана для ідентифікації жести.



Рис 2. Зображення жести захоплене з веб-камери. Контур долоні знайдений та виділений.

Цю інформацію не можна назвати вичерпною, оскільки в більшості випадків вона не включає в себе дані про положення окремих пальців. Звісно

для відкритої долоні це не є проблемою, але коли вона хоча б частково закрита, що характерно для реальних жестів дактильної мови, дані про положення та степінь зігнутості окремих пальців отримати складно. Загалом, ця проблема може бути частково вирішена за допомогою застосування алгоритмів скелетизації. В пропонованій статті розглядається варіант розв'язання задачі розпізнавання жестів дактильної мови шляхом аналізу контуру долоні на зображенні.

Аналізувати контур можна декількома способами, наприклад, як послідовність взаємопов'язаних точок. Крім того, існує ряд числових характеристик, що можуть бути обчислені для контуру: моменти, ланцюги Фрімена тощо. Далі ж мова піде про представлення отриманого для жесту контуру у матричній формі.

Перехід до матричного вигляду починається з пошуку найменшого охоплюючого прямокутника для зображення контуру долоні.



Рис 3. Найменший охоплюючий прямокутник для зображення (контуру) долоні.

Отримавши його координати, можна вирізати з загального зображення ту частину, яка відповідає цьому прямокутнику та конвертувати її у бінарну матрицю. Детально процес конвертації буде описано у наступному розділі статті.

«Характеристичні» матриці

Ідентифікація слова у задачі розпізнавання мовного сигналу може бути здійснена як шляхом побудови тих чи інших числових векторів ознак, так і безпосередньо аналізом спектрограми у матричній формі. Аналогічне твердження має

місце і для задачі розпізнавання жестів, адже, як було згадано у попередньому розділі, контур жесту так само може бути поданий у вигляді матриці. Проте для обох випадків, актуальною є проблема стандартизації розмірності такої матриці. Одним з можливих варіантів вирішення цієї проблеми є побудова характеристичної матриці: захоплення зображення (двовимірного, кольорового) матриці спектрограми (чи контуру відповідно) та його подальше стискання чи розтягнення до встановленого стандартного розміру з подальшою конвертацією у матричну форму за певними правилами.

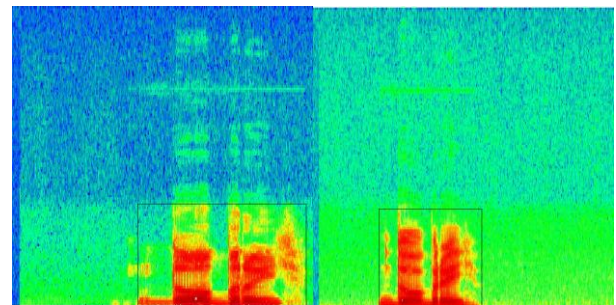


Рис 4. Приклад записів одного й того ж слова промовленого в різному середовищі.

Слід зазначити, що можна помітити схожість у під задачах для розпізнавання мови та жестів, а саме той факт, що на даному етапі постає питання стандартизації зображень – матриці спектрограми та мінімального охоплюючого прямокутника контуру – з подальшою їх конвертацією у матричну форму.

Пропонований підхід можна назвати методом «характеристичних матриць», тобто аналізу сигналу на основі матриць, побудованих певним чином за спектрограмою як зображенням (чи зображенням мінімального охоплюючого прямокутника для контуру) з використанням форматів перетворення зображень. До згаданих перетворень зображень відноситься, зокрема масштабування.

Сам процес формування характеристичної матриці для задачі розпізнавання мови починається з побудови двовимірного кольорового зображення спектрограми. З метою відокремлення інформативно цінної складової на зображенні виділяється прямокутник, який охоплює частину зображення, що відповідає сказаному слову. Пошук аналогічного прямокутника для задачі розпізнавання жестів був згаданий у попередньому розділі. Такий підхід дозволяє відкинути все до сказаного слова та після нього, а також залишити лише ті

частоти, на яких є сенс проводити аналіз. В свою чергу у застосування до зображення контуру залишається лише та область яка безпосередньо охоплює зображення долоні. Частина зображення, що потрапляє в прямокутник зберігається та направляється на подальшу обробку. Слід зазначити, що розмір розглянутих прямокутників для одного й того ж слова промовленого різними дикторами, на різній швидкості, з різною інтонацією буде відрізнятися. Так само, як будуть відрізнятися розміри мінімальних охоплюючих прямокутників для зображення контуру жесту, адже вони залежать від багатьох факторів: розмір долоні людини, відстань від долоні до записуючого пристрою ітд. Тому має місце задача стандартизації, яка може бути вирішена масштабуванням зображення. Іншими словами, розміри збереженого зображення приводяться до певних стандартних величин по висоті та ширині. Доцільним також може бути застосування згладжування та інших операцій над зображеннями.

Проте, є принципова різниця у техніці масштабування зображень для задачі розпізнавання мови та для задачі розпізнавання жестів. Коли мова йде про масштабування за віссю часу для зображення мовного сигналу, то достатньо його просто стиснути чи розтиснути, адже загалом ширина зображення відповідає швидкості вимови слова і від такого роду масштабування інформація не буде втрачена. Якщо ж розглянути задачу розпізнавання жестів, то виникає потреба у більш складному варіанті масштабування.



Рис 5. Літера «в»

На рисунку 5 наведений приклад жесту, який красномовно свідчить про необхідність внесення змін у вищезгаданий алгоритм стандартизації.

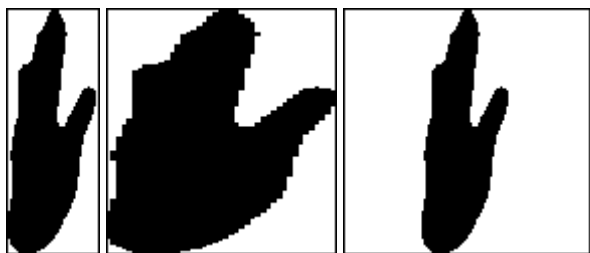


Рис 6. Різні варіанти стандартизації зображення

Виділимо на рисунку 6 3 частини: перша – побудований мінімальний охоплюючий

прямокутник для жесту, інші дві - варіанти його стандартизації. Припустимо, що в якості розміру еталону було обрано квадрат певного розміру. В такому разі розтягнувши зображення звичайним чином ми отримаємо результат, що наведений у другій частині рисунку 6. Не важко помітити, що в такому випадку зображення жесту в значній мірі втратило свою інформативну цінність, адже співвідношення ширини та висоти, яке є важливим у цій задачі, було змінено. Більш правильним є підхід ілюстрований у третій частині рисунку 6. У цьому разі додаткові порожні зони були розміщені зліва та справа від основної частини. Розмір цих зон однаковий і знайдений таким чином, щоб результуюче зображення відповідало встановленому для еталонів стандарту.

На наступному кроці пропонується зробити зворотній перехід від зображення до матриці. У зв'язку з цим нагадаймо, що RGB – формат представлення кольору, як комбінації червоного, зеленого та синього. Експериментальним шляхом встановлюються допустимі значення RGB, які дозволяють прийняти рішення: чи слід розглядати піксель на зображенні як значущий чи ні. Саме перетворення зображення полягає у заміні пікселів, які задовольняють встановлені допустимі значення RGB одиницями і нулями всіх інших. В результаті отримується матриця з нулів та одиниць, яку називатимемо характеристичною. Її зображення можна відтворити у чорно-білому вигляді, що характерно для бінарних матриць.



Рис 7. Приклад зображення «характеристичних» матриць одного слова.

Перетворення спектрограм як матриць через використання перетворень за форматами зображень має свої переваги. Зокрема, такий підхід дозволяє розв'язати проблему стандартизації спектрограми за довжиною мовного сигналу і отримати придатні до аналізу об'єкти. Аналогічним чином отримані в результаті стандартизації «характеристичні» матриці для зображень жестів дактильної мови можуть бути використані для подальшого їх розпізнавання. Єдиною відмінністю процесу

конвертації для цієї задачі є той факт, що зображення жесту може бути розглянуто у чорно-білому варіанті і встановлення допустимих значень RGB не є необхідністю, адже фактично кожен піксель однозначно є чорним або білим. Всі точки які потрапили в контур маркуються як чорні, всі інші – як білі. Тому на виході процесу конвертації отримується бінарна характеристична матриця.

Відстані відповідності: еліпсоїдальна відстань

Після формування векторів ознак на етапі кластеризації виникає необхідність у їх порівнянні, встановленні так званої відстані відповідності між двома векторами. У пропонованій статті розглянута можливість застосування еліпсоїдальної та ортогональної відстаней.

Їх основною особливістю є навчання системи шляхом роботи не з одним еталоном, а з набором (для різних умов середовища).

Еліпсоїдальна відстань будується засобами псевдо обернення для тих чи інших варіантів лінійних операторів. Така відстань спирається на концепцію «мінімальні еліпси групування». Власне, це ті еліпси які «накривають» кожен із навчальних вибірок «мінімальним» і «оптимальним» чином. Еліпсоїдальні відстані будуються для матриць як матриць лінійних операторів між матричними евклідовими просторами засобами псевдообернення для згаданих просторів. Вони реалізуються, як і у випадку векторних евклідових просторів, через так звані «групуючі оператори» теорії псевдообернення. Такі оператори визначаються за матрицею оператора A , оператора між векторними евклідовими просторами, і визначаються співвідношенням:

$$R(A) = A^+ A^{+T}, R(A^T) = (A^T)^+ (A^T)^{+T} = A^{+T} A^+.$$

Принципова роль групуючих операторів полягає в тому, що вони дають можливість побудувати «мінімальні еліпси групування»: еліпсоїди, які містять всі вектори набору $a_k, k = \overline{1, n}$ і є оптимальними в певному сенсі. Зміст оптимальності еліпсів групування полягає у тому, що його вісі утворені ортонормованим набором векторів, суми квадратів проєкцій на які є максимальними, а квадрати довжин відповідних вісей співпадають із відповідними сумами квадратів проєкцій. Точніше є справедливими наступні чотири теореми [4].

Теорема 1. Для довільного набору векторів $a_k \in R^m, k = \overline{1, n}$ розв'язком оптимізаційної задачі пошуку максимуму суми квадратів проєкцій на підпростір, породжений нормованим вектором $u \in R^m: \|u\|=1$ є вектор u_1 із сингулярності $\{u_1, \lambda_1^2\}$ сингулярного розкладу матриці $A = (a_1, \dots, a_n)$:

$$u_1 = \arg \min_{u \in R^m: \|u\|=1} \sum_{k=1}^n \|Pr_u a_k\|^2,$$

$$\min_{u \in R^m: \|u\|=1} \sum_{k=1}^n \|Pr_u a_k\|^2 = \lambda_1^2.$$

Теорема 2. Для довільного набору векторів $a_k \in R^m, k = \overline{1, n}$ розв'язком оптимізаційної задачі пошуку максимуму суми квадратів проєкцій на підпростір, породжений нормованим вектором $u \in R^m: \|u\|=1$ є вектор u_1 із сингулярності $\{u_1, \lambda_1^2\}$ сингулярного розкладу матриці $A = (a_1, \dots, a_n)$:

$$u_k = \arg \min_{u \in R^m: \|u\|=1, u \perp \{u_1, \dots, u_{k-1}\}} \sum_{k=1}^n \|Pr_u a_k\|^2,$$

$$\min_{u \in R^m: \|u\|=1, u \perp \{u_1, \dots, u_{k-1}\}} \sum_{k=1}^n \|Pr_u a_k\|^2 = \lambda_{k+1}^2,$$

$$k = \overline{1, r-1},$$

де $\{u_k, \lambda_k^2\}, k = \overline{1, r}$, як і в попередній теоремі сингулярності сингулярного розкладу матриці, побудованої із елементів досліджуваного набору векторів.

Теорема 3. Для довільного набору векторів $a_k \in R^m, k = \overline{1, n}$

$$a_k^T R(A^T) a_k \leq r_{\max}^2 < r$$

$$r_{\max}^2 = \max_{k=1, n} a_k^T R(A^T) a_k$$

де, як і в двох попередніх теоремах A матриця складена із векторів набору як із стовпчиків.

Еліпсоїд теорема 3 групує вектори набору за центральним розташуванням еліпсу групування: за еліпсом, що має центром початок координат. В практичних застосуваннях центром еліпсу вибирають середнє значення \bar{a} елементів набору:

$$\bar{a} = \frac{1}{n} \sum a_k.$$

У цьому випадку групуючий оператор будується за матрицею \tilde{A} , що будується за центрованими середніми векторами набору

$\tilde{a}_k : \tilde{a}_k = a_k - \bar{a}, k = \overline{1, n}$. Отже, справедлива наступна теорема.

Теорема 4. Для довільного набору векторів $a_k \in R^m, k = \overline{1, n}$

Виконуються наступні нерівності

$$(a_k - \bar{a})^T R(\tilde{A}^T)(a_k - \bar{a}) \leq \tilde{r}_{\max}^2 \leq r, k = \overline{1, n}$$

$$r_{\max}^2 = \max_{k=\overline{1, n}} \tilde{a}_k^T R(\tilde{A}^T) \tilde{a}_k$$

В якості набору векторів використовуються, як правило, навчальні вибірки тих чи інших класів $Kl_l, l = \overline{1, L}$. В якості відстаней відповідності (точніше їхнім квадратам): функціоналів $\rho^2(x, Kl_l), x \in R^m, l = \overline{1, L}$, за мінімумом значення яких здійснюється віднесення до класу, - можна використовувати мінімальні еліпси групування. Це означає, що відстані відповідності визначаються співвідношеннями:

$$\rho^2(x, Kl_l) = (x - \bar{a}_l)^T \frac{R(A_l^T)}{r_{1\max}^2} (x - \bar{a}_l), x \in R^m$$

$$l = \overline{1, L}$$

Саме така еліпсоїдальна відстань використовується для характеристичних матриць.

Відстані відповідності: ортогональна відстань

Разом з еліпсоїдальною відстанню у статті пропонується ортогональна відстань відповідності. Вона надає можливість перенести властивості псевдообернення та SVD- розкладу на випадок матричних векторів ознак.

$$(\varphi_\alpha y, X)_{tr} = \left(\sum_{k=1}^K y_k A_k, X \right)_{tr} = \sum_{k=1}^K y_k (A_k, X)_{tr} = \sum_{k=1}^K y_k (tr A_k^T X) = \left(y, \begin{pmatrix} tr A_1^T X \\ \dots \\ tr A_K^T X \end{pmatrix} \right)$$

,що й доводить теорему

Теорема 6[5]. Добуток двох операторів є лінійний оператор $\varphi_\alpha^* \varphi_\alpha : R^K \rightarrow R^K$, що задається матрицею, що її ототожнюватимемо із самим оператором, яка визначається співвідношенням:

$$\varphi_\alpha^* \varphi_\alpha = \begin{pmatrix} tr A_1^T A_1, \dots, tr A_1^T A_n \\ \dots \\ tr A_n^T A_1, \dots, tr A_n^T A_n \end{pmatrix} \quad (1)$$

Позначатимемо $R^{(m \times n), K}$ - евклідов простір $m \times n$ матричних кортежів довжини K $\alpha = (A_1 : \dots : A_K) \in R^{(m \times n), K}$ з «природним» по компонентним скалярним добутком:

$$(\alpha, \beta) = \sum_{k=1}^K (A_k, B_k)_{tr} = \sum_{k=1}^K tr A_k^T B_k,$$

$$\alpha = (A_1 : \dots : A_K), \beta = (B_1 : \dots : B_K) \in R^{(m \times n), K}.$$

Позначимо також $\varphi_\alpha : R^K \rightarrow R^{m \times n}$ лінійний оператор між відповідними евклідовими просторами, що задається матричним кортежем $\alpha = (A_1 : \dots : A_K) \in R^{(m \times n), K}$ і визначається за допомогою матричних кортежних операцій у відповідності із співвідношенням:

$$\varphi_\alpha y = \sum_{k=1}^K y_k A_k, \alpha = (A_1 : \dots : A_K) \in R^{(m \times n), K}, y = \begin{pmatrix} y_1 \\ \dots \\ y_K \end{pmatrix} \in R^K$$

Теорема 5[5]. Спряженим φ_α^* до оператора $\varphi_\alpha : R^K \rightarrow R^{m \times n}$ є лінійний оператор, що, очевидно, діє у зворотному до φ_α напрямку: $\varphi_\alpha^* : R^{m \times n} \rightarrow R^K$, і визначається співвідношенням:

$$\varphi_\alpha^* X = \begin{pmatrix} tr A_1^T X \\ \dots \\ tr A_K^T X \end{pmatrix}.$$

Доведення.

Дійсно,

Зауважимо, що матриця, визначена співвідношенням (1) є матрицею Грама елементів A_1, \dots, A_K матричного кортежу $\alpha = (A_1 : \dots : A_K)$, що задає оператор φ_α .

Доведення. Дійсно,

$$\wp_{\alpha}^* \wp_{\alpha} y = \wp_{\alpha}^* (\wp_{\alpha} y) = \begin{pmatrix} \text{tr} A_1^T \sum_{i=1}^n A_i y_i \\ \dots \\ \text{tr} A_n^T \sum_{i=1}^n A_i y_i \end{pmatrix} = \begin{pmatrix} \text{tr} \sum_{i=1}^n A_i^T A_i y_i \\ \dots \\ \text{tr} \sum_{i=1}^n A_n^T A_i y_i \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \text{tr} A_i^T A_i y_i \\ \dots \\ \sum_{i=1}^n \text{tr} A_n^T A_i y_i \end{pmatrix} = \begin{pmatrix} \text{tr} A_1^T A_1, \dots, \text{tr} A_1^T A_n \\ \dots \\ \text{tr} A_n^T A_1, \dots, \text{tr} A_n^T A_n \end{pmatrix} \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} = (\text{tr}(A_i^T A_j)) y$$

, що й доводить теорему.

Сингулярний розклад для матриці (1) є очевидним: це симетрична та невід'ємно визначена матриця. Він визначається набором сингулярностей $(v_i, \lambda_i^2), i, j = \overline{1, r}$:

ортонормованим набором векторів $\|v_i\| = 1, v_i \perp v_j, i \neq j; i, j = \overline{1, r}; \lambda_1 > \lambda_2 > \dots > \lambda_r > 0$,

що є власними для оператора $\wp_{\alpha}^* \wp_{\alpha} : R^K \rightarrow R^K$:

$$\wp_{\alpha}^* \wp_{\alpha} v_i = \lambda_i^2 v_i, i = \overline{1, r}.$$

Визначені за сингулярностями $(v_i, \lambda_i^2), i = \overline{1, r}$

матриці $U_i \in R^{m \times n} : U_i = \frac{1}{\lambda_i} \wp_{\alpha} v_i, i = \overline{1, r}$ є

елементами набору сингулярностей $(U_i, \lambda_i^2), i = \overline{1, r}$ оператора $\wp_{\alpha} \wp_{\alpha}^*$.

Сингулярний розклад кортежного оператора. Сингулярності двох операторів: $\wp_{\alpha}^* \wp_{\alpha}, \wp_{\alpha} \wp_{\alpha}^*$, очевидним чином визначають сингулярний розклад оператора \wp_{α} .

Теорема 7 [5]. (сингулярний розклад кортежного оператора).

$$\wp_{\alpha} = \sum_{k=1}^K \lambda_k U_k V_k^T.$$

Варіант сингулярного розкладу. Приймаючи до уваги співвідношення

$$U_i \in R^{m \times n} : U_i = \frac{1}{\lambda_i} \wp_{\alpha} v_i, i = \overline{1, r} \text{ та його}$$

наслідок $\lambda_i U_i = \wp_{\alpha} v_i, i = \overline{1, r}$, маємо

$$\wp_{\alpha} = \sum_{k=1}^K \lambda_k U_k V_k^T = \sum_{k=1}^K (\wp_{\alpha} v_k) v_k^T$$

Зауваження загального характеру: потрібний загальний варіант теореми про сингулярний розклад. Це твердження мусить стосуватися загальних евклідових просторів. Його потрібно формулювати для лінійних операторів над загальними евклідовими просторами. Його вигляд мусить бути таким.

Теорема 8 [5]. Для довільного лінійного оператора $\wp_E : E_1 \rightarrow E_2$ над парою евклідових

просторів $(E_i, (\cdot, \cdot)), i = 1, 2$ існує набір сингулярностей

$(v_i, \lambda_i^2), (u_i, \lambda_i^2), i = \overline{1, r}, r = \text{rank} \wp_E$ операторів

$\wp_E^* \wp_E, \wp_E \wp_E^*$ відповідно із спільним набором власних чисел $\lambda_i^2, i = \overline{1, r}$, що

$$\wp_E x = \sum_{i=1}^r u_i \lambda(v_i, x), \wp_E^* y = \sum_{i=1}^r v_i \lambda(u_i, y)_2.$$

Крім того, виконуються співвідношення:

$$u_i = \lambda_i^{-1} \wp_E v_i, i = \overline{1, r},$$

$$v_i = \lambda_i^{-1} \wp_E^* u_i, i = \overline{1, r}.$$

Основні оператори ПДО теорії для кортежних операторів: псевдообернення за SVD-розкладом.

У відповідності до SVD-визначення ПДО кортежного оператора задається [5] співвідношенням

$$\wp_{\alpha}^+ = \sum_{k=1}^K \lambda^{-1} v_k (U_k, \cdot)_{tr} = \sum_{k=1}^K \lambda^{-2} v_k (\wp_{\alpha} v_k, \cdot)_{tr}$$

Стандартним чином за SVD-поданням кортежного оператора визначаються ортогональні проектори базових підпросторів оператора та, відповідно, - групуючі оператори.

Теорема 9. Оператори, позначені як $P(\wp_{\alpha}^*), P(\wp_{\alpha})$ та визначені співвідношеннями

$$P(\wp_{\alpha}^*) = \sum_{k=1}^r U_k (U_k, \cdot)_{tr}$$

$$P(\wp_{\alpha}) = \sum_{k=1}^r v_k (v_k, \cdot) = \sum_{k=1}^r v_k v_k^T.$$

є ортогональними проекторами $P_{L_{\wp_{\alpha}^*}}, P_{L_{\wp_{\alpha}}}$ на підпростори $L_{\wp_{\alpha}^*}, L_{\wp_{\alpha}}$ можливих значень операторів $\wp_{\alpha}^*, \wp_{\alpha}$ відповідно:

$$P(\wp_{\alpha}^*) = P_{L_{\wp_{\alpha}^*}}, P(\wp_{\alpha}) = P_{L_{\wp_{\alpha}}}$$

а самі ці підпростори є лінійними оболонками відповідних ортонормованих наборів:

$$L_{\wp_{\alpha}^*} = L(U_1, \dots, U_r), L_{\wp_{\alpha}} = L(v_1, \dots, v_r)$$

Доведення. Доведення проводиться так само, як у випадку лінійних операторів між евклідовими просторами числових векторів:

нескладно перевіряється симетричність та ідемпотентність обох операторів.. Так само очевидними є твердження про те, що $U_k \in L_{\varphi_\alpha}, v_k \in L_{\varphi_\alpha^*}$, а, отже із міркувань розмірності $L_{\varphi_\alpha} = L(U_1, \dots, U_r), L_{\varphi_\alpha^*} = L(v_1, \dots, v_r)$.

Крім того, як впливає із визначення $P_{L_{\varphi_\alpha}}, P_{L_{\varphi_\alpha^*}}$ останні простори є для них просторами можливих значень відповідно. Для завершення доведення залишається зауважити, що підпростір, на який здійснює ортогональне проектування ортогональний проектор, може бути описаний, зокрема, як простір можливих значень для нього.

Теорема 10. Оператори $Z(\varphi_\alpha^*), Z(\varphi_\alpha)$, що є доповненнями до тотожного оператора ортогональних проекторів:

$P(\varphi_\alpha^*), P(\varphi_\alpha)$ відповідно:

$$Z(\varphi_\alpha^*)X = X - P(\varphi_\alpha^*)X, \quad Z(\varphi_\alpha) = E_K - P(\varphi_\alpha).$$

є ортогональними проекторами на ядра операторів відповідно.

Доведення. Доведення впливає із того, що, по-перше, к $\varphi_\alpha^*, \varphi_\alpha$ кожен із операторів $Z(\varphi_\alpha^*), Z(\varphi_\alpha)$ є симетричним та ідемпотентним .

$$\rho^2(X, L_{\varphi_\alpha}) = \|X_{L_{\varphi_\alpha^\perp}}\|^2 = \|Z(\varphi_\alpha^*)X\|_{tr}^2 = (Z(\varphi_\alpha^*)X, Z(\varphi_\alpha^*)X)_{tr} = (X, Z(\varphi_\alpha^*)Z(\varphi_\alpha^*)X)_{tr} = (X, Z(\varphi_\alpha^*)X)_{tr}$$

Оскільки ортонормований набір $U_i, i = \overline{1, r}$ є ортонормованим базисом в $L_{\varphi_\alpha} = L(U_1, \dots, U_r)$, а $X_{L_{\varphi_\alpha^\perp}} = \sum_{i=\overline{1, r}} X_{U_i}$ за цим

координатами розкладу $X_{L_{\varphi_\alpha^\perp}}$ за цим ортонормованим базисом, то

$$\|X_{L_{\varphi_\alpha^\perp}}\|^2 = \sum_{i=1}^r (X, U_i)_{tr}^2.$$

Залишається зауважити, що за теоремою Піфагора в абстрактному варіанті

$$\|X\|^2 = \|X_{L_{\varphi_\alpha}}\|^2 + \|X_{L_{\varphi_\alpha^\perp}}\|^2, \text{ а, отже,}$$

$$\rho^2(X, L) = \rho^2(X, L_{\varphi_\alpha}) = (X, Z(\varphi_\alpha^*)X)_{tr} = \|X\|^2 - \sum_{k=1}^r (X, U_k)_{tr}^2$$

для кортежного оператора φ_α , породженого набором A_1, \dots, A_K :

$$\varphi_\alpha = (A_1, \dots, A_K).$$

Доведення. Доведення впливає із того, що підпростори $L = L(A_1, \dots, A_K)$ та L_{φ_α} співпадають між собою

Крім того вони є ортогональними проекторами на ортогональні доповнення до підпросторів $L_{\varphi_\alpha} = L(U_1, \dots, U_r), L_{\varphi_\alpha^*} = L(v_1, \dots, v_r)$, відповідно. А саме це ортогональні доповнення і є ядрами операторів $\varphi_\alpha^*, \varphi_\alpha$ відповідно.

Теорема 11. Квадрат відстані $\rho^2(X, L_{\varphi_\alpha})$ довільної $m \times n$ матриці X до лінійного підпростору L_{φ_α} , що є множиною можливих значень кортежного оператора φ_α задається формулою

$$\rho^2(X, L_{\varphi_\alpha}) = (X, Z(\varphi_\alpha^*)X)_{tr} = \|X\|^2 - \sum_{k=1}^r (X, U_k)_{tr}^2$$

Доведення. Дійсно,

$$\rho^2(X, L_{\varphi_\alpha}) = \|X_{L_{\varphi_\alpha^\perp}}\|^2$$

у розкладі $X = X_{L_{\varphi_\alpha}} + X_{L_{\varphi_\alpha^\perp}}$ за

розкладом $R^{m \times n} = L_{\varphi_\alpha} + L_{\varphi_\alpha^\perp}$. Очевидним чином

$$X_{L_{\varphi_\alpha^\perp}} = Z(\varphi_\alpha^*)X,$$

а, отже,

$$\|X_{L_{\varphi_\alpha^\perp}}\|^2 = \|X\|^2 - \|X_{L_{\varphi_\alpha}}\|^2 = \|X\|^2 - \sum_{k=1}^r (X, U_k)_{tr}^2$$

теорема доведена.

Теорема 12. Квадрат відстані $\rho^2(X, L)$ довільної $m \times n$ матриці X до лінійного підпростору $L = L(A_1, \dots, A_K)$, що є лінійною оболонкою набору $m \times n$ матриць A_1, \dots, A_K

Задається формулою

Теорема 13. Квадрат відстані $\rho^2(X, \Gamma(\bar{a}, L))$ довільної $m \times n$ матриці X до гіперплощини $\Gamma(\bar{a}, L)$:

$$\bar{a} = \frac{1}{K} \sum_{k=1}^K A_k, L = L(\tilde{A}_1, \dots, \tilde{A}_K), \tilde{A}_k = A_k - \bar{a}, k = \overline{1, K}$$

, породженого набором

$m \times n$ матриць A_1, \dots, A_K задається формулою

$$\rho^2(X, \Gamma(\bar{a}, L)) = (X - \bar{a}, Z(\varphi_{\bar{a}}^*)(X - \bar{a}))_{tr} = \|X - \bar{a}\|^2 - \sum_{k=1}^r (X - \bar{a}, \tilde{U}_k)_{tr}^2$$

, в якій кортежний оператор $\varphi_{\bar{a}}$ визначається

співвідношенням $\varphi_{\bar{a}} = (\tilde{A}_1, \dots, \tilde{A}_K)$, а

$\tilde{U}_i, i = \overline{1, r}$ ортонормований набір власних матриць оператора $\varphi_{\bar{a}}^*$.

Доведення. Доведення очевидним чином випливає з того, що

$$\rho^2(X, \Gamma(\bar{a}, L)) = \rho^2(X - \bar{a}, L)$$

та попередньої теореми.

Алгоритм розпізнавання

В статті пропонується алгоритм розпізнавання «порівняння з еталоном». Після того як сформований словник слів чи жестів, для кожного з елементів зі словника формується та зберігається набір характеристичних матриць, які відповідають записам слів чи зображенням жестів в різних умовах середовища. Цей словник використовується у процесі кластеризації. Після перетворення вхідного сигналу у відповідну характеристичну матрицю, ця матриця, використовуючи одну з розглянутих у статті відстаней відповідності, перевіряється на близькість до кожного елементу зі словника. Елемент, який виявляється найближчим у термінах відповідної відстані відповідності приймається за результат.

Тестування та результати

Для тестування довільним чином була відібрана множина із 32 – ох слів (діалог). На базі цієї множини були проведені експериментальні дослідження. Тестування розпізнавання жестів проводилося на множині дактилів.

Спеціально розроблений модуль навчання сформував навчальну вибірку (базу еталонів) для кожного елементу зі словника, використовуючи характеристичні матриці. В залежності від конфігурації системи, використовувалась еліпсоїдальна чи ортогональна відстань. Також для порівняльної характеристики у тестуванні була перевірена і евклідова.

Для реалізації програмної частини задачі розпізнавання мови було обране середовище Java. Це об'єктно-орієнтована мова програмування, яка охоплює багатий набір інструментарію для розробки програмного забезпечення різних напрямків. Крім того, Java підтримує роботу з CUDA- технологіями та має ряд бібліотек, зокрема JCublas, JCufft, в яких реалізовані основні математичні операції із застосуванням паралельних обчислень.

Для реалізації програмної частини задачі розпізнавання жестів була обрана мова програмування C#. Для цього середовища розроблена оболонка бібліотеки Open CV, яка називається Emgu CV. Вона включає в себе багатий інструментарій, який дозволяє працювати з потоками даних, що отримується з записуючого пристрою у реальному часі. Крім того, в ній реалізований ряд функцій та класів, які можуть бути ефективно застосовані при розпізнаванні.

Хоча запропоновані у статті відстані відповідності вимагають подальшого вивчення та оптимізації, вже на даному етапі для задач розпізнавання мови одного диктора на обмеженій множині слів та розпізнавання жестів дактильної мови, отримані математичні результати проілюстровані застосуванням у даних задачах розпізнавання показали свою працездатність. Зокрема, в задачі розпізнавання дактильної мови жестів.

Недолік застосування характеристичних матриць для задачі розпізнавання мови полягає у чутливості до вибору RGB параметрів, адже для цієї задачі конвертація пікселя малюнку в елемент 0|1 матриці не є такою тривіальною задачею як для розпізнавання жестів, де вхідний малюнок чорно-білий і може бути легко конвертований у бінарну матрицю. Тому при правильній конфігурації та ряду додаткових етапів обробки можна досягти значного покращення.

Загалом, перспективним напрямком є застосування багаторівневої кластеризації, де ряд технік та алгоритмів застосовується поетапно. Запропоновані у статті матричні вектори ознак та відстані відповідності можуть скласти основу одного з таких етапів.

Отже, в пропонованій статті були розглянуті проблеми класифікації мовних сигналів та мови жестів. Зазначимо, що природними представниками об'єктів класифікації в згаданих задачах є матриці. Запропонований розвиток математичного

апарату псевдообернення аналізу таких об'єктів на основі розвитку теорії псевдообернення для матричних евклідових просторів.

Список використаних джерел

1. *Golik A.O.* Designing the feature vectors: implementation of speech recognition algorithms using parallel computing capabilities. *Visn., Ser. Fiz.-Mat. Nayky, Kyiv Univ. im. Tarasa Shevchenka.* – 2012. – N 3. – P. 141-150. (in Ukrainian).
2. *Donchenko V.S., Golik A.O.* Implementation of algorithms of speech recognition using parallel computing capabilities // XVI International Conference “Problems of Decision Making under Uncertainties”(PDMU - 2010),– October 4-8, 2010.–Yalta, Ukraine
3. *Donchenko V.S., Golik A.O.* Using of parallel computing technology in speech recognition problems // XVII International Conference “Problems of Decision Making under Uncertainties”(PDMU -2011),– May 23-27, 2011.–Skhidnytsia, Ukraine
4. *Donchenko V.S.* Euclidian space of vectors and matrixes: constructive methods of description of basic structures and their usage // *International Journal “Information technologies & Knowledge”*.- 2011.- Vol. 5.- Number 3.-P.203-216.
5. *Donchenko V., Zinko T., Skotarenko F.* Feature vectors’ in grouping information problem in applied mathematics: vectors and matrixes /Problems of Computer Intellectualization...- Editors Vitali Velichko, Aleksey Voloshin O, Krassimir Markov.-Kyiv,Ukrain- Sofia Bulgaria: NASU, V.M.Glushkov Institute of Cybernetics, ITHEA : 2012- p.111-124.
6. *Donchenko V., Krivonos U., Omardibirova V.* Basic structures of Euclidian space: methods of description and usage / *Information Models of Knowledge*.- Editors: Krassimir Markov, Vitalii Velichko, Oleksy Voloshin - ITHEA.- Kiev, Ukraine – Sofia, Bulgaria. Number 16.– P. 155-169.- ISBN 978-954-16-0048-1
7. *Kirichenko N.F., Donchenko V.S.* Pseudoinverse in problems of grouping of information // X International Conference “System analyses and information technologies”– 20-24 May 2008 y. Kyiv. – p.86.
8. *Kirichenko N.F., Donchenko V.S.* Linear structures in problems of clustering and pseudoinverse // Abstracts: International Conference ‘Problems of Decision making under uncertainties (PDMU-2008)’.– May, 13-17, 2008. – Kyiv-Rivne, Ukraine. – 2008. – P.72-75.
9. *Kirichenko N.F., Donchenko V.S.* Hyperset clustering and pseudoinverse // Abstracts: International Conference ‘Problems of Decision making under uncertainties (PDMU-2007)’.– May, 21-25, 2007. – Chernivtsi, Ukraine.– 2007.– P.143 -144.
10. *Kirichenko N.F., Donchenko V.S.* Pseudoinverse in clustering // *Cybernetics and System Analyses*– 2007 – №4.– p.73-92.
11. *Kirichenko N.F.* Analytical representation of pseudoinverse matrixes // *Cybernetics and System Analyses* - №2. –1997.– p.98-122.
12. *Kirichenko N.F., Donchenko V.S.* Pseudoinverse and linear structures in problems of prediction and clustering // Abstracts/ International Conference “Problems of Decision Macing under Uncertainties”(PDMU -2009),– April 27-30, 2009.–Skhidnyysia, Ukraine.– p.114.
13. *Kirichenko N.F., Donchenko V.S.* Pseudoinverse in clustering // *Cybernetics and System Analyses*- №4, 2007– p.98-122.
14. *Bondhugula, V., Govindaraju, N. and Manocha, D.* 2006. Fast Singular Value Decomposition on Graphics Processors. Technical report. University of North Carolina
15. *Moore E.H.* On the reciprocal of the general algebraic matrix.//*Bull. Amer. Math. Soc.* – 26, 1920. – p. 394-395.
16. *Penrose R.* A generalized inverse for matrices.// *Proc. Cambr. Philosophical Soc.*- 51, 1955.– p.406-413.
17. *William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery.* 2.6 Singular Value Decomposition // *Numerical Recipes in C.* — 2nd edition. — Cambridge: Cambridge University Press.

Надійшла до редколегії 17.02.13