

У роботі описано розроблений метод передискретизації у випадку двох вхідних зображень на основі псевдообертання виродженого матричного оператора дивергенцій. Процедура передбачає побудову виродженого квадратного матричного оператора дивергенцій, пошук векторів-ознак на основі псевдообертання Мура-Пенроуза і синтез нового передискретизованого зображення. Сам синтез може відбуватися декількома способами. Основні два з них – це додавання синтезованого вектора значень функції інтенсивності в одне із вхідних зображень, або синтез нового зображення з використанням отриманих векторів-ознак та відповідних рядків/стовпців із вхідних зображень. Експериментально встановлено, що другий спосіб показує найкращі результати.

ЛІТЕРАТУРА:

1. Рашкевич Ю.М. Зміна роздільної здатності зображення засобом псевдообертання виродженого матричного оператора відносних симетричних мір конвергенції / Ю.М. Рашкевич, І.В. Ізонін, Д.Д. Пелешко, І.О. Малець // Вісник Національного університету “Львівська політехніка”. – 2015. – № 826: Комп’ютерні науки та інформаційні технології. – С. 259–266.
2. Пелешко Д.Д. Автоматична первинна сегментація мовного сигналу на основі симетричної матриці відстаней / Д.Д. Пелешко, М.І. Рашкевич, Ю.М. Пелешко // Наукові праці: науково-методичний журнал. – Комп’ютерні технології. – Миколаїв: Вид-во ЧДУ ім. Петра Могили. – 2014. – Вип. 225., Т.237. – С. 66–72.

СИСТЕМА АВТОМАТИЧНОГО МОНІТОРИНГУ МЕДІЙНОГО ПРОСТОРУ НА ОСНОВІ ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ СЛУХОВИХ І ЗОРОВИХ ОБРАЗІВ

Сажок М.М.², Маріковський О.В.², Мартиненко М.Р.²,
Робейко В.В.^{1,2}, Селюх Р.А.², Федорин Д.Я.²

¹Відділ розпізнавання та синтезу звукових образів,

Міжнародний науково-навчальний центр інформаційних технологій та систем, м. Київ, Україна

²ТОВ “Інтелектуальні медійні системи”, м. Київ, Україна

{ sazhok, marikovski, marynka.martynenko, valya.robeiko, vxml12, dmytro.fedoryn } @gmail.com

Сучасний світ швидко масштабує будь-які комунікаційні форми спілкування з аудиторією. Телебачення, радіо та інші сучасні мережі потокового або записаного контенту все більше потребують доступних методів пошуку в них релевантної інформації для орієнтування у тому, що відбувається в медійному просторі. Враховуючи постійне зростання аудіо-візуальних каналів комунікацій, постає питання автоматизації пошуку в них релевантної інформації та розроблення ефективного цільового методу видачі результатів. При цьому параметри запиту можуть мати мультимодальний (різнорідний) характер: слова з прямої мови, текстового ряду, наявність зображення людини, обличчя якої широко відоме тощо.

Технології розпізнавання слухових і зорових образів знаходять все нові сфери застосування в інформаційному суспільстві. Однією з таких сфер якраз і є оброблення медійної інформації, зокрема теле- і радіомовлення.

У світі існує ряд переважно експериментальних і допоміжних систем оброблення телерадіоэфіру, в яких автоматизовані генерування субтитрів і пошук інформації для англійської та ряду інших мов, здебільшого європейських [1], [2]. В основі таких систем лежить технологія розпізнавання мовленнєвого сигналу, призначена для перетворення в текст мовлення, яке надходить із певних джерел медійної інформації та відповідає наперед відомому набору телерадіопрограм (новини, інтерв’ю, телешоу, трансляції засідань парламенту та ін.). Зоровий канал сприйняття медійної інформації моделюється методами розпізнавання зображень або комп’ютерного зору. При цьому постає ряд нових задач, пов’язаних із динамікою відеопотоку та специфікою телеэфіру 3.

Авторами взято за основу розроблену раніше експериментальну систему перетворення українського телерадіомовлення в текст 4. При надходженні вхідного сигналу з деякого джерела (в нашому випадку — це мережа IPTV або файлова система) компонента реального часу — Розпізнавач — звертається до Бази даних і знань (БД і З), що формується у відкладеному часі за допомогою ряду програмних засобів на підставі звукових і текстових даних. При цьому отримана відповідь розпізнавання відображається у вигляді тексту синхронно зі звуковим сигналом для п’ятихвилинних сегментів. У представленій роботі автори модернізували БД і З, запропонували та реалізували спосіб пошуку відповідних сегментів медійної інформації за ключовими словами на підставі результатів розпізнавання, розпочали інтеграцію з зоровим каналом сприйняття інформації та здійснили ряд заходів щодо дослідної експлуатації системи моніторингу телерадіоэфіру.

БД і 3 має акустичну та лінгвістичну складові. Розвиток акустичної складової полягає в оцінці параметрів акустичних моделей фонем за мовленнєвим корпусом, в якому текстова анотація має різну ступінь наближення. Так для частини корпусу (близько 20 годин на кожну мову) проведено пофразову сегментацію — від паузи до паузи. Решта корпусу (понад 300 годин) супроводжуються текстом, прив'язаним до сюжетів цілком. Отже, для цієї частини корпусу виявилось необхідним застосування методів “напівсліпого” навчання 5.

Лінгвістична складова БД і 3 була поповнена новим текстовим матеріалом із новинних джерел. Проведено автоматизацію процедури поповнення словника новими словами за рахунок передбачення словесних наголосів з відповідним переходом до фонемної транскрипції слів. Окремо введено деревоподібний словник, що дає змогу здійснювати ітерацію серед слів на підставі часткового написання. Останнє необхідне для дотримання сучасних стандартів інтерфейсу користувача при здійсненні пошуку.

Результати розпізнавання індексуються, тобто обробляються з метою забезпечення знаходження акустичних сегментів, які містять задані при пошуку слова, а саме, — кожному слову ставляться у відповідність сегменти, де це слово спостерігається. Під час пошуку одного слова або об'єднання слів (альтернативних ключових слів), результатом пошуку оголошуються сегменти, які містять хоча б одне з заданих слів. Слова, задані частково, перетворюються на об'єднання слів за деревоподібним словником. При пошуку перетину слів (ключових фраз), результатом пошуку оголошуються сегменти, які містять усі слова в межах заданого інтервалу, наприклад, 10 секунд. При цьому враховується, що інтервал, в якому цілком містяться слова з ключової фрази, може перебувати в різних сегментах. Аналогічно здійснюється пошук об'єднання перетину слів. Більш складні пошукові конструкції реалізуються рекурсивно.

Розпочата інтеграція зорового каналу полягає у розробці підсистеми виявленні зміни сюжету 6. Це необхідно для того, щоб перейти від п'ятихвилинних фрагментів, до фрагментів, що відповідають різним сюжетам і телепередачам. Для формування навчальної вибірки було проведено ручну розмітку 15 новинних передач ТСН загальною тривалістю 20 годин. При цьому експерт позначав момент відеопотоку, що відповідав переходу до сегменту іншого змісту, наприклад, від кадрів, що знімаються у студії, до сюжету на вулиці. Надійність автоматичного виявлення таких переходів склала в середньому близько 90%.

Більшість описаних функцій реалізовано в експериментальній системі моніторингу медійного простору 7. Розроблена система передбачає обмін даними між клієнтом і сервером розпізнавання через мережу (Інтернет або локальну) за протоколом TCP/IP. Обмін відбувається з використанням REST інтерфейсу. Серверне програмне забезпечення розроблено мовами C++ (розпізнавання мовленнєвих сигналів), PERL (взаємодія з аудіовідеоданими, здійснення індексації та пошуку) та PHP (оброблення запитів від клієнта). В інтерфейсі клієнта використовуються можливості JavaScript і HTML5. На теперішній час система обробляє 10 телевізійних і 5 каналів радіомовлення. Користувач задає ключові слова, об'єднання слів або ключові фрази зі слів або об'єднань слів. При цьому джокер “зірочка” дає змогу задавати початкову частину слів. Результати пошуку подаються в хронологічному порядку, починаючи з найновішого сегменту. Дослідна експлуатація системи показала релевантність пошуку в середньому на рівні 90%.

Цінність системи полягає у тому, що вона долає останню перепону на шляху глобального обміну інформацією, оскільки текстові дані, що генеруються із контенту дуже легко індексуються будь-якою пошуковою системою світу. Тим самим можна сказати, що технологія автоматичного розпізнавання відіграє важливу роль у зборі мультимедійної інформації за ключовими словами. Система є масштабованою і тому універсальною у розгортанні на нові джерела медійної інформації.

Подальший розвиток системи полягає в забезпеченні лематизації при пошуку, введенні метаданих, таких як ідентифікатор мовця, подальшої інтеграції з підсистемами комп'ютерного зору та у підвищенні надійності розпізнавання в цілому за рахунок збільшення словника, розширення навчальної вибірки та розвитку багатомовного аспекту.

ЛІТЕРАТУРА:

1. <http://voxalead.labs.exalead.com/>
2. http://tech.ebu.ch/docs/events/metadata15/Petr_Vitek_and_Pavel_Ircing_CT_UWB.pdf
3. R. Yan, A. Hauptmann. A review of text and image retrieval approaches for broadcast news video. *Information Retrieval*, 2007, Volume 10, Issue 4-5, pp 445-484.
4. Сажок Н.Н., Робейко В.В., Федорин Д.Я., Селюх Р.А. Система преобразования телерадиовещания в текст для украинского языка. – УСиМ. №6, 2015, с. 55-73.
5. M. Gales and S. Young. “The Application of Hidden Markov Models in Speech Recognition”. *Foundations and Trends in Signal Processing*, 2007, 1(3), pp. 195- 304.
6. <http://speech.com.ua/>
7. <http://cybermova.com/technology/synchrophone.html>