Сучасні методи покращення робасності діалогових систем типу „людина - комп'ютер”

Богдан Власенко

Університет Магдебурга, Інститут електроніки та інформатики,
факультет обробки сигналів та комунікаційних систем,
кафедра когнітивних систем

bogdan.vlasenko@gmail.com

Abstract

Herin I present overview of the most popular Prosody Recognition methods. They are used for improvement of robustness for automatic human-machine dialog systems. Structure and relation between different Prosody types were present too. Overview and comparison of performance reliability in Verbmobil and Smartkom projects, illustrated too. At the end I introduce you modern tasks in Prosody Recognition field.

Вступ

За умов стрімкого розвитку комунікаційних систем та інформаційних технологій з однієї сторони, та збільшення потреби в доступі до великої кількості різноманітної інформації з іншої, перед інженерами з’явилася проблема розробки нового типу систем, які б дозволяли людині, користуючись властивими їй засобами спілкування (мовлення, жести, міміка й т. ін.) отримувати необхідну інформацію. Ці системи отримали назву діалогових систем типу „людина-комп'ютер”. Зокрема, відомі приклади діалогових систем, що дозволяють отримувати інформацію про розклад руху транспорту, поточний стан банківського рахунку та бронювати квитки на потяг. Також існують системи типу „людина-комп'ютер-людина”, зокрема Verbmobil, що здійснює переклад з німецької мови на англійську шляхом „мовлення-мовлення”.

В даній статті розглянуто основні напрямки розробки мультимодальних систем розпізнавання й розуміння мовленнєвого сигналу. В таких системах мультимодальність досягається шляхом використання декількох підсистем розпізнавання, що обробляють інформацію різних типів, наприклад: розпізнавання лексичної складової мовлення, розпізнавання просодичної (емоційної та розмежувальної просодії) складової мовлення, розпізнавання руху контурів обличчя під час артикуляції, розпізнавання допоміжних жестів.

Передумовою для переходу до мультимодальних систем, стало розширення задачі з розпізнавання мовлення до розуміння мовлення. Адже в практичному використанні більшу цінність представляють системи, що здатні опрацьовувати довільне злите мовлення.

Найбільш відомими підсистемами, що використовуються в мультимодальних системах розпізнавання мовлення є: розпізнавання емоційної та розмежувальної просодії, розпізнавання намірів, розпізнавання мовлення на основі акустичного та лінгвістичного аналізу, семантичні мережі впорядкування тематичних доменів.

1. Діалогові системи „людина-комп'ютер”

В діалогових системах типу „людина-комп'ютер” можна виділити три взаємопов'язані підсистеми: підсистема акустичного аналізу, підсистема лінгвістичного аналізу, підсистема озвучення тексту. В деяких випадках замість підсистем акустичного та лінгвістичного аналізаторів, використовують підсистему розуміння мовленнєвого сигналу та підсистему побудови діалогу.

1.1. Загальна структура

Загальна структурна діалогової системи типу „людина-комп'ютер” зображена на рисунку 1.

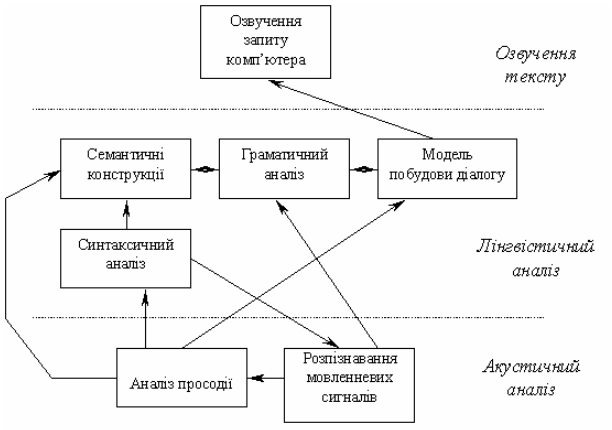

Рис. 1. Загальна структура діалогової системи типу „людина-комп'ютер”.

У найпростішому випадку діалогової системи типу „людина-комп'ютер” діалог відбувається лише шляхом мовленнєвого спілкування. Тобто вхідними даними для такої системи є мовленнєвий сигнал, що містить в собі запит на певну інформацію, а вихідними – мовленнєвий сигнал, що містить в собі або уточнення відносно очікуваної інформації, або саму інформацію.

1.2. Акустичний аналіз

До складу акустичного аналізатора входять дві підсистеми: підсистема аналізу просодії та підсистема розпізнавання мовленнєвого сигналу [1].

В залежності від поставленої задачі підсистема розпізнавання мовленнєвого сигналу може бути орієнтована на великий словник (наприклад Verbmobil [2]) і відносно невеликий, обмежений певним тематичним доменом (наприклад SmartKom [3]). Як правило, діалогові системи орієнтовані на розпізнавання злитого мовлення.

Для покращення стійкості та робасності діалогових систем використовується підсистема аналізу просодії [4]. Більш детально її структура буде описана в розділі 2.

1.3. Лінгвістичний аналіз

Основною передумовою використання лінгвістичного аналізатора є можливість накопичення інформації стосовно порядку слів, синтаксичного аналізу, ключових слів фрази, семантичної структури мови для покращення результатів розпізнавання та розуміння мовлення [2]. Зокрема, відповідно до рисунку 1, на вхід до лінгвістичного аналізатора подаються вихідні дані з акустичного аналізатора. Задачею лінгвістичного аналізатора є вибір найбільш коректної, з точки зору лінгвістичної змістовності, фрази з числа можливих альтернатив, отриманих на виході з акустичного аналізатора. Відповідно до коректно визначеної фрази модель побудови діалогу формує відповідь або допоміжне запитання комп’ютера, яке подається на підсистему озвучення тексту.

Розглянемо детальніше складові підсистеми лінгвістичного аналізу та їх функції.

Синтаксичний аналіз проводить синтаксичний розбір фрази на базі аналізу просодії. В результаті оцінюємо:

a. Тип речення за метою висловлювання: розповідне, питальне, спонукальне.

b. Тип речення за емоційним забарвленням: окличне, неокличне.

c. Розділові знаки в реченні.

Семантичні конструкції містять у собі інформацію про можливі конструкції мови, що характеризують взаємозв'язки між членами речення.

Граматичний аналіз, беручи до уваги апріорну інформацію про порядок слів, формує найбільш коректну послідовність слів з числа можливих альтеративних розміщень.

Модель побудови діалогу, враховуючи результат граматичного аналізатора, формує необхідну відповідь або фразу уточнення.

1.4. Озвучення тексту

Для озвучення текстових повідомлень, які генерує комп'ютер, використовується озвучувач тексту. У вузькому колі прикладних задач, можна користуватися заздалегідь накопиченою базою фонограм відповідей та уточнень. Але в цьому випадку втрачається гнучкість рішення.

2. Використання просодичних характеристик мовленнєвого сигналу

Просодична складова мовленнєвого сигналу відіграє дуже важливу роль під час спілкування. Просодія буває двох типів: розмежувальна та емоційна. Розмежувальна просодія сприяє розбірливості мовлення, характеризує розмічення пауз між словами та реченнями, відповідає за наголоси всередині слова та всередині речення, виявляє мету висловлювання. Емоційна просодія виражає емоційний стан, в якому перебуває співрозмовник, а також допомагає зрозуміти його наміри.

2.1. Знаходження просодичних характеристик

На рисунку 2 зображено приклад знаходження базових характеристик просодичної складової мовленнєвого сигналу. А саме, енергії мовленнєвого сигналу на короткому вікні аналізу та траєкторії основного тону мовленнєвого сигналу.

Загалом існує три групи базових просодичних характеристик: ті, що відносяться до основного тону, ті, що відносяться до енергії на короткому вікні аналізу, та темпоральні [4].

Серед характеристик, орієнтованих на основний тон, можна виділити наступні:

- стандартне відхилення тривалості;
- середнє значення частоти основного тону;
- стандартне відхилення основного тону;
- відносний максимум основного тону;
- відносний мінімум основного тону;
- позиція максимуму основного тону;
- позиція мінімуму основного тону;
- максимум відхилення основного тону;
- середня відстань між точками перегину;

- середнє значення похідної від основного тону.

Серед характеристик, орієнтованих на енергію, можна виділити наступні:

- відносний максимум похідної від енергії;
- позиція максимуму похідної від енергії;

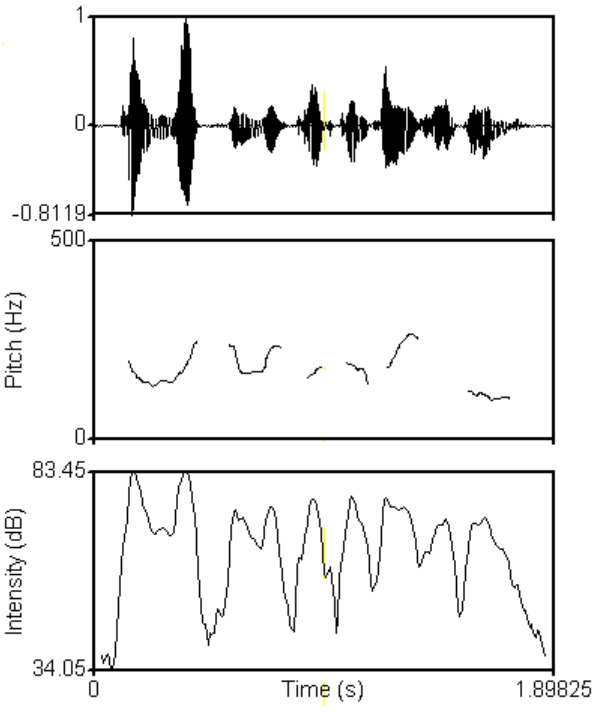

Рис. 2. Базові характеристики просодії

- середнє значення похідної від енергії;
- стандартне відхилення похідної від енергії;
- максимум другої похідної від енергії;
- середня відстань між точками перегину;
- стандартне відхилення відстані між точками перегину.

Серед темпоральних характеристик можна виділити такі:

- середня тривалість вокалізованих звуків;
- середня тривалість пауз між словами;
- середнє відхилення тривалості пауз між словами;
- середнє значення відношення тривалості слів до кількості складів.

2.2. Можливі моделі просодичних аналізаторів

Серед існуючих методів моделювання просодії можна виділити наступні: метод опорних векторів, метод найбільшої правдоподібності, що використовує Гаусівські суміші, нейромережеві методи, а також метод на основі прихованих Марківських моделей.

Існують два типи аналізаторів просодії: з відомим змістом фрази (коли на виході з підсистеми “розпізнавання мовленнєвих сигналів” отримуємо набір можливих послідовностей слів зі словника) а також з невідомим змістом фрази. Зосередимо увагу на першому типі.

2.3. Взаємодія з іншими підсистемами

Розглянемо наступний рисунок:

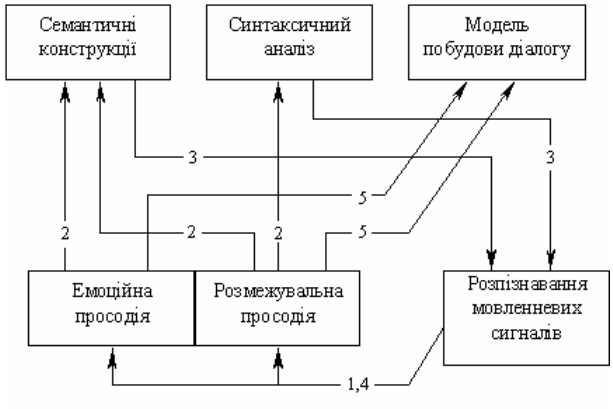

Рис. 3. Взаємозв'язки з підсистемою “аналіз просодії”

Кожна дуга, що характеризує зв'язок між системами, помічена порядковим номером від 1 до 5. Номер відповідає порядку запуску підсистеми.

На рисунку 3 можна побачити замкнений цикл 1-2-3. Розглянемо його детальніше, зосередивши увагу на функціях, що виконуються на кожному з етапів:

1-ий етап: Побудова можливих гіпотез стосовно послідовності слів та їх розміщення в часі.

2-ий етап: Передача можливих просодичних оцінок до набору семантичних конструкцій. Передача гіпотетичного графу слів на синтаксичний аналізатор.

3-ій етап: Отримання узгоджених між собою семантичних і синтаксичних конструкцій.

Після замкненого циклу знову запускається підсистема “Розпізнавання мовленнєвих сигналів”, яка також опрацьовує отриману раніше семантичну й синтаксичну інформацію. Таким чином, на 4-ому етапі передаємо кінцевий варіант послідовності слів та їх розміщення в часі для остаточного аналізу на емоційну та розмежувальні просодії.

На 5-ому етапі до кінцевого варіанту послідовності слів додаємо остаточну оцінку емоційного й розмежувального просодичних аналізаторів та передаємо їх на модель побудови діалогу.

3. Приклади існуючих діалогових систем

Серед найбільш відомих прикладів діалогових систем можна виділити Verbmobil і SmartKom.

Verbmobil – діалогова система типу „людина-комп'тер-людина". Вона виконує переклад довільного мовлення, що промовляє певний користувач системи, та озвучує його іншому. Здійснює переклад з німецької мови на англійську й навпаки. Система розуміння мовленнєвого сигналу реалізована завдяки використанню просодичних характеристик для лінгвістичного аналізу [2].

В Verbmobil просодичний аналіз працює з наступними класами:

- Просодично відображений наголос фрази (основний, допоміжний, емоційний, порівняльний наголоси);
- Просодично відображені межі (повна інтонаційна межа, допоміжні межі, межі слів);
- Просодично відображена мета висловлювання (розповідне, питальне спонукальне).

SmartKom – мультимодальна діалогова система, що комбінує в собі інтерфейси введення та виведення на базі мовлення, міміки та жестів. Основним досягненням даного проекту є накопичена база даних, що містить відео- та аудіозаписи взаємодії споживача та системи, керованої людиною. При чому для отримання більш реалістичних даних, споживач повинен бути впевнений, що він працює з інтелектуальною системою. Подібний метод накопичення даних має назву Wizard-of-Oz [3].

Так існує три версії SmartKom, для різного прикладного застосування:

- SmartKom-Public: діалогова система, що допомагає знайти інформацію стосовно готелів, ресторанів, громадських подій;
- SmartKom-Mobile: діалогова система, що орієнтована на інтеграцію в кишенькові комп'ютери і дозволяє здійснювати доступ до Інтернету, а також спрощує керування GPS системою;
- SmartKom-Home/Office: діалогова система, що дозволяє користувачеві отримувати інформацію стосовно розладу телепередач, керувати побутовою електротехнікою, здійснювати доступ до електронної поштової скриньки та телефону.

Після накопичення база даних була розмічена групою експертів. На відміну від Verbmobil, експерти також оцінювали емоційний стан кожної фрази. Так було виділено 7 класів емоцій:

1) нейтральний;
2) веселий;
3) злий;
4) потребуючий допомоги;
5) розмірковуючий;
6) здивований;
7) невизначений.

Для кожної емоції з 2 по 7 її оцінна інтенсивність:

1) сильно виражена;
2) слабо виражена.

4. Сучасні задачі розпізнавання просодії

Проблема розпізнавання просодичної складової залишається актуальною й у наш час. Просодія грає дуже важливу роль під час розробки систем розпізнавання та розуміння мови, що використовують лінгвістичні та паралінгвістичні дані. Також просодія грає дуже важливу роль у системах озвучування тексту.

Висновки

Практичні дослідження показали, що просодія значно покращує надійність роботи діалогових систем. Вона грає важливу роль під час лінгвістичного аналізу, що застосовується в системах розуміння мовлення.

Література

[1] Batliner, Anton ; Nöth, Elmar: Prosody and Automatic Speech Recognition - Why not yet a Success Story and where to go from here . *In: Proceedings of the 2nd Plenary Meeting and Symposium on Prosody and Speech Processing,* Tokyo 2003, pp. 357-364.

[2] Elmar Nöth, Anton Baltiner, Andreas Kiessling, Ralf Kompe: Verbmobil: The Use of Prosody in the Linguistic Components of a speech Understanding System. *IEEE 2000*, pp 519-533.

[3] Wolfgang Wahlster, A. Blocher, N. Reithinger SmartKom: Multimodal Communication with a Life-Like Character *In: Proceedings of Eurospeech 2001, 7th European Conference on Speech Communication and Technology*, Aalborg, Denmark, September 2001, Vol. 3, pp. 1547 – 1550

[4] Mixdorff, H. Speech Technology, ToBI and Making Sense of Prosody. *Invited talk at Speech Prosody 2002*, pp. 31-38, Aix, France.