

УДК 004.934

Имамвердиев Я.Н.¹, Сухостат Л.В.²

Институт Информационных Технологий НАНА, Баку, Азербайджан

¹yadigar@lan.ab.az, ²lsuhostat@hotmail.com**РАСПОЗНАВАНИЕ ДИАЛЕКТОВ НА ОСНОВЕ АКУСТИЧЕСКОЙ МОДЕЛИ**

Определение регионального диалекта является важным в области речевых технологий. Оно находит широкое применение в телефонных справочных системах, при адаптации выхода синтезированной речи в диалоговых системах, в криминалистической экспертизе при профилировании диктора в судебных или военных ситуациях и т.д. В статье описываются различные подходы, которые позволяют использовать несколько источников информации от акустического сигнала для построения системы распознавания диалектов. В частности, рассматриваются акустический, просодический, фонетический и фонотактический подходы.

Ключевые слова: распознавание диалектов, акустическая модель, речевой сигнал, UBM-GMM-модель.

Введение

Одной из основных проблем современных исследований по речеобразованию и речевым технологиям являются понимание и моделирование индивидуальных вариаций в разговорном языке. Физические лица имеют свои собственные стили разговора в зависимости от многих факторов, таких, как диалект и акцент, а также социально-экономическое положение. Эти различия, как правило, демонстрируют трудности моделирования крупномасштабных дикторнезависимых систем, предназначенных для обработки информации на определенном языке. Люди на протяжении многих лет учатся в определенной мере идентифицировать и интерпретировать большинство из этих аспектов речи.

За последние несколько десятилетий значительный прогресс был достигнут в автоматической идентификации языка говорящего по данному речевому образцу. Распознавание акцента и диалекта совсем недавно начало привлекать внимание ученых в области речевых технологий [1–4]. Задача идентификации диалекта – это распознавание регионального диалекта говорящего в пределах заданного языка на основании речевого сигнала.

Современные системы распознавания речи сосредоточены на изменениях речи в зависимости от диалекта или акцента определенного языка. Диалект рассматриваемого языка является моделью произношения или словаря языка, используемого в сообществе носителей языка, принадлежащих к той же географической области. Исследования показывают, что применение распознавания диалекта говорящего перед автоматическим распознаванием речи помогает в улучшении производительности систем путем адаптации к соответствующим моделям (акустической и языковой) [5]. Это связано с тем, что дикторы, говорящие на разных диалектах, произносят некоторые слова по-разному, последовательно изменяя определенные фонемы и даже морфемы.

Имеющиеся ресурсы [6, 7] показывают, что исследования по автоматической идентификации диалекта проводились для различных языков западных и восточных стран.

Диалекты, акценты и стили

Существуют три различные языковые вариации, которые появляются в любом языке. Две из этих категорий определяются региональными вариациями как в произношении (акцентах), выборе слов, грамматических (диалекты) и социологических вариациях, так и в разных стилях разговора в зависимости от возраста, ситуации и пола. Зная все эти переменные, можно создать представление о социальных, исторических и географических

факторах используемого языка [8].

Диалект – это разновидность речи в пределах указанного языка. Различия между диалектами в основном возникают за счет региональных и социальных факторов, и эти различия отличаются с точки зрения произношения, словарного запаса и грамматики [7].

Акценты определяются как разновидности произношения определенного языка и относятся к звукам, существующим в языке [1]. В отличие от диалектов, акценты охватывают лишь небольшую группу вариаций, которые могут возникнуть в определенном языке.

Стили обычно относятся к настроению говорящего и ситуации, в которую помещен диктор. Этот фактор отличается от диалекта и акцента, так что диалект и акцент – это то, как люди говорят на определенном языке в обществе, в то время как стили относятся к разговорному языку того же человека в различных ситуациях.

О диалектологии

Изучением звуков, слов и грамматических форм, варьирующихся в языке, занимается диалектология. Этот термин обычно используется для описания изучения как акцентов (различия в произношении звуков, использующихся в языке), так и диалектов (отличающихся грамматическими структурами и словами). В целом, диалектология сосредоточена на географическом распределении различных акцентов и диалектов, хотя она начала исследовать также социальные факторы (такие как возраст, пол и положение в обществе).

Традиционные исследования в диалектологии, как правило, направлены на создание диалектных карт, в результате чего воображаемые линии были нарисованы на карте, чтобы указать различные области диалектов. Это привело к началу изучения социальных, а также региональных различий языков. Первым в мире в Германии в 1821–1837 гг. Johann Andreas Schmeller исследовал диалект Баварии, который включал лингвистический атлас [9]. Лингвистический атлас США (1930-е годы) был одним из первых исследований диалектов с учетом социальных факторов. В 1950 году в Университете Лидса (*University of Leeds*) был проведен обзор английских диалектов в Англии и восточных районах Уэльса.

Начало исследований русских диалектов относится к XIX в. «Толковый словарь живого великорусского языка» (1863–1866) В.И. Даля сыграл огромную роль в их изучении. Первая диалектологическая карта русского языка была составлена в 1915 г. [10, 11].

В Азербайджане на протяжении многих лет ведется большая работа по изучению азербайджанских диалектов [12–16].

Распознавание диалектов

Различные методы были предложены для решения проблем в задаче распознавания диалектов. Большинство из них имеют сходство с соответствующими методами, используемыми в задаче распознавания языка. Также много исследований было проведено в области автоматического распознавания регионального акцента [2, 3, 17–20].

Распознавание диалектов может быть выполнено на различных уровнях, таких как акустический (например, спектральная информация), просодический, фонотактический (например, языковые модели) и лексический [21]. Что касается акустического уровня, то спектральная информация речевого сигнала извлекается с помощью методов параметризации речи и далее применяются алгоритмы классификации, такие как модель гауссовых смесей [6], метод опорных векторов [22] и нейронные сети [23]. В работе [24] представлен акустический подход по распознаванию четырех языков Индии: индийского, английского, хинди, ассамского, бенгальского. В качестве параметров используются мел-частотные кепстральные коэффициенты (*Mel-frequency Cepstral Coefficients, MFCC*), распознавание осуществляется с использованием смешанной гауссовой модели (*Gaussian Mixture Model, GMM*). На просодическом уровне рассматриваются продолжительность фонетических единиц [6, 21, 24, 25] и ритм [26].

В [27, 28] были использованы просодические параметры. При этом речевой сигнал разделяется на невокализованные/вокализованные сегменты, где определяются частота основного тона, длительность невокализованного/вокализованного участка и др. Подход основан на применении N-граммных моделей. Суть метода состоит в раздельном моделировании фразовых и локальных акцентов.

В работе [29] проведено исследование системы распознавания языка на основе фонотактического подхода. Фонотактический подход в распознавании акцента и диалекта базируется на гипотезе, что диалекты или акценты отличаются распределением последовательностей фонем. Иными словами, тексты в одном и том же языке могут быть распознаны по этим распределениям символов.

Предлагаемый подход

Для распознавания диалектов языка в данной работе предлагается GMM-UBM-модель. Универсальная фоновая модель (*Universal Background Model, UBM*) – это GMM-представление характеристик всех диалектов, создаваемое с использованием части обучающих данных.

GMM – это модель плотности распределения, которая представляет собой модель диалекта или языка. Она определяет различные гауссовы распределения, которые имеют свое математическое ожидание, дисперсию и вес в GMM. Предположим, что M – это число малых гауссовских распределений модели. Следующее уравнение пытается моделировать плотность распределения вероятности N -мерного случайного вектора x , добавляя взвешенное сочетание взвешенных многомерных гауссовых плотностей:

$$p(x | \lambda_d) = \sum_{i=1}^M p_i b_i(x), \quad (1)$$

где

$$b_i(x) = \frac{1}{(2\pi)^{N/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) \right\},$$

μ_i – вектор математических ожиданий и Σ_i – ковариационная матрица.

Эти параметры представлены следующей записью:

$$\lambda = \{ p_i, \mu_i, \Sigma_i \}, \quad i = 1, \dots, M, \quad (2)$$

где p_i – смешанные веса, удовлетворяющие условию $\sum_{i=1}^M p_i = 1$.

У каждой группы диалектов своя модель λ_d .

Параметры модели каждого диалекта оцениваются с помощью алгоритма Expectation-Maximization (EM) по последовательности векторов признаков $X = \{x_1, x_2, \dots, x_T\}$, которые извлекаются из набора речевых высказываний данного диалекта d .

При этом для каждой группы диалектов находятся следующие значения:

$$p_i = \frac{1}{T} \sum_{t=1}^T pr(i | x_t, \lambda), \quad (3)$$

$$\mu_i = \frac{\sum_{t=1}^T pr(i | x_t, \lambda) x_t}{\sum_{t=1}^T pr(i | x_t, \lambda)}, \quad (4)$$

$$\Sigma_i = \frac{\sum_{t=1}^T pr(i|x_t, \lambda) x_t^2}{\sum_{t=1}^T pr(i|x_t, \lambda)} - \mu_i^2, \quad (5)$$

где апостериорная вероятность для компонента i имеет вид

$$pr(i|x_t, \lambda) = \frac{p_i b_i(x)}{\sum_{k=1}^M p_k b_k(x)}. \quad (6)$$

GMM-параметры определяются путем оценки обучения максимального правдоподобия:

$$\lambda_d = \arg \max_{\lambda_i} \left\{ \prod_{t=1}^T p(x_t | \lambda_i) \right\}. \quad (7)$$

Преимуществом использования UBM в системах идентификации диалектов является значительное уменьшение количества обучающих данных. Вместо отдельного обучения диалектозависимых моделей в подходе используется байесовская адаптация из UBM на основе речевых образцов, обученных на определенном диалекте. В качестве параметров предлагается использовать MFCC-коэффициенты. Общая схема предлагаемой модели распознавания диалектов показана на рис. 1.



Рис. 1. Система идентификации диалектов на основе GMM-UBM

Во время идентификации неизвестное речевое высказывание X классифицируется после вычисления логарифмического правдоподобия, которое определяется по формуле:

$$L_d(X) = \log p(X | \lambda_d) = \frac{1}{T} \sum_{t=1}^T \log p(x_t | \lambda_d). \quad (8)$$

До автоматической идентификации диалекта речевые сигналы сначала предварительно обрабатываются методом фильтрации по нулевой частоте (Zero Frequency Filtering, ZFF) [30] для удаления изменяющегося во времени низкочастотного шума.

Учитывая, что GMM и UBM имеют M -смесей, мы выбрали для тестирования первые N -смесей для D -диалектов. Для системы GMM количество смешанных испытаний составило $N_{mixture} = M \times D$. Для пяти групп диалектов при использовании 512 GMM-смесей число испытаний составляет $N_{mixture} = 512 \times 5 = 2560$.

Заключение

Исследование было направлено на автоматическую идентификацию диалектов языка на основании речевых образцов. Для достижения задачи идентификации диалектов были исследованы акустический, просодический, фонетический, фонотактический и

лексический подходы. Наиболее широкое распространение получили модели на основе GMM. В данной статье была предложена акустическая модель на основе GMM-UBM для обеспечения достоверности распознавания.

Таким образом, можно сделать вывод, что наиболее перспективным направлением автоматического распознавания диалектов языка является разработка подходов, направленная на первоначальное разделение диалектов на группы с последующим определением конкретного диалекта.

Литература

1. Clark J. An Introduction to phonetics and phonology. Oxford: Blackwell Publishing, 2007, 504 p.
2. Biadys F. Automatic dialect and accent recognition and its application to speech recognition: Ph.D. dissertation. Columbia University, 2011, 171 p.
3. Omar M.K., Pelecanos J. A novel approach to detecting non-native speakers and their native language / Proc. of IEEE ICASSP, 2010, pp.4398–4401.
4. Roy P., Das P.K. Language identification of Indian languages based on gaussian mixture models // International Journal of Wisdom Based Computing, 2011, vol.1, no.3, pp.54–59.
5. Liu M.K., Xu B., Huang T.Y., Deng Y.G., Li C.R. Mandarin accent adaptation based on context-independent/context-dependent pronunciation modeling / Proc. of ICASSP, 2000, vol.2, pp.1025–1028.
6. Hazen, T., Zue, V. Segment-based automatic language identification // Journal of the Acoustic Society of America, 1997, vol.101, no.4, pp.2323–2331.
7. Akmajian A., Demers R.A., Farmer A.K, Harnish R.M. Linguistics: an introduction to language and communication. Massachusetts: MIT Press, 2001, 604 p.
8. Nerbonne J. Linguistic variation and computation / Proc. of the 10th conference on European chapter of the Association for Computational Linguistics, 2003, vol.1, pp.1–3.
9. Petyt K.M. The study of dialect: an introduction to dialectology. London: Westview Press, 1980, 235 p.
10. Словарь русских народных говоров, вып. 1–43. М. Л.: Наука, 1965, 306 с.
11. Бромлей С.В., Булатова Л.Н., Захарова К.Ф. Русская диалектология. М.: Просвещение, 1989, 223 с.
12. “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı təsdiq edilməsi” haqqında Azərbaycan Respublikası Prezidentinin Sərəncamı, "Azərbaycan" qəzeti, Bakı, 2013-cü il, 9 aprel.
13. Ширалиев М.Ш. О диалектной основе азербайджанского национального литературного языка. М.: Изд-во АН СССР, 1958, с.78–84.
14. Azərbaycan dilinin dialektoloji lüğəti. Bakı: “Şərq-Qərb”, 2007, 568 səh.
15. Bayramov İ. Qərbi Azərbaycan şivələrinin leksikası. Bakı: “Elm və Təhsil”, 2011, 440 səh.
16. Kazımov Q. Azərbaycan dili dialekt və şivələrinin tədqiqi tarixindən // Filologiya məsələləri, 2014, no.1, s.3–27.
17. Teixeira C., Trancoso I., Serralheiro A. Accent identification / Proc. of INTERSPEECH, 1996, vol.3, pp.1784–1787.
18. Hanani A., Russell M.J., Carey M.J. Human and computer recognition of regional accents and ethnic groups from british english speech // Computer Speech and Language, 2013, vol.27, no.1, pp.59–74.
19. Huang R., Hansen J.H.L., Angkititrakul P. Dialect/accent classification using unrestricted audio // IEEE Trans. on Audio, Speech and Language Processing, 2007, vol.15, no.2, pp.453–464.
20. Mporas I., Ganchev T., Fakotakis N. Phonotactic recognition of greek and cypriot dialects from telephone speech // SETN 2008, Advances in Artificial Intelligence, Lecture Notes in Computer Science. Berlin: Springer-Verlag, 2008, pp.173–181.
21. Tong, R., Ma, B., Li, H., Chng, E.S. Integrating acoustic, prosodic and phonotactic features for spoken language identification / Proc. of IEEE ICASSP, 2006, pp.205–208.

22. Campbell W. M., Singer E., Torres-Carrasquillo P.A., Reynolds D.A. Language recognition with support vector machines / Proc. of Odyssey, 2004, pp.285–288.
23. Braun J., Levkowitz H. Automatic language identification with perceptually guided training and recurrent neural networks / Proc. of the 5th International Conference on Spoken Language Processing (ICSLP), 1998, pp.3201–3205.
24. Ghesquiere P.J., Compennolle D.V. Flemish accent identification based on formant and duration features // Proc. of ICASSP, 2002, pp.749–752.
25. Lin C.Y., Wang H.C. Fusion of phonotactic and prosodic knowledge for language identification / Proc. of the 9th International Conference on Spoken Language Processing (ICSLP), 2006, pp.425–428.
26. Farinas J., Pellegrino F., Rouas J.L., Andre-Obrecht R. Merging segmental and rhythm features for automatic language identification / Proc. of IEEE ICASSP, 2002, pp.753–756.
27. Rouas J.-L., Farinas J., Pellegrino F., Andre-Obrecht R. Modeling prosody for language identification on read and spontaneous speech / IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003, vol.6, pp.1–4.
28. Rouas J.-L. Automatic prosodic variations modeling for language and dialect discrimination // IEEE transactions on audio, speech and language processing, 2007, vol.15, no.6, pp.1904–1911.
29. Torres-Carrasquillo P. A., Singer E., Kohler M. A. Approaches to language identification using gaussian mixture models and shifted delta cepstral features / Proc. of IEEE ICASSP, 2002, pp.757–760.
30. Murty K. S. R., Yegnanarayana B. Epoch extraction from speech signals // IEEE Transactions on Audio, Speech, and Language Processing, 2008, vol.16, no.8, pp.1602–1613.

UOT 004.934

İmamverdiyev Yadigar N.¹, Suxostat Lyudmila V.²

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹yadigar@lan.ab.az, ²lsuhostat@hotmail.com

Akustik model əsasında dialektlərin tanınması

Regional dialektin təyini nitq texnologiyaları sahəsində əhəmiyyətlidir. Dialektlərin tanınması telefon məlumat sistemlərində dialoq sistemlərinin sintezləşdirilmiş nitqinin çıxışının adaptasiyası məsələsində, kriminalistik ekspertiza sahəsində, məhkəmə və ya hərbi vəziyyətlərdə dikturun profilləşdirilməsi məsələsində və s. geniş tətbiq olunur. Məqalədə dialektlərin tanınması sisteminin qurulması üçün akustik signalın bir neçə informasiya mənbəyindən istifadə etməyə imkan verən müxtəlif yanaşmalar təsvir edilir. Xüsusi halda, akustik, prosodik, fonetik və fonotaktik yanaşmalara baxılır.

Açar sözlər: *dialektlərin tanınması, akustik model, nitq signalı, UBM-GMM model.*

Yadigar N. Imamverdiyev¹, Lyudmila V. Sukhostat²

Institute of Information Technology of ANAS, Baku, Azerbaijan

¹yadigar@lan.ab.az, ²lsuhostat@hotmail.com

Dialects recognition based on acoustic model

Regional dialect recognition is important in the field of speech technologies. It is widely used in telephone reference systems, in adapting the output synthesized speech in dialog systems, and also in forensics for profiling speaker in judicial or military situations, etc. The article describes different approaches that allow the usage of multiple information sources from the acoustic signal for the construction of dialects recognition system. In particular, acoustic, prosodic, phonetic, and phonotactic approaches are considered.

Keywords: *dialects recognition, acoustic model, speech signal, UBM-GMM model.*