УДК 553.98:551

ПРИМЕНЕНИЕ МОДЕЛЕЙ ГАУССОВОЙ СМЕСИ В ЛИНГВОМЕТРИЧЕСКИХ ЗАДАЧАХ

Берзинь А. У.

APPLICATION OF GAUSSIAN MIXTURE MODELS IN LINGUOMETRIC TASKS

Bērziņš A. A.

Рассматриваются результаты применения методов распознавания речи, основанных на моделях гауссовой смеси, для задания расстояния между устными идиомами (т. е., любыми устными естественно-языковыми системами, в том числе языками, диалектами, говорами, социолектами и т. п.). В качестве входных данных используются фонограммы спонтанной речи. Эксперименты проводятся на звукозаписях латышских и латгальских говоров, но методы применимы и к любым другим идиомам. В рамках исследования проводились четыре эксперимента – с прямо созданными моделями на полных фонограммах и на уравновешенных по продолжительности фонограммах, а также с моделями с MAP-адаптацией тоже на одних и других вариантах фонограмм. Каждый из экспериментов проводился для векторов средних значений, ковариационных матриц и весовых векторов. Рассчитывались евклидова метрика, метрика L1, расхождение Кульбака-Лейблера и метрика Чебышёва. Оценка результатов проводилась при помощи метода экспертной оценки для дендрограмм (т.е., специфических бинарных графов), полученных путём агломеративной иерархической категоризации результатов, пользуясь метрикой иерархического выбора. Вывод о применимости метода положителен.

Ключевые слова: гауссова смесь, модель, речь, распознавание, спектр, признак, нормальное распределение, язык, идиом, диалект, говор.

The article describes results of application of GMM-based speech recognition methods to define a kind of a distance between idioms. Spontaneous speech recordings of many enough speakers of same idioms are used on the input of the method. The experiments were carried out on recordings of Latvian and Latgalian subdialects, but the method is applicable to any other spoken idioms too. Euclidean, L1 (or city block) and Chebyshov (or Jordan) metrics and Kullback-Leibler divergence were tried out for mean vectors, covariance matrices and weight vectors. We did four cases of experiments: directly created models on full length recordings, models, created using MAP-adaptation on full length recordings, directly created models on recordings with balanced length, and models, created using MAP-adaptation on recordings with balanced length. The results were evaluated by the method of expert evaluation for the output dendrograms (that is, specific binary graphs) of agglomerative hierarchical clustering of the results of the experiments, using the metric of hierarchical choices. The main conclusions are: 1) the MAP-adaptation is making things much worse; 2) the balancing is making things a little bit worse; 3) the method is working and GMMs could be used for calculations of a distance between idioms; 4) usage of Euclidean distance on mean vectors and covariance matrices is recommended, since it returns the best results and is balancing-stable.

Keywords: Gaussian mixture, model, speech, recognition, spectrum, feature, normal distribution, language, idiom, dialect, speech.

Введение. Гауссова смесь – это совокупность, точнее – взвешенная сумма (со скалярными коэффициентами веса), конечного числа распределений Гаусса, называемого также нормальным распределением. Такую модель описывают три величины – вектор математического ожидания, ковариационная матрица и вектор коэффициентов веса.[1] Поскольку сумма неза-

[1] Модель гауссовых смесей представляет собой взвешенную сумму M компонент и может быть записана выражением

$$p(\overline{x} \mid \lambda) = \sum_{i=1}^{M} p_i b_i(\overline{x}), \quad (1)$$

где $\overline{x}$ – это D-мерный вектор случайных величин; $b_i(\overline{x})$, i = 1..M – функции плотности распределения составляющих модели и p_i, i = 1..M – веса компонентов модели. Каждый компонент является D-мерной гауссовой функцией распределения вида

$$b_i(\overline{x}) = \frac{1}{(2\pi)^{D/2} \left|\Sigma_i\right|^{1/2}} \exp\left\{-\frac{1}{2}(\overline{x} - \overline{\mu}_i)^T \sum\nolimits_i^{-1} (\overline{x} - \overline{\mu}_i)\right\} \quad (2)$$

где $\overline{\mu}_i$ — вектор математического ожидания и Σ_i — ковариационная матрица. Веса смеси удовлетворяют выражению $\sum_{i=1}^{M} p_i = 1$.

Полностью модель гауссовой смеси определяется векторами математического ожидания, ковариационными матрицами и весами смесей для каждого компонента модели. Эти параметры все вместе записываются в виде

$$\lambda = \{p_i, \overline{\mu}_i, \Sigma_i\} \qquad i=1..M \quad (3)$$

В задаче распознавания диктора каждый диктор представляется моделью гауссовых смесей и ставится в соответствие со своей моделью λ. Модель гауссовой смеси может иметь несколько различных форм в зависимости от вида ковариационной матрицы. Модель может иметь одну ковариационную матрицу для каждого компонента модели, как определено в (3), одну ковариационную матрицу для всех гауссовых компонент в модели или одну ковариационную матрицу, используемую всеми дикторами во всех моделях. Ковариационная матрица также может быть полной или диагональной. [3]

висимых нормальных распределений является нормальным распределением, то и гауссова смесь является таковым[2].

Модели гауссовых смесей (МГС) широко используются в задачах классификации данных (например, в экономике, демографии, экологии и других областях), если есть основания полагать, что каждый из этих классов соответствует нормальному распределению. Следовательно, МГС также используют в задачах распознавания разного рода (изображений, звука и обьектов других видов), потому что они, по сути, есть выяснение наиболее вероятной принадлежности объекта к какому-то из указанных классов.

Так как нас интересует моделирование речи, то мы работали с распределениями спектральных информационных признаков речи. Векторы признаков формируются из коэффициентов косинусного преобразования Фурье, или т.н. *MFCC* (с английского *Mel Frequency Cepstral Coefficients*)[3], используя в каждом кадре по 14 коэффициентов.

Модели формируются с помощью алгоритма максимизации ожидаемой стоимости или *EM*[4], но двумя способами – как полностью независимые, отдельно обученные модели, каждая из которых формируется только на речевых данных одного соответствующего говора, а также как модели, произведённые путём MAP-адаптации[5] из общей, созданной на частичных данных всех говоров, универсальной фоновой модели или UBM[6].

Эксперименты проводились, пользуясь программным пакетом *MatLab*. Скрипты мы писали сами, на основе информации, предоставленной аспирантом Голосовой лаборатории Инженерного факультета Мексиканского национального автономного университета Хосе Бенито Транголом, о его экспериментах в области распознавания речи и используемых им библиотеках, функциях и параметрах.

Данные. В нашем распоряжении были собранные (записанные) нами звукозаписи спонтанной речи

---

[2] Teorēma. Ja gadījuma lielums X ir neatkarīgu normālu gadījuma lielumu summa, tad arī tam ir normālais sadalījuma likums, pie kam tā matemātiskā cerība ir vienāda ar saskaitāmo matemātisko cerību summu, bet dispersija – ar saskaitāmo dispersiju summu. [8]

[3] Первая нам известная публикация, в которой о них было заявлено и они применялись:

The parametric representations evaluated in this study may be divided into two groups: those based on the Fourier spectrum and those based on the linear prediction spectrum. The first group comprises the mel-frequence cepstrum coefficients (MFCC) ..

For the MFCC computation, 20 triangular bandpass filters were simulated ... The MFCC were computed as

$$MFCC_i = \sum_{k=1}^{20} X_k \cos\left[i\left(k - \frac{1}{2} \right) \frac{\pi}{20} \right] \quad \text{i=1,2,...,M,}$$

where M is the number of cepstrum coefficients, and X_k, k=1,2,...,20, represents the log-energy output of the kth filter. [7]

[4] We define the EM (Expectation-Maximization) algorithm for Gaussian mixtures as follows. The algorithm is an iterative algorithm that starts from some initial estimate of Θ (e.g., random), and then proceeds to iteratively update Θ until convergence is detected. Each iteration consists of an E-step and an M-step.

E-Step: Denote the current parameter values as Θ. Compute w_{ik} (using the equation above for membership weights) for all data points $\underline{x}_i$, $1 \le i \le N$ and all mixture components $1 \le k \le K$. Note that for each data point $\underline{x}_i$ the membership weights are defined such that $\sum_{k=1}^{K} w_{ik} = 1$. This yields an N × K matrix of membership weights, where each of the rows sum to 1.

M-Step: Now use the membership weights and the data to calculate new parameter values. $N_k = \sum_{i=1}^{N} w_{ik}$, i.e., the sum of the membership weights for the kth component—this is the effective number of data points assigned to component k.

...

After we have computed all of the new parameters, the M-step is complete and we can now go back and recompute the membership weights in the E-step, then recompute the parameters again in the E-step, and continue updating the parameters in this manner. Each pair of E and M steps is considered to be one iteration. [14]

[5] In the GMM-UBM system, we derive the hypothesized speaker model by adapting the parameters of the UBM using the speaker's training speech and a form of Bayesian adaptation. (This is also known as Bayesian learning or maximum a posteriori (MAP) estimation. We use the term Bayesian adaptation since, as applied to the speaker-independent UBM to estimate the speaker-dependent model, the operation closely resembles speaker adaptation used in speech recognition applications.) Unlike the standard approach of maximum likelihood training of a model for the speaker independently of the UBM, the basic idea in the adaptation approach is to derive the speaker's model by updating the well-trained parameters in the UBM via adaptation. This provides a tighter coupling between the speaker's model and UBM which not only produces better performance than decoupled models, but, as discussed later in this section, also allows for a fast-scoring technique. Like the EM algorithm, the adaption is a two step estimation process. The first step is identical to the expectation step of the EM algorithm, where estimates of the sufficient statistics 7 of the speaker's training data are computed for each mixture in the UBM. Unlike the second step of the EM algorithm, for adaptation these new sufficient statistic estimates are then combined with the old sufficient statistics from the UBM mixture parameters using a data-dependent mixing coefficient. The data-dependent mixing coefficient is designed so that mixtures with high counts of data from the speaker rely more on the new sufficient statistics for final parameter estimation and mixtures with low counts of data from the speaker rely more on the old sufficient statistics for final parameter estimation. [13]

Между прочим, адаптировать можно разные параметры, при чём не только МГС, но и скрытых моделей Маркова: Model adaptation can also be accomplished using a maximum a posteriori (MAP) approach. This adaptation process is sometimes referred to as Bayesian adaptation. MAP adaptation involves the use of prior knowledge about the model parameter distribution. Hence, if we know what the parameters of the model are likely to be (before observing any adaptation data) using the prior knowledge, we might well be able to make good use of the limited adaptation data, to obtain a decent MAP estimate. This type of prior is often termed an informative prior. Note that if the prior distribution indicates no preference as to what the model parameters are likely to be (a non-informative prior), then the MAP estimate obtained will be identical to that obtained using a maximum likelihood approach. For MAP adaptation purposes, the informative priors that are generally used are the speaker independent model parameters. [15]

[6] A Universal Background Model (UBM) is a model used in a biometric verification system to represent general, personindependent feature characteristics to be compared against a model of person-specific feature characteristics when making an accept or reject decision. For example, in a speaker verification system, the UBM is a speaker-independent Gaussian Mixture Model (GMM) trained with speech samples from a large set of speakers to represent general speech characteristics. Using a speaker-specific GMM trained with speech samples from a particular enrolled speaker, a likelihood-ratio test for an unknown speech sample can be formed between the match score of the speaker-specific model and the UBM. The UBM may also be used when training the speaker-specific model by acting as a the prior model in MAP parameter estimation. [13]

пяти идиомов (латвийских говоров) – один из Курляндии: Дундажской волости, и четыре из Латгалии: Аулеи, Бальтинова, Вилека и Рудзатов. Курляндия исторически была под немецким игом, поэтому местные говоры подверглись влиянию (нижне)немецкого языка, а северокурляндские говоры, в том числе и дундажский, содержут большой субстрат ливонского языка (принадлежащего к прибалтийско-финской подгруппе финно-угорских языков). Латгалия, в свою очередь, была под поляками, поэтому в латгальских говорах присутствует влияние польского языка, также – в силу близкого соседства и наличия белорусских и старообрядческих деревень – белорусского и русского. Бальтиновский и вилекский являются говорами северолатгальскими, которые от западнолатгальского рудзатского и южнолатгальского аулейского отличаются существенно – и морфологически, и лексически.

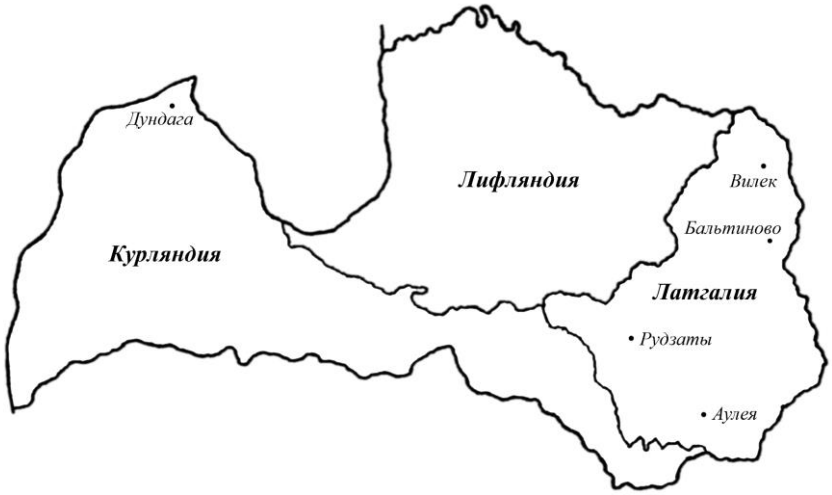

Рис. 1. Расположение записанных говоров на карте Латвии

Все звукозаписи собирались согласно заданным нами принципам сбора информации для автоматизированного анализа фонограмм [4], т.е., все записи были однородными, записанными однотипной аппаратурой (использовался динамический микрофон одностороннего направления, фиксированый на голове информанта), в условиях уменьшенного влияния внешних шумов. Все записи были мануально вычищены, удалению подверглись все посторонние звуки и голоса, оставив только прямую речь информанта. Качество записи – 44,1 кГц / 16 битов.

Таблица 1

Характеристика набора фонограмм, используемого в эксперименте

Говор	Минут	Информантов	Мужчин	Женщин
Аулея	95	14	8	6
Бальтиново	140	23	9	14
Дундага	161	17	4	13
Рудзаты	246	28	11	17
Вилек	238	30	11	19

Всех информантов просили рассказывать о быте, родителях, бабушках, дедушках, братьях, сёстрах, детях, других членах семьи, учёбе, работе, хозяйстве, службе в армии, свадьбах, праздниках, соседях и т. п. Т.е., в ходе сбора данных на традиционность и гомогенность лексики обращалось пристальное внимание.

Поэтому, учитывая однородность нашего многоговорного корпуса и в техническом, и в содержательном смысле, мы даже можем не постесняться его считать сопоставимым[7]. Доселе этот термин применялся только к текстовым корпусам, но мы считаем, что его можно применять и к речевым, и при таком применении наш корпус соответствует смыслу сопоставимости.

Эксперимент № 1: прямое создание моделей на полных фонограммах. Прежде всего, мы провели эксперименты, создавая модели гауссовой смеси непосредственно, без адаптации, то есть каждая модель создавалась только на фонограммах соответствующего говора и полностью независимо от моделей других говоров.

В этом и последующих экспериментах мы рассчитывали следующие метрики: евклидову, L1 (или

[7] A comparable corpus is one which selects similar texts in more than one language or variety. There is as yet no agreement on the nature of the similarity, because there are very few examples of comparable corpora. ... The possibilities of a comparable corpus are to compare different languages or varieties in similar circumstances of communication, but avoiding the inevitable distortion introduced by the translations of a parallel corpus. [10]

A comparable corpus is a pair of corpora in two different languages, which come from the same domain. [6]

городского квартала) и Чебышёва, а также расхождение Кульбака-Лейблера, для всех трёх составляющих моделей – векторов средних значений, ковариационных матриц и весовых векторов. В нашем распоряжении имеются таблицы всех расстояний, но такой объём был бы слишком большим для статьи, посему мы приведём лишь некоторые примеры, а с остальными можно ознакомиться в нашей диссертации [5].

Таблица 2

Пример значений евклидовой метрики для векторов средних значений моделей гауссовой смеси, созданных на полных фонограммах без адаптации

Аулея	0,0000	111,1847	125,0873	114,4163	113,9320
Бальтиново	111,1847	0,0000	123,1723	113,5741	93,4037
Дундага	125,0873	123,1723	0,0000	116,1056	115,0827
Рудзаты	**114,4163**	**113,5741**	116,1055	0,0000	**103,6138**
Вилек	113,9320	93,4036	115,0826	103,6138	0,0000
	Аулея	Бальтиново	Дундага	Рудзаты	Вилек

При неформальной оценке соответствия результатов для **векторов средних значений** нашему интуитивному пониманию близости говоров, евклидова метрика ведёт себя почти хорошо (с одним небольшим несоответствием), L1 – также как евклидова, Чебышёва – очень плохо, а расхождение Кульбака-Лейблера – полностью бессмысленно.

На **ковариационных матрицах** евклидова метрика ведёт себя также, как на векторах средних значений, L1 – немного хуже, расхождение Кульбака-Лейблера – в одном направление содержит небольшие, но явные несоответствия, от которых не удаётся избавится путём симметризирования. Интересно, что и метрика Чебышёва в данном случае оказывается хорошей.

Евклидова и L1 метрики, и расхождение Кульбака-Лейблера (в обоих направлениях) на **весовых векторах** производят более-менее приемлемое впечатление, но всё-таки содержат и явные несоответствия похожего характера. Чебышёва же метрика содержит меньше несоответствий, но зато они выражены ярче.

Эксперимент № 2: модели с адаптацией на полных фонограммах. Во втором эксперименте мы решили провести такие же расчеты, как и в первом, но с помощью моделей, созданных при помощи MAP (*maximum a posteriori*) адаптации. Для этого мы создали общую фоновую модель, которая была построена на 30% записей всех информантов всех говоров. Потом, пользуясь оставшимися 70% данных соответствующего говора, эта фоновая модель адаптировалась и создавались модели всех говоров. Т.е., в процессе подготовки данных все речевые файлы были разделены на две части – 30% начальную и 70% конечную.

Теоретически адаптация может проводиться не только по всем трём, но и по двум или одному параметру МГС, чаще всего – по средним значениям. Однако в таком случае результаты неинтересны – расстояния, основанные на различиях средних значений, совпадают с расстояниями моделей, которые адаптированы по всем трём параметрам, а расстояния, основанные на ковариационных или весовых различиях, равны нулю. Поэтому мы решили экспериментировать с моделями, которые адаптируются по всем трём параметрам.

Интересно, что евклидово расстояние для **векторов средних значений** моделей, созданных с адаптацией, хуже, чем у моделей без адаптации. У L1 те же проблемы, что и у евклидовой, Чебышёва – неплоха, с одним небольшим несоответствием, а расхождение Кульбака-Лейблера – полностью бессмысленно.

На **ковариационных матрицах** евклидовой метрике присущи те же проблемы, что и на векторах средних значений, но в более выраженной форме, а L1 – даже на одно несоответствие больше. Расхождение Кульбака-Лейблера в данном случае осмысленно, но с теми же проблемами, что и евклидово. У метрики Чебышёва тоже такие же проблемы, но проявляются они на других комплектах говоров.

Перейдём к **весовым векторам**. Евклидова метрика показывает неожиданно хорошие результаты, которые полностью соответствуют интуитивному представлению. Интересно, что, в отличие от 1-го эксперимента, евклидова метрика на весовых векторах показала лучшие результаты, чем на ковариационных матрицах. При этом результаты более похожи на результаты на ковариационных матрицах без адаптации. Логического объяснение этому пока не находим, поскольку в процессе настройки в весовых векторах не аккумулируются данные ковариационных матриц данных адаптации, которые могут подобным образом повлиять[8]. L1 ведёт себя похоже, но всё-же похуже.

---

8 Lastly, these new sufficient statistics from the training data are used to update the prior sufficient statistics for mixture i to create the adapted parameters for mixture i with the equations:

$\hat{w}_i = [\alpha_i^w n_i / T + (1 - \alpha_i^w) w_i]\gamma$ adapted mixture weight,

$\hat{\mu}_i = \alpha_i^m E_i(x) + (1 - \alpha_i^m)\mu_i$ adapted mixture mean,

$\hat{\sigma}_i^2 = \alpha_i^\upsilon E_i(x^2) + (1 - \alpha_i^\upsilon)(\sigma^2{}_i + \mu^2{}_i) - \hat{\mu}_i^2$ adapted mixture variance.

The adaptation coefficients controlling the balance between old and new estimates are $\{\alpha_i^w, \alpha_i^m, \alpha_i^\upsilon\}$ for the weights, means and variances, respectively. The scale factor, γ, is computed over all adapted mixture weights to ensure they sum to unity. Note that the sufficient statistics, not the derived parameters, such as the variance, are being adapted.

For each mixture and each parameter, a data-dependent adaptation coefficient α_i^ρ, $\rho \in \{w, m, \upsilon\}$, is used in the above equations. This is defined as

$$\alpha_i^\rho = \frac{n_i}{n_i + r^\rho},$$

where r_ρ is a fixed "relevance" factor for parameter ρ. It is common in speaker recognition applications to use one adaptation coefficient for all parameters $\alpha_i^w = \alpha_i^m = \alpha_i^\upsilon = n_i/(n_i + r)$ and further to only adapt certain GMM parameters, such as only the mean vectors. [11]

Таблица 3

Пример значений симметризованного расхождения Кульбака-Лейблера для весовых векторов моделей гауссовой смеси, созданных на полных фонограммах с MAP-адаптацией

Аулея	0,000000	0,018859	0,023787	0,019783	0,018035
Бальтиново	0,018859	0,000000	0,021157	0,019594	0,014987
Дундага	0,023787	0,021157	0,000000	0,021521	0,020433
Рудзаты	0,019783	0,019594	0,021521	0,000000	0,018447
Вилек	0,018035	0,014987	0,020433	0,018447	0,000000
	Аулея	Бальтиново	Дундага	Рудзаты	Вилек

Расхождение Кульбака-Лейблера имеет лишь одно несоответствие, при том только в одном направлении, и оно гасится путём симметризации. Метрика Чебышёва ведёт себя иначе, с небольшими несоответствиями.

Эксперимент № 3: прямые модели на уравновешенных фонограммах. Количество спонтанного речевого материала никогда не будет одинаковым – оно зависит от обстоятельств, успеха и др. факторов, играющих роль во время сбора. Из 1-й таблицы видно, что и количество информантов, и продолжительность записей отличается от говора к говору. Это нормально, и в этом нет ничего плохого, однако количественная разница может повлиять на результаты экспериментов – модели, сделанные на существенно большем количестве речи, могут отличаться от других лишь потому, что количество речи было больше, в следствии чего различия, отображающие лингвистические свойства, могут нивелироваться и приводить к искажению результатов.

Поэтому мы решили провести эксперименты, в которых количество речи изначально уравновешивается, т.е., для говоров с большей общей продолжительностью собранной речи из фонограммы каждого информанта берётся только её начальная часть таким образом, чтобы общая продолжительность речи говора оказалась примерно такой же, как у самого «короткого» говора. Наша гипотеза заключалась в том, что результаты должны быть лучше, чем в случае неуравновешенных речевых данных.

Этот эксперимент такой же, как 1-й, только вместо полных фонограмм используются «уравновешенные».

На **векторах средних значений** евклидова – хороша, Чебышёва – очень плоха, а Кульбака-Лейблера – бессмысленно, т.е., ведут себя также, как в неуравновешенном случае, а у L1 появляются несоответствия, которых не было, т.е., можем сделать вывод, что евклидова более стабильна к уравновешиванию, чем L1.

Таблица 4

Пример значений евклидовой метрики для векторов средних значений моделей гауссовой смеси, созданных на уравновешенных фонограммах без адаптации

Аулея	0,000	109,163	122,115	119,603	116,402
Бальтиново	109,163	0,000	122,018	120,539	103,089
Дундага	122,115	122,018	0,000	125,585	117,541
Рудзаты	119,603	120,539	125,585	0,000	116,709
Вилек	116,402	103,089	117,541	116,709	0,000
	Аулея	Бальтиново	Дундага	Рудзаты	Вилек

На **ковариационных матрицах** евклидова метрика – хороша, расхождение Кульбака-Лейблера – терпимо, т.е., ведут себя похоже, как в неуравновешенном случае, правда, несоответствия Кульбака-Лейблера касаются других говоров, нежели в нём. L1 – портится. И портится также и Чебышёва, т.е, она тоже нестабильна к уравновешиванию.

На **весовых векторах** у евклидовой остаются те же проблемы, что и в неуравновешенном случае, правда – в менее выраженной форме, у Кульбака-Лейблера – тоже, и в обоих направлениях, правда, частично меняются проблемные говоры, у L1 – на одно несоответствие меньше, а у Чебышёва их численно вдвое больше, хоть суть не меняется. Расхождение Кульбака-Лейблера имеет лишь одно несоответствие, при том только в одном направлении, и оно гасится путём симметризации. Метрика Чебышёва ведёт себя иначе, с небольшими несоответствиями.

Эксперимент № 4: модели с адаптацией на уравновешенных фонограммах. В свою очередь, этот эксперимент такой же, как 2-й, но вместо полных фонограмм используются «уравновешенные».

На **векторах средних значений** евклидова ведёт себя также, как без уравновешивания, т.е., хуже, чем без адаптации. Кульбака-Лейблера – также бессмысленно. L1 – тоже также, но несоответствия не так ярко выражены. Чебышёва выглядит более-менее прилично, но присуствует одно сильное несоответствие.

Таблица 5

Пример значений метрики Чебышёва для ковариационных матриц моделей гауссовой смеси, созданных на уравновешенных фонограммах с MAP-адаптацией

Аулея	0,000000	1,553342	**1,298674**	1,300741	**1,939994**
Бальтиново	1,553342	0,000000	1,897925	1,705610	1,243552
Дундага	1,298674	1,897925	0,000000	2,039151	1,444145
Рудзаты	1,300741	1,705610	2,039151	0,000000	2,037265
Вилек	**1,939994**	1,243552	**1,444145**	2,037265	0,000000
	Аулея	Бальтиново	Дундага	Рудзаты	Вилек

На **ковариационных матрицах** нет существенных отличий от неуравновешенного случая, разве что у Чебышёва меняются несоответствующие говоры и уменьшается выраженность.

На **весовых векторах** уравновешевание портит и евклидову, и L1 (она как и в неуравновешенном случае, немного хуже евклидовой), и Кульбака-Лейблера (количество несоответствий не увеличивается, но они становятся симметричными), и Чебышёва (вместо одной проблемы теперь их целая куча).

Экспертная оценка. Во время проведения экспериментов мы результаты оценивали «на глаз» (ибо обладаем некоторыми знаниями в балтийской диалектологии и имеем представление о близости и отдалённости говоров). Но, строго говоря, для научного доказательства нужна более формальная оценка. Так как численные абсолютные значения метрик и расхождений различаются, а нас интересуют лишь относительные отношения – что чему ближе, то естественным решением было провести иерархическую категоризацию результатов и сравнивать уже результаты категоризации – т. н. дендрограммы (бинарные деревья с мечеными листьями и немечеными внутренними узлами). Путём опытов и анализа мы пришли к выводу, что наиподходящей является аггломеративная иерархическая категоризация.

Для полученных результатов категоризации единственным возможным способом формализации их оценивания было проведение экспертной оценки, и так как готового решения для такого рода данных (дендрограмм) мы не нашли, то разработали сами свою метрику, которую описали в [1], и назвали её метрикой иерархического выбора.

В литературе описано, что в качестве экспертов можно привлекать только хороших специалистов в оцениваемой области, при том их оптимальное число – от 10 до 20[9]. Поскольку эксперименты проводились на латышских говорах, то мануальную оценку провести были способны провести лишь специалисты очень узкого круга, которые в Латвии занимаются или занимались диалектологией. Мы отобрали 17 таких людей и попросили провести оценку наших подготовленных данных. 12 из них просьбу выполнили.

Перед использованием экспертной оценки необходимо убедиться, что уровень их компетенции достаточно высок. Есть несколько способов, как это сделать[10], но для нас единственным доступным вариантом была оценка компетенции, основываясь на данных самого опроса – эту оценку называют также определением степени единомыслия[11], которую при этом можно и улучшить путём исключения сильно отличающихся. Оценку мы проводили своей новозаданной метрикой. В результате из 12 экспертов мы исключили одного, который слишком выделялся, и в оценке результатов наших экспериментов пользовались оценками 11 экспертов.

---

9 Par ekspertiem parasti uzaicina savas nozares labākos speciālistus. Bet vēlams, lai tie būtu zinoši arī blakus nozarēs. ... Ņemot vērā iepriekš teikto un pieredzi, var izteikt rekomendāciju ekspertu kolektīvā iekļaut ne vairāk par 20 ekspertiem, parasti skaitļi ir 15, 12 vai pat 10. To pamato iepriekšējā paragrāfā teiktais, ka iterāciju metodē skaitļi pārstāj mainīties pēc 9-11 ekspertu iekļaušanas kolektīvā. [9]

10 Методы определения степени компетентности экспертов принято делить на такие группы … :

- оценка компетентности экспертов в зависимости от их оценки объектов;
- взаимооценка;
- самооценка;
- оценка по объективным документальным данным об эксперте.

Кратко рассмотрим некоторые методы, относящиеся к указанным группам. Компетентность эксперта, например, определяют в зависимости от того, насколько его оценки согласованы с оценками большинства.

Однако в том случае, когда относительную значимость некоторого множества альтернатив оценивают, к примеру, m экспертов, из которых m-1 экспертов совершенно некомпетентны в рассматриваемом вопросе, а один является высококвалифицированным специалистом, то их оценки с высокой степенью вероятности могут сильно отличаться друг от друга. [2]

11 Literatūrā ir sastopamas metodes, kas tīri formāli ļauj uzlabot ekspertu vienprātības pakāpi. Viena no tām … paredz, ka no kopējās ranžējumu tabulas … pēc kārtas izslēdz kāda eksperta doto ranžējumu. [9]

Таблица 6

Значения метрики иерархического выбора между результатами расчётов рассматриваемых методов и экспертной оценкой

№ эксперимента	*Составляющая модели: расстояния*	*Эксперт №1*	*Эксперт №2*	*Эксперт №3*	*Эксперт №4*	*Эксперт №5*	*Эксперт №6*	*Эксперт №7*	*Эксперт №8*	*Эксперт №9*	*Эксперт №10*	*Эксперт №11*	*Среднее арифм.*
1	вект. ср. знач.: евкл.; L1	0,33	0,00	0,33	0,33	0,33	0,33	0,33	0,33	0,00	0,33	0,33	0,27
1	вект. ср. знач.: Чеб.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
1	ков. матр.: евкл.; L1	0,33	0,00	0,33	0,33	0,33	0,33	0,33	0,33	0,00	0,33	0,33	0,27
1	ков. матр.: КЛ-симм.	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
1	ков. матр.: Чеб.	0,33	0,00	0,33	0,33	0,33	0,33	0,33	0,33	0,00	0,33	0,33	0,27
1	вес. вект.: евкл.; L1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
1	вес. вект.: Чеб.	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
2	вект. ср. знач.: евкл.; L1	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
2	вект. ср. знач.: Чеб.	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
2	ков. матр.: евкл.	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
2	ков. матр.: L1	1,00	0,67	1,00	1,00	1,00	1,00	1,00	1,00	0,67	1,00	1,00	0,94
2	ков. матр.: Чеб.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2	вес. вект.: евкл.; L1	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33
2	вес. вект.: КЛ-симм.	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33
2	вес. вект.: Чеб.	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
3	вект. ср. знач.: евкл.	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33	0,33
3	вект. ср. знач.: L1	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
3	вект. ср. знач.: Чеб.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
3	ков. матр.: евкл.	0,33	0,00	0,33	0,33	0,33	0,33	0,33	0,33	0,00	0,33	0,33	0,27
3	ков. матр.: КЛ-симм.	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
3	ков. матр.: L1	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
3	ков. матр.: Чеб.	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
3	вес. вект.: евкл.	0,33	0,67	0,33	0,33	0,33	0,33	0,33	0,33	0,67	0,33	0,33	0,39
3	вес. вект.: L1	0,33	0,67	0,33	0,33	0,33	0,33	0,33	0,33	0,67	0,33	0,33	0,39
3	вес. вект.: Чеб.	0,67	1,00	0,67	0,67	0,67	0,67	0,67	0,67	1,00	0,67	0,67	0,73
4	вект. ср. знач.: евкл.; L1	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
4	вект. ср. знач.: Чеб.	0,00	0,33	0,00	0,00	0,00	0,00	0,00	0,00	0,33	0,00	0,00	0,06
4	ков. матр.: евкл.; L1	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
4	ков. матр.: Чеб.	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,67
4	вес. вект.: евкл.	0,67	0,33	0,67	0,67	0,67	0,67	0,67	0,67	0,33	0,67	0,67	0,61
4	вес. вект.: L1	0,33	0,00	0,33	0,33	0,33	0,33	0,33	0,33	0,00	0,33	0,33	0,27
4	вес. вект.: Чеб.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Из таблицы № 6 видно, что достаточно большое число результатов по экспертной оценке могут считаться хорошими (чем меньше число в правом столбце, тем лучше).

Выводы. Можем сделать вывод, что в целом метод даёт хорошие результаты и может использоваться по назначению. Попытаемся проанализировать и понять, как он употребим лучше всего.

Из таблицы № 6 видно, что MAP-адаптация, в целом, вредит. Это мы могли себе представить, ибо, по сути, адаптация – это облегчение процесса подготовки данных за счёт разнообразия моделей, правда, в достаточно малой степени, чтобы это не влияло на результаты распознавания. Но то, что не влияет на распознавание речи, может всё-же повлиять при расчёте расстояний между идиомами.

Также видим, что гипотеза об улучшении результатов при помощи уравновешивания объёма данных оказалась ложной – результаты не только не улучшились, но и немного ухудшились. Возможно, потому что больший объём данных отдельных говоров позволял создавать модели этих говоров более качественно, таким образом более ярко выделяя объективные различия говоров. Из чего следует сделать вывод, что увеличение объёма данных может положительно повлиять на результаты.

При анализе конкретных метрик и данных, на которых они рассчитывались, то хорошей и стабильной к уравновешиванию (а значит и к другим изменениям

объёма данных) является евклидова метрика на векторах среднего значения и ковариационных матрицах. Очевидно, можно пользоваться любой из них, а также попытаться на их основе создать новую евклидообразную метрику, которая-бы учитывала и те, и другие данные.

Таким образом, можем сформулировать главный вывод нашей статьи: модели гауссовой смеси могут использоваться для оценки степени близости идиомов, при чём это расстояние может рассчитываться простой и доступной метрикой – евклидовой.

Литература

1. Берзинь А.У. Метрика иерархического выбора и возможности её применения в компьютерной лингвистике // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог 2021». Вып. 20 (27), дополнительный том. М.: РГГУ, 2021.
2. Полегенько А.Ф., Князский О.В. Оценка относительной компетентности экспертов в экспертной группе с использованием матриц парных сравнений // Озброєння та військова техніка. № 3. Київ: Центр. НДІ озброєння та військ. техніки ЗС України, 2014.
3. Садыхов Р.Х., Ракуш В.В. Модели гауссовых смесей для верификации диктора по произвольной речи // Доклады БГУИР, № 4/2003. Минск: Белорусский государственный университет информатики и радиоэлектроники, 2003.
4. ბერზინი ა. ინფორმაციის მოპოვების პრინციპები ფონოგრამების ავტომატური ანალიზისთვის / Принципы сбора информации для автоматизированного анализа фонограмм // ქართული ენა და თანამედროვე ტექნოლოგიები – 2011. თბილისი: „მერიდიანი“, 2011.
5. Bērziņš A.A. Dabisko valodu automatizēta salīdzināšana: disertācija zinātnes doktora (PhD) grāda iegūšanai informācijas tehnoloģijas nozarē. Rēzekne: Rēzeknes Tehnoloģiju akadēmija, MMXX.
6. Comparable Corpora. // MT Research Survey Wiki. University of Edinburgh. Режим доступа: http://www.statmt.org/survey/Topic/ComparableCorpora, свободный – (5.XI.2019).
7. Davis S., Mermelstein P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences // IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980, Vol. 28, No. 4, pp. 357-366.
8. Dobkeviča M. Varbūtību teorijas un matemātiskās statistikas elementi. Daugavpils: RTU DF, 2004.
9. Markovičs Z. Ekspertu novērtējuma metodes. R.: RTU izdevniecība, 2009.
10. Preliminary recommendations on Corpus Typology. EAGLES – Expert Advisory Group on Language Engineering Standards Guidelines, 1996. Режим доступа:: http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html, свободный – (5.XI.2019).
11. Reynolds D. Gaussian Mixture Models. MIT Lincoln Laboratory. Режим доступа:
1. https://www.ll.mit.edu/mission/cybersec/publications/publication-files/full_papers/0802_Reynolds_Biometrics-GMM.pdf, свободный – (21.IX.2016).
12. Reynolds D., Quatieri T., Dunn R. Speaker Verification Using Adapted Gaussian Mixture Models // Digital Signal Processing 10, pp. 19-41, 2000.
13. Reynolds D. Universal Background Models. MIT Lincoln Laboratory. Режим доступа:
2. https://www.ll.mit.edu/mission/cybersec/publications/publication-files/full_papers/0802_Reynolds_Biometrics_UBM.pdf, свободный – (21.IX.2016).
14. Smyth P. The EM Algorithm for Gaussian Mixtures. // Probabilistic Learning: Theory and Algorithms. Irvine: University of California. Режим доступа:
3. http://www.ics.uci.edu/~smyth/courses/cs274/notes/EMnotes.pdf, свободный – (21.IX.2016).
15. Young S., Evermann G., Gales M., Hain Th., Liu X., Moore G., Odell J., Ollason D., Povey D., Valtchev V., Woodland Ph. The HTK Book (for HTK Version 3.4). Cambridge: Cambridge University Engineering Department, 2009.

Анс Улдович Берзинь, Столичный университет Норвегии, e-mail: ansis@latnet.lv

Ans Uldovich Berzin, Oslo Metropolitan University, e-mail: ansis@latnet.lv