КЛАСТЕРИЗАЦИЯ ПАРАМЕТРОВ РЕЧЕВОГО СИГНАЛА С УЧЕТОМ ИХ ДИНАМИЧЕСКИХ ХАРАКТЕРИСТИК

Стукалов Д.Н.

Рязанская государственная радиотехническая академия.
391000, Рязань, ул. Гагарина, 59/1, каф. вычислительной и прикладной математики.
тел. (0912) 21-46-67, e-mail: albatros@dialup.etr.ru

Реферат. Рассматриваются пути повышения эффективности векторного квантования параметров речевого сигнала. Особенности эволюции вектора акустических признаков позволяют применять модифицированные метрики для построения кодовой книги методами кластеризации. При этом возможно значительное снижение негативного влияния эффектов коартикуляции, приводящих к неоднозначной интерпретации векторов обучающей выборки. Сформированная кодовая книга содержит дополнительную информацию, которая позволяет повысить качество функционирования систем связи и распознавания речи.

Векторное квантование широко используется как в технике низкоскоростной передачи речевых сигналов, так и при построении акустических моделей систем распознавания речи. К числу основных преимуществ данного метода кодирования следует отнести:

- возможность выявить подобие между повторяющимися фонетическими элементами языка;
- возможность устранить корреляционные связи между координатами вектора признаков;
- возможность учесть нелинейные зависимости между координатами вектора признаков.

Все это позволяет существенно снизить избыточность цифрового представления сигнала, а, следовательно, и сократить алфавит возможных состояний акустической модели.

Реализация векторного квантования требует применения двухэтапной процедуры. На первом этапе (этапе обучения) происходит формирование кодовой книги векторного квантователя. На втором этапе (этапе квантования) наблюдаемому вектору признаков ставится в соответствие наиболее подходящий вектор кодовой книги. Эффективность функционирования квантователя существенно зависит от качества проведения операции обучения. Как правило, для этой цели используют процедуру кластеризации. Традиционный подход кластерного анализа заключается в сортировке множества векторов обучающей выборки на группы с похожими свойствами - кластеры. Критерием сходства обычно является мера расстояния между вектором обучающей выборки и центром кластера. Пусть обучающая выборка содержит K векторов $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, \ldots \mathbf{o}_K\}$. Тогда, задача кластеризации состоит в нахождении P векторов кодовой книги $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_P\}$, представляющих обучающую выборку с минимальной погрешностью

$$\mathbf{V} \leftarrow \min_{\mathbf{V}} \Delta = \min_{\mathbf{V}} \left[\frac{1}{K} \sum_{i=1}^{K} \min_{j} \rho(\mathbf{o}_i, \mathbf{v}_j) \right], \tag{1}$$

где $\rho(\mathbf{o}_i, \mathbf{v}_j)$ – метрика расстояния соответствующего пространства признаков, Δ – погрешность представления обучающей выборки. Наиболее употребительной принято считать декартову метрику

$$\rho(\mathbf{o}_i, \mathbf{v}_j) = \sqrt{\sum_{m=1}^{M} (o_m - v_m)^2}\ , \tag{2}$$

где o_m, v_m – координаты соответствующих векторов, а M – размерность пространства признаков.

В тоже время не следует забывать, что речевой сигнал в общем случае характеризуется не дискретным набором состояний вектора признаков, а его непрерывной эволюцией. Данный факт приводит к тому, что существует значительная статистическая зависимость между измерениями вектора признаков, взятыми через небольшие отрезки времени. В силу изменчивости голоса, относительно небольшие смещения вектора признаков могут приводить к тому, что более поздний вектор будет поставлен в соответствие другому кластеру. Кроме того, явление коартикуляции, то есть взаимного влияния соседних фонетических единиц, зачастую также приводит к их некорректной идентификации. Основная причина описанных ошибок векторного квантования заключается в том, что процедуры кластеризации и квантования не учитывают динамические свойства вектора признаков речевого сигнала. Для адекватного описания, кроме координат необходимо указывать направление и скорость изменения вектора. Тогда выборка наблюдения может быть представлена в виде последовательности векторов $\mathbf{O}^* = \{\mathbf{o}_1^*, \mathbf{o}_2^*, \ldots \mathbf{o}_K^*\}$. Модифицированный вектор $\mathbf{o}_i^*$ представляет совокупность $\mathbf{o}_i^* = \{\mathbf{o}_i, \mathbf{d}_{oi}, w_{oi}\}$, где вектор $\mathbf{o}_i$ – соответствует координатам вектора наблюдения в i-й момент времени, вектор $\mathbf{d}_{oi}$ – характеризует направление изменения вектора наблюдения, а параметр w_{oi} – скорость.

Избыточность описания поведения вектора параметров в пространстве признаков приводит к тому, что процедура обучения может строиться различными способами, в зависимости от преследуемых целей. Для задач связи вполне приемлемым является критерий (1), обеспечивающий минимально возможную ошибку квантования при заданном количестве кластеров. Основная проблема связана с выбором метрики, соответствующей расширенному описанию вектора наблюдения. Естественно, что мера сходства векторов внутри кластера должна вычисляться не только по их взаиморасположению, но и по схожести направления. Возможный вариант – использование взвешенной метрики

$$\rho^*(\mathbf{o}_i^*, \mathbf{v}_j^*) = a\rho(\mathbf{o}_i, \mathbf{v}_j) + b\rho(\mathbf{d}_{oi}, \mathbf{d}_{vj}) + c\rho(w_{oi}, w_{vj}), \tag{3}$$

где a, b, c – весовые коэффициенты при соответствующих метриках по величине, направлению и скорости.

Сложнее обстоит дело с задачами акустического моделирования в системах распознавания речи. Здесь требуется учитывать не только статистический, но и фонетический состав речевого сигнала. Это значит, что в переделах кластера необходимо группировать такие компоненты вектора наблюдения, которые обладают похожими фонетическими характеристиками. Для этого необходимо сделать некоторые предположения о соответствии поведения характеристик вектора параметров реальному фонетическому окружению. В том случае, если вектор признаков изменяется незначительно (паузы в речи, длительные вокализованные участки и т.п.), удельный вес направления в общей мере сходства может быть снижен. С другой стороны, при быстрых флюктуациях вектора признаков важно сохранить информацию о направлении изменения внутри кластера. Модифицированная метрика, соответствующая данным предположениям может быть записана в следующем виде

$$\rho^*(\mathbf{o}_\mathbf{i}^*, \mathbf{v}_\mathbf{j}^*) = a\rho(\mathbf{o}_i, \mathbf{v}_j) + bw_{oi}\rho(\mathbf{d}_{oi}, \mathbf{d}_{vj}) + c\rho(w_{oi}, w_{vj}), \quad (4)$$

где скорость изменения вектора наблюдения является взвешивающим фактором для метрики направления. Это означает, что чем выше скорость изменения вектора параметров, тем большее влияние на общую меру сходства оказывает направление изменения данного вектора. И наоборот, невысокая скорость соответствует объединению различных по направлению векторов в одинаковые кластеры.

Разница при использовании традиционного подхода (соотношение (2)) и модифицированных метрик (соотношения (3) и (4)) наглядно представлена на рис.1. Здесь представлены векторы обучающей выборки в двухмерном пространстве признаков. Стрелки показывают направление изменения вектора признаков. Результаты кластеризации на два кластера показаны цветом. В левой части рисунка показан результат

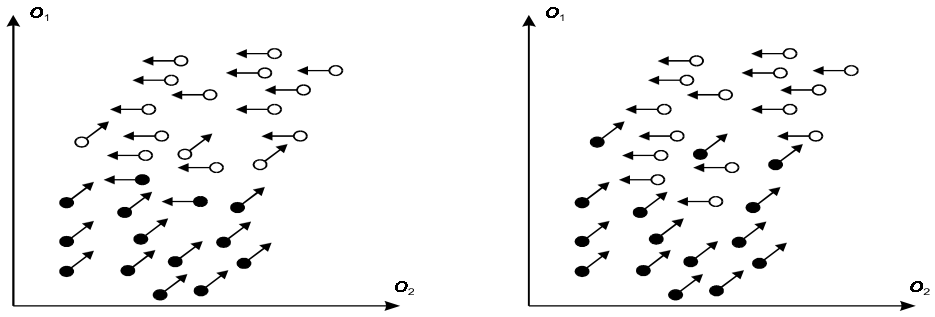

Рис. 1

кластеризации при традиционном подходе. Как видно в один и тот же кластер попадают вектора с различными направленными свойствами. В правой части рисунка представлены результаты кластеризации в соответствии с модифицированной метрикой. Особенно необходимо отметить, что области кластеров могут пересекаться. Это может быть использовано для отражения явления коартикуляции в речи. Дополнительная информация о направлении изменения вектора признаков может быть успешно использована для точного восстановления квантованного сигнала при синтезе речи, а также для интерполяции искаженных при воздействии помех сегментов сигнала. Особо следует отметить сходство векторов кодовой книги и реальных фонетических элементов речи, что позволяет использовать их в большинстве акустических моделей на основе полифонных структур (дифоны, трифоны, слоги). Упрощается также процедура обучения марковских акустических моделей речи.

CLUSTERIZATION OF SPEECH SIGNAL PARAMETERS WITH PROVISION FOR THEIR DYNAMIC CHARACTERISTICS

Stukalov D.N.

Ryazan state radio engineering academy.
391000, Ryazan, Gagarin st., 59/1,
phone: (0912) 21-46-67, e-mail: albatros@dialup.etr.ru

Abstract. A ways of efficiency increasing of vector quantisation of speech signals are considered. Evolution particularities of acoustic features vector allow to apply a modified metrics to building of code book by clusterization methods. Significant decreasing of co-articulation effects, which bring to ambiguous interpretation of study set feature vectors is possible. Formed code book is kept additional information, which allows to rise an operation quality of communications networks and speech recognitions systems.

Vector qvantisation is broadly used as in low speed speech transmitting technique as by acoustic modeling in speech recognition systems. Several of main advantages of given coding method are:

- possibility to reveal resemblance between reiterative phonetic elements of language;
- possibility to avoid correlation between coordinates of feature vector;
- possibility to take nonlinear dependencies account into feature vector coordinates.

All this allows greatly to reduce redundancy of numerical presentation of signal, and, consequently, shorten an alphabet of possible conditions of acoustic models. Realization of vector qvantisation requires using an two-stage procedure. On the first stage (educating stage) occurs a forming of a quantizer code book. On the second stage (qvantization stage) is fixed a correspondence between observed feature vector and more suitable code book vector. Efficiency of quantizer operation greatly depends on quality of education stage.

As a rule, on this stage is used a clusterization procedure. Traditional approach of cluster analysis is sorting an ensemble of study set to groups with similar characteristics - clusters. Criterion of resemblance usually is a measure of distance between study set vector and cluster center.

Let, study set is kept K vectors $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, \ldots \mathbf{o}_K\}$. Then, problem of clusterizations is finding P vectors of code book $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_P\}$, which image a study set with minimum inaccuracy

$$\mathbf{V} \leftarrow \min_{\mathbf{V}} \Delta = \min_{\mathbf{V}} \left[\frac{1}{K} \sum_{i=1}^{K} \min_{j} \rho(\mathbf{o}_i, \mathbf{v}_j) \right], \tag{1}$$

where $\rho(\mathbf{o}_i, \mathbf{v}_j)$ – distance metrics of corresponding feature space, Δ – image inaccuracy of study set. The most common in use is cartesian metrics

$$\rho(\mathbf{o}_i, \mathbf{v}_j) = \sqrt{\sum_{m=1}^{M} (o_m - v_m)^2}\ , \tag{2}$$

where o_m, v_m – coordinates of corresponding vectors, and M – dimensionality of feature space.

In too time should not forget that speech signal, in general, is characterized not discrete set of feature vector conditions, but its unceasing evolution. This fact brings about that that exists a significant statistical dependency between measurements of feature vector, taken through the small time intervals. On the strength of voice variability, comparatively small displacing a feature vector can bring about that that more late vector will be putted in the correspondence to other cluster. Besides, a co-articulation phenomena, that is to say the mutual influence of nearby phonetic units, for-frequent also brings their incorrect identifications. Main reason of described vector quantization mistakes is concluded in that that procedures to clusterizations and quantization do not take dynamic characteristics of into account feature vector of speech signal. For the identical description, except coordinates it is necessary to indicate a direction and velocity of changing of feature vector. Then, observing set can be presented as sequences of vectors $\mathbf{O}^* = \{\mathbf{o}_1^*, \mathbf{o}_2^*, \ldots \mathbf{o}_K^*\}$.

Modified vector $\mathbf{o}_i^*$ is collection $\mathbf{o}_i^* = \{\mathbf{o}_i, \mathbf{d}_{oi}, w_{oi}\}$, where vector $\mathbf{o}_i$ – corresponds to coordinates an observing vector in i-th time moment, vector $\mathbf{d}_{oi}$ – characterizes direction of changing an observation vector, and parametr w_{oi} – velocity.

Description redundancy of behavior of a parameter vector in feature space brings about that that procedure of educating can is built by different ways, depending on persecuted aims. For tasks of communication wholly acceptable is a criterion (1), ensuring minimum possible quantization error under given amount of clusters. Main problem is choice of metrics, corresponding extended description of observation vector. Naturally that measure of vectors resemblance inwardly cluster must be calculated not only upon their location, as well as on directions similarity. Possible variant - using a weighted metrics.

$$\rho^*(\mathbf{o}_\mathbf{i}^*, \mathbf{v}_\mathbf{j}^*) = a\rho(\mathbf{o}_i, \mathbf{v}_j) + b\rho(\mathbf{d}_{oi}, \mathbf{d}_{vj}) + c\rho(w_{oi}, w_{vj}), \tag{3}$$

where a, b, c – - weight coefficients by corresponding metricses of size, direction and velocity. More difficult is a deal with tasks of acoustic modeling in speech recognition systems. It is here required to take into account not only statistical, as well as phonetic composition of speech signal. This signifies that inside cluster boundaries it is necessary to group such components of observed vector, which have similar phonetic characteristics. For this it is necessary to do some suggestions about a correspondence of behavior a characteristics of parameter vector to the real phonetic encirclement. In that case, if feature vector changes small (pauses in speeches, long vocalized intervals etc.), weight of direction in the general measure of resemblance can be reduced. On the other hand, under a quick

fluctuations of feature vector it is important to save information on the change direction inwardly clusters. Modified metrics, corresponding given suggestions can be written as next expression

$$\rho^*(\mathbf{o}_i^*, \mathbf{v}_j^*) = a\rho(\mathbf{o}_i, \mathbf{v}_j) + bw_{oi}\rho(\mathbf{d}_{oi}, \mathbf{d}_{vj}) + c\rho(w_{oi}, w_{vj}), \quad (4)$$

where velocity of changing an observing vector is a weighting factor for the direction metrics. This means that than above velocity of changing a parameter vector, that greater influence upon the general resemblance measure renders a direction of changing a given vector. Conversely, low velocity corresponds an association of direction different vectors in alike clusters. Difference between the traditional approach (expression (2)) and modified metricses (expression (3) and (4)) graphically shown on fig.1. Here presented study set vectors in a two-dimensional feature space. Arrows show a direction of changing a feature vector. The clusterizations results on two clusters are shown by the color. In left part of the drawing shown result of clusterizations under the traditional approach. As seen, one cluster get a vector with different directed by characteristics. In the right of part of the drawing presented the clusterizations results under modified metrics. Particularly it is necessary to note that a clusters area can be crossed. This can be used for reflecting a co-articulations phenomena in speech. Additional information on the direction of

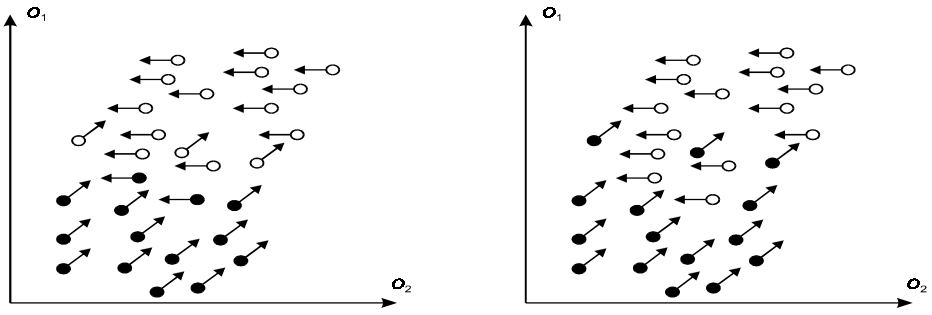

Fig. 1

changing a feature vector can be successfully used for exact recovering a quantizing signal at the syntheses of speech, as well as for interpolation the lost segments of signal. Specifically should note similarity of code book vectors and real phonetic elements of speech that allows to use them in the majority of acoustic models on the base polyphone structures (diphones, threephones, syllables). Also procedure of educating the markov acoustic speech models can be simplified.