Generalized arithmetic coding on simplexes

Dmitriy Kovalev
Novosibirsk State University
Russia
Email: dmitriy.kovalev@gmail.com

Sergey Krendelev
Novosibirsk State University
Russia
Email: tak@mail.nsk.ru

Abstract

In this paper we describe a uniform approach for arithmetic coding understanding. This approach is based on considering n-dimensional simplex and its mappings onto itself.

I. Introduction

Consider the following process of information transmission. Let there be some segment that has a marked beginning, which means that this segment is oriented. Let someone put a mark on this segment and send it to an addressee. This process can be called a message transmission.

For the above actions to really define a message transmission, the mark on the line segment needs to have some meaning. For example, addressee can measure the length of the segment from the segment beginning to the mark. Obtained distance can be interpreted as number in some positional numeral system. Then each digit of this number can be interpreted as a symbol of a message.

II. Main part

Before passing any message with described message transmission algorithm, we should define what does it mean *to measure* the length of the segment. The Greeks would do it the following way. Consider working with an alphabet that consists of 5 symbols. Then original segment can be divided into 5 sub-segments with the help of compass and ruler, using the Thales theorem. In this case the original segment is a unit segment. To determine the first letter of the message, numbered from 0 to 4 , it is necessary to find the number of sub-segment that contains the mark, sub-segments are also numbered from 0 to 4. To decode the rest of the message the sub-segment that contains the mark is divided again and again recursively to provide next symbols of the message. This process is infinite according to Greeks, who believed that line segment can be divided infinite number of times.

This process has two important properties. The fist property is that the result is independent from stretching of the segment. Original segment can be uniformly stretched and the resulting message will still be the same. The second property is that the segment can be divided not only into equal parts but also into commensurable parts according the Thales theorem. It is possible to perform different divisions at each step as well, so the number of sub-segments can be different. Of course, both sender and addressee should know all steps of division process to be able to reconstruct the correct message. Succession of these actions defines the measuring process, what it means to measure the distance from the beginning of the segment to the mark.

Let's translate these actions into mathematical language. Consider two points p_0, p_1 in space and a line segment that connects them. This object is a one dimensional simplex or 1-simplex. Let's define it as $[p_0, p_1]$ and consider that orientation is defined and p_0 is the beginning. Let

$\lambda \geq 0, \mu \geq 0, \lambda + \mu = 1$. Consider point q that lies on the simplex. It can be represented as $q = \lambda p_0 + \mu p_1$. Any pair (λ, μ) defines a point that belongs to the simplex. Such pair is called barycentric coordinates of the point q. It is obvious that p_0 can be represented as $(1, 0)$ and p_1 is represented as $(0, 1)$. If the space where the simplex is defined is a metric space, then barycentric coordinates have the following meaning: μ is the ratio between the distance from p_0 to q and the length of the simplex. In a similar way, λ is the ratio between distance from q to p_1 and the length of the simplex:

$$|q - p_0| = |\lambda p_0 + \mu p_1 - p_0| = |\lambda p_0 + \mu p_1 - \lambda p_0 - \mu p_0| = \mu|p_1 - p_0|,$$
$$|p_1 - q| = |p_1 - \lambda p_0 - \mu p_1| = |\lambda p_1 + \mu p_1 - \lambda p_0 - \mu p_1| = \lambda|p_1 - p_0|.$$

Consider two points that belong to the simplex:

$$q_0 = \lambda_1 p_0 + \mu_1 p_1, \quad \lambda_1 \geq 0, \mu_1 \geq 0, \lambda_1 + \mu_1 = 1$$
$$q_1 = \lambda_2 p_0 + \mu_2 p_1, \quad \lambda_2 \geq 0, \mu_2 \geq 0, \lambda_2 + \mu_2 = 1$$

If point q_0 is closer to beginning of the simplex than q_1, then it follows that $\mu_1 < \mu_2$. Now we have two simplexes $[p_0, p_1]$ and $[q_0, q_1]$. Let's describe a mapping process between them. For this we need to define how to map each point from the first simplex to another point from the second simplex. Consider a point $s = \alpha p_0 + \beta p_1 (\alpha \geq 0, \beta \geq 0, \alpha + \beta = 1)$ that belongs to $[p_0, p_1]$.Its barycentric coordinates are (α, β). The image of point s is point r that lies on $[q_0, q_1]$ and has the same barycentric coordinates as point s. It means that:

$$\begin{aligned} r &= \alpha q_0 + \beta q_1 \\ &= \alpha(\lambda_1 p_0 + \mu_1 p_1) + \beta(\lambda_2 p_0 + \mu_2 p_1) \\ &= \alpha\lambda_1 p_0 + \beta\lambda_2 p_0 + \alpha\mu_1 p_1 + \beta\mu_2 p_1 \\ &= (\alpha\lambda_1 + \beta\lambda_2)p_0 + (\alpha\mu_1 + \beta\mu_2)p_1. \end{aligned}$$

This mapping is linear, so it can be represented as right-multiplication of matrix and vector:

$$(\alpha_1, \beta_1) = (\alpha, \beta)\begin{pmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{pmatrix}.$$

Or in a more familiar form as left-multiplication:

$$\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \lambda_1 & \lambda_2 \\ \mu_1 & \mu_2 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

Thus the mapping between two simplexes can be defined with either

$$\begin{pmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{pmatrix} \text{or} \begin{pmatrix} \lambda_1 & \lambda_2 \\ \mu_1 & \mu_2 \end{pmatrix}.$$

Let's name it as simplex-to-simplex transformation matrix. In particular, point p_0 with original barycentric coordinates $(1, 0)$ is mapped into the point with barycentric coordinates (λ_1, μ_1).

It means that for each simplex mapping to itself there exists a matrix with special properties. The sum of all elements in each row is 1. Such matrices are called stochastic matrices. The multiplication of stochastic matrices is a stochastic matrix:

$$\begin{pmatrix} 1-a & a \\ 1-b & b \end{pmatrix}\begin{pmatrix} 1-p & p \\ 1-q & q \end{pmatrix} = \begin{pmatrix} (1-a)(1-p) + a(1-q) & (1-a)p + aq \\ (1-b)(1-p) + b(1-q) & (1-b)p + bq \end{pmatrix},$$

$$(1-a)(1-p)+a(1-q)+(1-a)p+aq=(1-a)(1-p+p)+a(1-q+q)=1,$$

$$(1-b)(1-p)+b(1-q)+(1-b)p+bq=(1-b)(1-p+p)+b(1-q+q)=1.$$

If a simplex exists in a metric space, then each such mapping is also a contracting mapping.

Let's assign a list of matrices for each partition of the line segment into simplexes. Let $p_0=q_0<q_1<\ldots<q_{n-1}<q_n=p_1$ be a simplex partition, then for each simplex $[q_i,q_{i+1}]$, $i=0,1,\ldots,n-1$ there exists a matrix A_i that maps simplex $[p_0,p_1]$ to simplex $[q_i,q_{i+1}]$. Consider an alphabet of n symbols, where for each i-th symbol matrix A_i is defined. Then for each string of symbols there is pattern that consists of multiplications of matrices that map simplex onto itself. If we use right-multiplication of matrices then we will get a matrix $B=A_{i_1}A_{i_2}\ldots A_{i_s}$, where s is the size of the string that is being encoded. Matrix B uniquely identifies the sequence of matrices.

In other words, B is an encoding of the string. Considering that this is a contracting mapping, it is enough to consider a point with barycentric coordinates $(\alpha,\beta)=(\frac{1}{2},\frac{1}{2})B$. With that point we can uniquely reconstruct all A_i mappings. It is sufficient to keep only the value of α, since $\alpha+\beta=1$. This is the essence of arithmetic coding.

The information discussed above is well known, however, the goal of this paper was encoding of information represented as vectors. For example, consider an image where each pixel is represented as vector that contains red, green, blue color components.

Generalization of arithmetic coding can be explained the following way. Consider a simplex $[v_0,v_1,\ldots,v_n]$ that called S. The order of points defines the orientation of a simplex. Assume that vectors $v_1-v_0,v_2-v_0,\ldots,v_n-v_0$ are linearly independent. Point w that belongs to the simplex and has barycentric coordinates $\lambda_0,\lambda_1,\ldots,\lambda_n$, where $\lambda_i\geq 0$, $\lambda_0+\lambda_1+\ldots+\lambda_n=1$ and $w=\lambda_0v_0+\lambda_1v_1+\ldots+\lambda_nv_n$.

Assume that simplicial partition of simplex S is given. That means that there exist simplexes $S_1,S_2,\ldots,S_r$, such that their union is S. For each S_i there is simplicial mapping A_i that maps S to S_i. Then for each A_i there exists a corresponding mapping matrix.

Consider the alphabet with r elements. Let there be mapping A_i for each alphabet symbol a_i. Then each string of length l can be represented as a sequence of mappings $B=A_{i_1}A_{i_2}\ldots A_{i_l}$. Composition of simplicial mappings is a simplicial mapping as well, which means that original simplex is mapped to another new simplex. All strings of length l are mapped to the partition of simplex S into r^l parts. So, each point from the simplex is unambiguously mapped to a string. In this case, volume is the same as the length from one-dimensional case. For information compression, the first division should be such that the volumes of simplexes from the first partition would be proportional to the number of occurrences of each symbol.

This approach is reversible. Imagine that we need to define a point (or vector) in space. Then let's define a simplex that contains this point and divide such simplex. Afterwards lets create all the mappings A_i. We can find a sequence of nested simplexes that contain original point. This sequence of simplexes is mapped to the sequence of mappings. Each mapping has a number. So, the location of the point in space can be represented as rational number. Note that this is very similar to the case with one-dimensional simplex.