Calculus of Likelihood Ratios

Niko Brümmer

July 7, 2010

1 Introduction

Here we are working in an idealized (e.g. i-vector PLDA) world where we can compute well-calibrated likelihood ratios for a few basic speaker recognition problems. This note is to show how such likelihood ratios are related.

1.1 Data

The data (input) to all problems considered here will be a set of i-vectors, denoted $\mathcal{I}$.

1.2 Hypotheses

We shall express all hypotheses in terms of partitions of $\mathcal{I}$. A partition is a list of $n \geq 1$ non-overlapping subsets, denoted $\mathcal{S}_1 \cdots \mathcal{S}_n$, which together contain all of the elements of $\mathcal{I}$. By grouping elements into subsets, the hypothesis claims that all of the elements in a subset belong to the same speaker and that elements in different subsets belong to different speakers. In this terminology, the partition represents a speaker recognition hypothesis.

> N.B. Do not confuse partition with subset: *Partition denotes the list of all of the subsets.*

1.3 Likelihoods

We shall work with *likelihoods* for partitions. The likelihoods contain the relevant information that the generative model, say λ, extracts from the data, $\mathcal{I}$. If $\mathcal{S}_1 \cdots \mathcal{S}_n$ is a partition of $\mathcal{I}$ into n speakers, then we denote the likelihood as:

$$L(\mathcal{S}_1 \cdots \mathcal{S}_n) \propto P(\mathcal{S}_1 \cdots \mathcal{S}_n | \mathcal{I}, \lambda). \tag{1}$$

The likelihoods are un-normalized, so that a single likelihood is meaningless, unless compared to one or more likelihoods for alternative partitions.

1.4 Independence assumption

We assume the model, λ has fixed, known parameters and has suitable independence assumptions so that the likelihoods multiply as follows:

$$L(\mathcal{S}_1 \cdots \mathcal{S}_n) = \prod_{i=1}^{n} L(\mathcal{S}_i) \tag{2}$$

The LHS is the likelihood for the whole partition. Each factor $L(\mathcal{S}_i)$ in the RHS is the likelihood that the elements of $\mathcal{S}_i$ are all of the same speaker.

2 Definitions

2.1 Likelihood ratio

The information in the likelihoods are unchanged if we normalize them in some convenient way. If the normalizer is also a likelihood, we get a *likelihood ratio*. We will find the following likelihood-ratio to be convenient:

$$R(\mathcal{S}) = \frac{L(\mathcal{S})}{F(\mathcal{S})}, \tag{3}$$

$$F(\mathcal{S}) = \prod_{v \in \mathcal{S}} L(\{v\}) \tag{4}$$

where v denotes an element of $\mathcal{S}$; and where $F(\mathcal{S})$ is the likelihood of the *finest* partition of the elements of $\mathcal{S}$, or the likelihood that all elements have different speakers. The ratio, $R(\mathcal{S})$ therefore compares the speaker hypotheses represented respectively by the coarsest and the finest partitions of $\mathcal{S}$.

> **Example.** $R(\{a, b\})$ is the familiar speaker detection likelihood-ratio, which compares: (i) the hypothesis that a and b are of the same speaker, against (ii) the hypothesis that they are of two different speakers.

Finally, notice that $R(\{v\}) = 1$, where $\{v\}$ is a singleton set.

2.2 Conditional likelihood-ratio

Let $\mathcal{S}' = \mathcal{S} \cup \{v\}$, where v is a new element not in $\mathcal{S}$. We define the *conditional likelihood-ratio* as:

$$R(\mathcal{S}'|\mathcal{S}) = \frac{L(\mathcal{S}')}{L(\mathcal{S})L(\{v\})} \tag{5}$$

This ratio compares the likelihood that all members of the set $\mathcal{S}'$ are of the same speaker, compared to the likelihood that v is of one speaker and all the others in $\mathcal{S}$ are of another speaker.

This is just the familiar *multiple train* speaker detection likelihood-ratio where v is the test and all the rest in $\mathcal{S}$ are multiple training recordings for the target speaker.

3 Relationships

The following relationships can be derived easily from the above definitions and the independence assumption (2).

3.1 The ratios are all you need

Let $\mathcal{S}_1 \cdots \mathcal{S}_n$ be a partition of $\mathcal{I}$ and let $\mathcal{S}'_1 \cdots \mathcal{S}'_m$ be a different partition of $\mathcal{I}$. Then we can compare the speaker recognition hypothesis represented by these two partitions as:

$$\frac{L(\mathcal{S}_1 \cdots \mathcal{S}_n)}{L(\mathcal{S}'_1 \cdots \mathcal{S}'_m)} = \frac{\prod_{i=1}^{n} R(\mathcal{S}_i)}{\prod_{i=1}^{m} R(\mathcal{S}'_i)} \tag{6}$$

To see this note that $\prod_{i=1}^{n} F(\mathcal{S}_i) = \prod_{i=1}^{m} F(\mathcal{S}'_i) = F(\mathcal{I})$. Result (6) shows that we can always use the normalized ratios $R()$ instead of the un-normalized likelihoods $L()$ to do any comparison of hypotheses.

> **Example.** Assume that $\{a, b\}$ are of the same speaker x and that $\{c, d\}$ are of the same speaker y, but x and y may or may not be the same. The likelihood-ratio comparing these two hypotheses is:
>
> $$\frac{R(\{a, b, c, d\})}{R(\{a, b\})R(\{c, d\})}$$

3.2 What is new?

We have already noted that for two inputs $\{a, b\}$, $R(\{a, b\})$ is the familiar, canonical speaker detection likelihood-ratio.

But for three simultaneous inputs $\{a, b, c\}$, is $R(\{a, b, c\})$ a new quantity? How is it related to $R(\{a, b\})$, $R(\{a, c\})$ and $R(\{b, c\})$? The answer is model dependent. If for some simple model, $R(\{a, b, c\})$ is a function of $R(\{a, b\})$, $R(\{a, c\})$ and $R(\{b, c\})$, then this function is parametrized by the model parameters. We cannot answer this question without knowledge of the model and its parameters.

However, as long as (2) holds, we can express the new quantity in a model-independent way by using the *multiple-train* conditional likelihood-ratio (5):

$$R(\{a, b, c\}) = R(\{a, b, c\}|\{a, b\})R(\{a, b\}) \tag{7}$$

This can be generalized recursively:

$$\begin{aligned} R(\{a, b, c, d\}) &= R(\{a, b, c, d\}|\{a, b, c\})R(\{a, b, c\}) \\ &= R(\{a, b, c, d\}|\{a, b, c\})R(\{a, b, c\}|\{a, b\})R(\{a, b\}) \end{aligned} \tag{8}$$

and so on.

4 Summary

Assuming (2), we found $R(\mathcal{S})$ to be a fundamental building block, which can be multiplicatively combined to give likelihood-ratios for comparing any partitioning hypotheses. See (6).

If multiple training can be properly handled (i.e. via (5)), any $R(\mathcal{S})$ can be expressed in terms of the already familiar speaker-detection likelihood-ratios. See (8).