Music Informatics

Alan Smaill

Feb 29, 2016

Today

- Entropy and compactness of representation
- Multiple viewpoints in music

Information and entropy

One way of measuring the randomness or predictability of data arriving sequentially is in terms of entropy.

- If a (loaded) dice always gives a 3 each time it is rolled, it is completely predictable (and minimum entropy $= 0$);
- If there is no pattern to the outcome of the rolls (if every outcome is equally likely), then this is minimally predictable (maximum entropy).

These ideas are important in information theory, and can be used in many contexts where various sorts of information is in play.

Entropy ctd

Official definition:

> *The entropy of a data source is the average number of bits per symbol needed to encode it.*

The symbols are the units or atoms that are found in the data source.
The encoding here allows compression of the data that needs to be transmitted, when there are patterns or redundancies. The definition thus relates to the best way to encode the date; the best encoding by this measure.

What are the symbols?

The definition of entropy depends on what the symbols are:

- If the source gives ABABABABA... we get different answers depending on what the symbols are:
  - A, B as separate symbols?
  - AB as a single symbol?

Entropy ctd

Shannon was the originator of the idea of using entropy in describing information. (The notion of entropy was first introduce in physics (thermodynamics).)
He did some experiments with humans, asking them to predict the next character in English text, at successive characters. He found that entropy here is 1.3–1.6 bits per character. This says that there is a lot of redundancy in English (and character-based natural languages in general).

W cn stll mk prtt gd sns f ntrl lngg whn ll th vwls r mttd!

We can still make pretty good sense of natural language when all the vowels are omitted!

Measuring Entropy

Suppose

- there are n possible outcomes (symbols);
- the measured probability of outcome x_i is $p(x_i)$;

Then the overall entropy is given by H:

$$H = -\sum_{i=1}^{n} p(x_i) log_2(p(x_i))$$

Here the base of the log is arbitrary – base 2 gives entropy in bits.

Note that $0 \leq p(x) \leq 1$, so $log(p(x)) \leq 0$,
so entropy is positive (or 0).

Larger H indicates more entropy, more randomness.

Why the logarithms?

A way to think about the formula:

- $$\sum_{i=0}^{n} p(x_i) log(p(x_i))$$

  is
  - a weighted sum (weighted by $p(x_i)$)
  - of uncertainties $log(p(x_i))$
- Use *log* to get addition rather than multiplication when combining:

$$log(p(x_i).p(x_j)) = log(p(x_i)) + log(p(x_j))$$

Entropy based on context

The usual way to apply this to text is to consider a Markov model of the sequence of characters and use conditional probabilities; write

$$p(X|abc)$$

for the probability that X occurs, given that the last three events were abc (here three is the order of the Markov process).

Given sampled probabilities for different orders, such a model can be used to generate output by generating output randomly with the given probabilities.

This obviously gets better quality if larger order is used (but also need bigger set of probabilities, n^o when there are n symbols and model is of order o).

Musical example

See

`http://www.music-cog.ohio-state.edu/Music829D/Notes/Infotheory.html`

What this doesn't tell us is the representation used:

- absolute pitch, or place in key?
- bar lines as part of the representation?
- note lengths?

Different choices here make a big difference to how well this works.

Entropy of Markov model

The formula here is the generalisation of the basic case:
this means summing over the possible contexts, as well as the symbols.
This then gives a way of comparing different representations.

The idea of multiple viewpoints is that different representations can be compared by their entropy measures. The viewpoints with the best entropy measures tell us about which features are most important for the structure of the music in question.

Viewpoints

See Conklin & Anagnostopoulou, "Representation and discovery of multiple viewpoint patterns", at `http://tinyurl.com/4f79dem` . Some viewpoints:

start time	integer value from start of piece
pitch	integer semitone value
contour	rise, fall, stationary, compared to previous pitch
int	melodic interval with previous note, integer
ioi	inter-onset interval between notes

Some viewpoints can be derived from others – we know that the values of contour can be computed from pitch, for example. The derived viewpoint may be more helpful as a predictor, however.

Combining viewpoints

The Markov analysis can be done separately with any of the viewpoints. We can also look at what happens when combined information between viewpoints is used; write $v \otimes u$ for the combined viewpoint, where the joint occurrence of 2 viewpoints is taken into account; eg pitch, and pitch interval from previous note – something like (32,-2).
Get for example int $\otimes$ ioi and so on.

To measure entropy associated with int $\otimes$ ioi and order 3 for a given piece, need to build probability tables for

$$p(x_i|e_1 e_2 e_3)$$

for each outcome x_i (combination of int and ioi) depending on each preceding sequence $e_1 e_2 e_3$ of similar combinations.
Many of these will never occur.

Markov processes and locality

Because of the local nature of the predictions, Markov models are bad at capturing even medium-level structure. This is apparent in music where metrical structure is global, and where symmetries of phrasing are also common.

A way to help here is to include viewpoints that relate to the larger scale by allowing reference to features that we know will be helpful just for this structure; this is less plausible from a cognitive point of view for a listener, however.

Examples:

fb	boolean: is event first in bar?
fph	boolean: is event first in phrase?
isq	boolean: is event on crotchet (quarter-note) beat?

Some experiments

Conklin and Witten looked at the question of predicting the next note in the melody of a Bach chorale, given a context in the sense above.

They got some human data, by presenting truncated sequences to people and asking them to predict the next note —
in fact want to know more: how confident is the prediction?

Here organise some bets: incorporating this into the experiments gives an idea of the probabilities involved.

Note that the Markov model gives probabilistic estimates of the next event – if two outcomes are equally likely, that should be reflected in human estimation.

Some results

- Just using the basic viewpoints (pitch, onset-time) is not good.
- using some combined viewpoints can help
- best involves using 4 different viewpoints, including pitch, and intfref $\otimes$ fib;
  here intfref is interval from a chosen reference pitch (first/last pitch in melody?)

Another use of multiple viewpoints

See the paper mentioned above.

This uses different representations to look for patterns (according to the different representations) that occur significantly often in a given collection (here Bach chorales).

A similar statistical analysis is used here, not to compare viewpoints, but to measure when patterns are significant for a particular viewpoint.

Prediction via partial match

The use of multiple viewpoint as a predictor for the next event goes back to Cleary, J. G. and Witten, I. H. (1984), "Data compression using adaptive coding and partial string matching" IEEE Trans Communications.

- Look for most likely next event using maximum order data (3 above);
- Back off to lower order if no predictions found,
- . . . until . . .
- use default 0-order model (all outcomes equally probable?)

More recent work on the topic is reported in Whorley et al., "Multiple Viewpoint Systems:Time Complexity and the Construction of Domains for Complex Musical Viewpoints in the Harmonisation Problem":

http://research.gold.ac.uk/9620/1/domain.pdf

The paper describes how to gather information to guide the choice of a good viewpoint, and its use in the harmonisation task.

It also provides empirical data on the time complexity of the processes involved.

Recent work ctd.

The use of multiple viewpoints for harmonisation of hymn tunes is described in detail. The input training data is a set of 50 harmonised hymn tuned in MIDI format, with some hand annotations added for phrasing. The basic method needs to be adapted to work with simultaneous lines of music, rather than simply a single line.

It is crucial to find ways to restrict the domain of the viewpoints used, so that learning can be effective. The article points out several ways to improve this aspect.

The final system is capable of giving reasonable harmonisations in the learned style *without* any explicit overall harmonic rules!

Summary

- Entropy as measure of quality of representation
- Multiple viewpoints