Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Vojtěch Lanz
Název práce	Unsupervised segmentation of Gregorian chant melodies for exploring chant modality
Rok odevzdání	2023
Studijní program	Informatika **Studijní obor** Language Technologies and Computational Linguistics

Autor posudku	David Mareček **Role** oponent
Pracoviště	ÚFAL MFF UK

Text posudku:

The master thesis aims to explore segmentation methods for Gregorian chants and to analyze the relationship between the inferred segments (melody units) to chant modality.

After the introduction, the first chapter describes the chant properties and types and introduces eight chant modalities that differ in final tones and tone ranges. It also introduces the Volpiano format, the encoding of chants that is used in the experiments.

The second chapter describes the background statistical methods that are used in the experiments: Bayesian statistics, Dirichlet and Pitman-Yor process with its hierarchical variant, and the BERT neural model. The third chapter summarizes related work – processing of Gregorian chants and segmentation methods both for natural language texts and melodies. The description of the datasets used and their statistics are given in the fourth chapter.

The fifth chapter shows the list of evaluation metrics for segmentation, the standard ones such as perplexity, vocabulary size, and average segment length; and the chant specific such as alignment with words and phrases or the performance of a classifier using the segments as features to predict the chant mode. It also introduces the visualization methods.

The models used in experiments are described in the sixth chapter. The simpler unigram model with a blocked Gibbs sampler is followed by the complex model based on the Nested Hierarchical Pitman-Yor process. Both models also have their "mode extension", where eight models are trained for each mode separately. This requires the knowledge of the mode for each chant, however, only in the training phase. In the test time, eight segmentations are created using all the models and then, the one with the highest posterior probability is chosen as the result. Finally, a completely different approach using BERT is described.

The experimental results are shown in the seventh chapter. The Pitman-Yor-based models scored the best. Then, the properties of the acquired segments are discussed. They appeared to be

better than the natural segmentation on word boundaries. The segments at the beginning and the end of the chants are more important for the mode prediction. There are also nice visualizations of how the segments are shared among the individual modes.

The last chapter concludes the main observations and proposes future work.

Hodnocení práce:

The thesis is 116 pages long and it is written in very good English. It is well-structured and intelligible, everything is clearly described. There was only one place where I was missing some information to understand the process (see Question 1). I appreciate the excellent theoretical description of the Hierarchical Pitman-Yor process, this is a really tough matter. I also appreciate the very good discussion and clearly summarized results.

I have a slightly different opinion regarding the choice of the best model. In the thesis, the NHPYLMModes model is taken as the best for its higher accuracy in the mode classification task, however, if our goal is rather a good and consistent segmentation, I would prefer the NHPYLM model with lower perplexity and a smaller dictionary. The average length of the segments is also not much lower. The advantage of this model is that sharing the vocabulary would simplify comparisons between modes (see question 2).

Otázky a komentáře:

1. The first question is about the Volpiano format:

   - What do pairs/triples of tones not separated by a dash mean (Figure 1.7)?

   - Is this format directly given to the segmenter, or do you preprocess it somehow, such as removing dashes, bar lines, etc.? In the resulting segments shown in Table 7.13, I can see only letters without dashes. Additionally, for example, bar lines could be used as fixed segment boundaries. But I couldn't find anything about this in the text. Could you clarify it?

2. Another question concerns pairs of authentic/plagal modes.

   - I understand that the difference between them mainly concerns the range, while their final tone is the same. I would like to know to what extent they share segments. In Figure 7.19, we can see that especially the pairs of modes 1,2, and 7,8 share many of them. Also distinguishing between the modes in these pairs is the most difficult for the classifier (Figure 7.16). Unfortunately, these graphs are only shown for the NHPYLM-Modes model, which trains segmentation for each mode separately, and, therefore, it

can easily happen that very similar melodies in two modes would be segmented completely differently. It would be interesting to see segment similarities even for a model using joint segmentation.

- It is interesting that there are no low and high tones (a,b,n,o) within the important features for the classification of modes (Table 7.13). These could distinguish the problematic pairs of modes 1,2 and 7,8. Is it because they rarely occur in the corpus? Perhaps they would occur there when using the joint segmentation model.

3. Gibbs sampling is a stochastic method with random initialization. Have you experimented with several runs with the same parameter setting? How much do the resulting segmentations differ?

Závěr:

I find this thesis connecting Gregorian chants and Bayesian statistics very interesting. The author showed that he understands well the complex theories behind them and performed many experiments with thorough discussions. I definitely recommend this thesis to be defended.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 28. srpna 2023

Podpis: