Book-Adaptive and Book-Dependent Models to Accelerate Digitization of Early Music

Laurent Pugin, John Ashley Burgoyne, Douglas Eck & Ichiro Fujinaga

NIPS MBC Workshop, Whistler
December 2007
OMR on early music sources

Aruspix
Font shape variability

Tenor

Veni vi di, spera flores di.

Secunda pars.

Vedomina celorum, in experta virg
Font shape variability

TENOR

Aem vi di nispa flores di

Secunda pars.

Ve domina celorum in experta virt
Font shape variability

Tenor: Aem vi di, Hispa flores di

Secunda pars: In domina celorum in experta virf
Document degradation variability
Goal of this research

What?
  – Enable Aruspix to be able to handle variabilities more efficiently

How?
  – Make it adaptable to each book
  – Use a supervised adaptation of trained HMMs
  – Does involve manual correction

Why?
  – OMR output has to be corrected anyway in most applications
MAP (maximum a posteriori) adaptation

MAP adaptation in speech

- A speaker-independent model (SI) is built off-line on a large set of data
- During recognition, the SI model is optimized to obtain the speaker-dependent (SD) model using a small set of data

MAP adaptation in OMR

- A book-independent model (BI) is built off-line on a large set of pages
- During recognition, the BI model is optimized to obtain the book-dependent (BD) model using a small set of pages
MAP (maximum a posteriori) adaptation

MAP adaptation in speech

• A speaker-independent model (SI) is built off-line on a large set of data

• During recognition, the SI model is optimized to obtain the speaker-dependent (SD) model using a small set of data

MAP adaptation in OMR

• A book-independent model (BI) is built off-line on a large set of pages

• During recognition, the BI model is optimized to obtain the book-dependent (BD) model using a small set of pages
Experiment data

- 1 BI model trained from scratch on 457 pages from various printed books
- 5 books to experiment with MAP adaptation
  - 50 pages of ground-truth in each book
    - 5 training sets of 40 pages (first pages of the book)
    - 5 testing sets of 10 pages
- Cross-validated one of the books
What We Observe

(a) RISM M.0579 (R. Amadino, Venice, 1587) – baseline rec. rate = 83.91%
(b) RISM M.0580 (G. Vincenti, Venice, 1587) – baseline rec. rate = 67.64%
(c) RISM M.0583 (A. Gardano, Venice, 1603) – baseline rec. rate = 82.85%
(d) RISM M.0585 (P. Phalèse, Antwerp, 1607) – baseline rec. rate = 86.07%

feature vector

V₁ V₂ V₃ V₄ V₅ ... Vₙ
Experiment workflow
Experiment workflow

- Training set (457 pages)
- Training
- Book independent model
Experiment workflow

- Training set (457 pages)
- MAP training set (40 pages)

Training → Book independent model → Book dependent model

MAP adaptation
Experiment workflow

Training set (457 pages)

MAP training set (40 pages)

Training

MAP adaptation

Book independent model

Book dependent model

Baseline

Evaluate MAP adaptation

MAP testing set (10 pages)
Experiment workflow

1. **Training set** (457 pages)
2. **MAP training set** (40 pages)
3. **Book independent model**
4. **Book dependent model**
5. **Trained form scratch model**
6. **Baseline**
7. **Evaluate MAP adaptation**
8. **Evaluate training from scratch**

Flow:
- Training set -> MAP training set
- MAP training set -> MAP testing set (10 pages)
- MAP testing set -> MAP adaptation
- MAP adaptation -> Evaluate MAP adaptation
- Evaluate MAP adaptation -> Evaluate training from scratch
- Evaluate training from scratch -> Trained form scratch model
Results

(a) RISM M.0579

(b) RISM M.0580

(c) RISM M.0583

(d) RISM M.0585

Legend:
- BI model (baseline)
- BA model (MAP adaptation)
- BD model
Analysis of the results

- MAP adaptation improves both recall and precision
- MAP adaptation is faster than training from scratch in most cases
- Only 5 to 10 pages are needed to achieve optimal performance
- MAP adaptation can be used in real-time when alignment is good enough
Conclusion and future work

- Exploit vertical as well as horizontal information
- This comes in the form of harmonic structure
- More complex graphical model combining information from multiple staves
- Appropriate for polyphonic music
Thank you!

Social Sciences and Humanities Research Council of Canada

Canada Foundation for Innovation

McGill
Results - Precision

The graph shows the precision of different methods as a function of the number of pages. The methods compared include Baseline, Training from scratch, and Embedded cumulative MAP adaptation. The precision is measured on a scale from 70 to 100, with the number of pages ranging from 5 to 40.
Results - Recall

Graph showing the recall rate over the number of pages for three different methods: Baseline, Training from scratch, and Embedded cumulative MAP adaptation.