Reducing Costs for Digitising Early Music with Dynamic Adaptation

Laurent Pugin, John Ashley Burgoyne, and Ichiro Fujinaga

Centre for Interdisciplinary Research in Music and Media Technology
Schulich School of Music of McGill University
Montréal, Québec, Canada
{laurent,ashley,ich}@music.mcgill.ca

Abstract. Optical music recognition (OMR) enables librarians to digitise early music sources on a large scale. The cost of expert human labour to correct automatic recognition errors dominates the cost of such projects. To reduce the number of recognition errors in the OMR process, we present an innovative approach to adapt the system dynamically, taking advantage of the human editing work that is part of any digitisation project. The corrected data are used to perform MAP adaptation, a machine-learning technique used previously in speech recognition and optical character recognition (OCR). Our experiments show that this technique can reduce editing costs by more than half.

1 Background

Indexing music sources for intelligent retrieval is currently a laborious process that requires highly skilled human editors [1]. Optical music recognition (OMR), the musical analogue to optical character recognition (OCR), can speed this process and greatly reduce the labour cost. In the case of early documents, the originals for which may not be available to a particular library, it is also important to have a digitisation system that can work with microfilm.

Aruspix is a cross-platform software program for OMR on early music prints based on hidden Markov models (HMMs) [2]. Like the Gamera project [3], it distinguishes itself from most commercial tools for OMR in that it is adaptive. Adaptive systems require training, however, and in order to train them, it is necessary to annotate a large set of images, dozens of images in our case, with complete transcriptions, known as ground truth. Furthermore, early documents suffer from a high and unpredictable level of variability across sources. The font shape varies considerably from one printer to another, and the noise introduced by document degradation or changes in the scanning parameters (e.g., brightness or contrast) may affect the accuracy of the recogniser as well. In these conditions, no single set of models can be expected to perform well for all books, and furthermore, a custom set of models optimised for one book would not necessarily perform well for another.

In a digitisation workflow, the consequence of these problems is that, in order to obtain sufficiently reliable models, one would need several dozen pages to

L. Kovács, N. Fuhr, and C. Meghini (Eds.): ECDL 2007, LNCS 4675, pp. 471–474, 2007.

be transcribed by hand every time a new book was to be processed. Similar issues are encountered in other domains, such as in speech recognition, where the problem is to deal with speaker variability. One common approach to solve the problem is to use dynamic adaptation techniques, such as MAP adaptation [4]. Outside of speech recognition, MAP adaptation has been brought successfully to a number of other fields, including handwriting recognition [5].

In this paper, we present a novel approach in OMR using the MAP adaptation technique. In a preliminary phase, a book-independent (BI) system is trained using pages taken from a number of different books. The BI system gives acceptable results in general but is not specifically optimised for a particular source. During the editing process, the BI system is optimised with MAP adaptation for the book that is currently being digitised. The main idea of the approach is to exploit editing work that has to be done during the digitisation process anyway in order to improve the recognition system. As soon as the editor has corrected the recognition errors on a newly digitised page, that page is used as ground-truth to adapt the BI models. Thus, when starting to digitise a new book, a book-dependent (BD) system can be obtained after only a couple of pages. The adaptation procedure is performed in a cumulative way so that at each adaptation step, it reads all of the pages of the book that have been digitised and corrected up to that point.

2 Experiments and Results

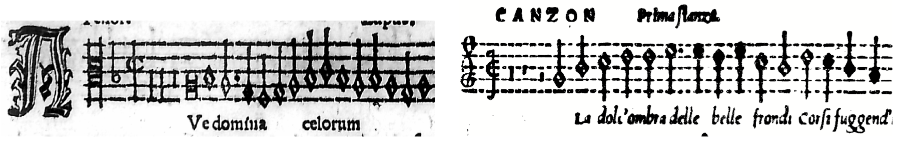
For our experiments, we used a set of microfilms of sixteenth-century music prints held at the Marvin Duchow Music Library at McGill University. They were scanned in 8-bit greyscale TIFF format at a resolution of 400 dots per inch using a Minolta MS6000 microfilm scanner. We used the Torch machine learning library[1] for both training of the BI system and MAP adaptation experiments. The BI system was trained using 457 pages taken from various music books produced by different printers. This set of pages was transcribed and represents a total of 2,710 staves and 95,845 characters of 220 different musical symbols (note values from *longa* to *semi-fusa*, rests, clefs, accidentals, *custodes*, dots, bar lines, coloured notes, ligatures, etc.).

To build BD models with MAP adaptation, we used five other printed music books: RISM[2] 1528-2, 1532-10, V-1421, R-2512 and V-1433 (see figure 1). For each of them, we transcribed 30 pages (150 in total), using 20 pages to build a training set and keeping the 10 remaining pages for a test set. For the training set, we took the first 20 pages of the book because the data become available in this order. For the same reason, we chose not to perform traditional cross-validation across the data set. The baseline for the evaluation was computed by using the BI models to recognise the pages of the 5 test sets.

OMR results are typically presented as symbol recognition rates. From a digitisation prospective, however, it is more beneficial to have an evaluation of the

[1] `http://www.torch.ch`

[2] `http://rism.stub.uni-frankfurt.de`



(a) RISM 1532-10 (Moderne, 1532) **(b)** RISM R-2512 (Gardano, 1575)

Fig. 1. Two prints used to experiment with MAP adaptation. Note the differences in font, line width, background, and overall scanning quality.

human costs. A human editor will always be required to correct the output to library standard, and the cost of this editor will outstrip the cost of the OMR processing time, software, and hardware in the long run. Based on our empirical experience with a human editor for this project, we estimated editing costs considering the following points: (1) deleting a wrongly inserted symbol is a straightforward operation, (2) changing the value of a misrecognised symbol takes twice the time of a deletion on average, and (3) adding a missing symbol is the most time-consuming operation, about four times the work of a deletion. In mathematical form, then, we propose the following average editing cost C per symbol:

$$C = 100 \left(\frac{{}^{1}\!/\!{}_{4}\, D + {}^{1}\!/\!{}_{2}\, S + I}{N} \right) \tag{1}$$

where D is the number of symbols to delete (i.e., wrongly added symbols), S the number of symbols to replace (i.e., misrecognised symbols), I is the number of symbols to insert (i.e., missing symbols), and N is the total number of symbols on the page. Transcribing a page by hand, i.e., without any automatic recognition, would be equivalent to an insertion for every symbol ($C = 100$).

Using MAP adaptation in the digitisation workflow reduced the editing cost on all five sets we used for our experiments (see table 1). In the best case (1532-10, figure 1a), the cost was reduced by a factor of 2.24 with nearly a 15 percent gain in recognition rate. Even for the book where the recognition rate was 95 percent at the beginning (R-2512, figure 1b), our highest baseline recognition rate, MAP adaptation improved the recognition system further, approaching a recognition rate of 97 percent and decreasing the editing cost by 26 percent. On average, the editing costs were decreased by 39 percent. When comparing the results after MAP adaptation to the baseline, in most cases the improvement is already significant after only 5 to 10 pages; only with R-2512, the best-recognised book before adaptation, did it take more than 10 pages to obtain an improvement.

3 Summary and Future Work

When digitising early music sources on microfilm, checking and correcting the OMR output is a highly time consuming step of the workflow. To reduce the editing costs, and to deal with the high variability in the data, we experimented with MAP adaptation within the digitisation workflow. Our results show that

Table 1. Recognition rates and editing costs (see equation 1) before and after MAP adaptation. Adaptation improves editing cost in all cases.

Book	Recognition rate		Editing cost	
	Baseline	MAP	Baseline	MAP
RISM 1529-1	84.16	91.90	9.21	4.99
RISM 1532-10	74.95	89.39	13.52	6.05
RISM V-1421	92.35	94.50	5.26	4.08
RISM R-2512	95.10	96.97	3.46	2.56
RISM V-1433	91.31	95.72	5.56	3.08

this approach can improve the recognition system and reduce the editing costs even when using only a couple of pages, which means that the editors can very quickly glean time-saving side effects from their required work when starting to digitise a new book. At this stage, the dynamic adaptation procedure has been fully implemented and integrated into Aruspix.

Although our experiments focused on the digitisation of early music, the efficiency of dynamic adaptation in handling data variability suggest that the approach could be used fruitfully for digitising early documents in general, including books and maps. Dynamically adaptive methods such as ours promise to be a great boon to digital libraries and should significantly reduce the labour costs that affect all major digisitation projects.

Acknowledgements

We would like to thank the Canada Foundation for Innovation and the Social Sciences and Humanities Research Council of Canada for their financial support. We also would like to thank G. Eustace and M. Reckenberg for their help.

References

1. Bruder, I., Finger, A., Heuer, A., Ignatova, T.: Towards a digital document archive for historical handwritten music scores. In: Sembok, T.M.T., Zaman, H.B., Chen, H., Urs, S.R., Myaeng, S.-H. (eds.) ICADL 2003. LNCS, vol. 2911, pp. 411–414. Springer, Heidelberg (2003)
2. Pugin, L.: Optical music recognition of early typographic prints using hidden Markov models. In: Proc. Int. Conf. Mus. Inf. Ret., Victoria, Canada, pp. 53–56 (2006)
3. MacMillan, K., Droettboom, M., Fujinaga, I.: Gamera: Optical music recognition in a new shell. In: Proc. Int. Comp. Mus. Conf., pp. 482–485 (2002)
4. Gauvain, J.L., Lee, C.H.: Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. IEEE Trans. SAP 2(2), 291–298 (1994)
5. Vinciarelli, A., Bengio, S.: Writer adaptation techniques in HMM based off-line cursive script recognition. Pat. Rec. Let. 23, 905–916 (2002)