

## Chapter 98

# Applying Machine Learning in Optical Music Recognition of Numbered Music Notation

**Fu-Hai Frank Wu**

*National Tsing Hua University, Hsinchu, Taiwan*

### **ABSTRACT**

*Although research of optical music recognition (OMR) has existed for few decades, most of efforts were put in step of image processing to approach upmost accuracy and evaluations were not in common ground. And major music notations explored were the conventional western music notations with staff. On contrary, the authors explore the challenges of numbered music notation, which is popular in Asia and used in daily life for sight reading. The authors use different way to improve recognition accuracy by applying elementary image processing with rough tuning and supplementing with methods of machine learning. The major contributions of this work are the architecture of machine learning specified for this task, the dataset, and the evaluation metrics, which indicate the performance of OMR system, provide objective function for machine learning and highlight the challenges of the scores of music with the specified notation.*

### **INTRODUCTION**

Music notation is the most important media for encoding music to play and read, therefore many classical master pieces encoded in the staff notation, conventional western music notation (CWMN) (Rastall, 1983), are alive for hundreds of years. Nonetheless the notation is the most popular notation worldwide, other musical notations exist and are used in daily life. For example, the numbered music notation (Huang, 2008), named as ‘jiǎnpǔ’ in pinyin for Mandarin, is well accepted in Asian. The research focuses on the sheet music written in the notation. The authors develop the processing engines and build the whole ecosystem of Optical Music Recognition (OMR) and perform experiments. The ecosystem is comprised of dataset, evaluation metrics, groundtruth, OMR processing engines.

DOI: 10.4018/978-1-7998-2460-2.ch098

The major differences between the staff notation and the numbered musical notation are the types of musical glyphs, for example the representative staves and digits, respectively. To name a few of the differences for pitch, the note head position relative to the staves is substituted by numbers {1, 2, 3, 4, 5, 6, 7} and the octave ‘.’ under or above digit (see blue box in Figure 1); the variety of rest notation is simplified by combination of note length glyphs and digit ‘0’. The note length (duration) of note head and flag is switched to the numbers of ‘-’, ‘.’, and ‘\_’ related to note digits (see red circles in Figure 1). On the other hand, some of musical glyphs, for instances ties and Volta brackets, are common for each other.

The research dataset is consisted of first 110 music score manuscripts of a hands-on songbook for singing reference. Those manuscripts are divided into two parts: one is the training set with 100 pieces of short scores with the average of 74 digit notes per sheet of music; the other is the test set with 10 pieces of sheet music with 65 digit notes per sheet. In addition to the digit note, that sheet music includes dynamic notations, such as tie and slur, structural notations to indicate the flow of notes in sheet music and accidentals, flat or sharp, to modify the pitch of notes. Without surprising the number of accidentals and structural notations are quite few to be well studied. However, the system also includes them in the processing steps and evaluation metrics for completeness of notation recognition.

Nowadays sheet music has been digitized for preservation and distribution, but only image digitization could limit the possible applications of Music Information Retrieval (MIR). On the other hand, the symbolic representation, for example MIDI (MIDI association) and musicXML (musicXML), has encoded music into musical entities. With those musical symbols, musical computing devices, such as MIDI instruments, could parse and play music accordingly without human intervention. In order to interpret automatically, the OMR system provides an intelligent and convenient bridge to translate score image into symbolic presentation. While the literatures related to OMR of the numbered musical notation are hard to find, the authors only refer to the literatures of staff notations in terms of the challenge (Bainbridge & Bell, 2001), issue (Rebelo et al., 2012), evaluation (Byrd & Simonsen, 2015), implementation (Bellini et al., 2001; Pugin et al., 2007; Tardón et al., 2009; Coüasnon & Camillerapp, 1994; Fornés et al., 2005; Ng & Boyle, 1996; Pinto et al., 2011; Rebelo et al., 2011; Toyama et al., 2006), improvement with machine learning (Fujinaga, 1996; Pugin et al., 2007; Rossant & Bloch, 2007) for possible commonality between different notations.

In an OMR system, the evaluation usually is executed on the recognized targets, which different researchers could define dissimilarly, for example note pitch and length. In staff music notation, the correctness of note pitch implied the correctness of the related musical symbols: note head (head vertical position on the staves), accidental, and possible interaction with key signature. If the evaluation result only show the results of note pitch, there are no ways to identify the problems of recognition of note head, staff, accidental, and accidental. Therefore, the authors propose detailed evaluation considering both the processing stages, for example staff recognition, and musical semantic, such as note pitch.

Another common issue of evaluation is that performance metrics are conducted with intensive labor. Nonetheless, Szwoch (Szwoch, 2008) proposed the automatic methodology to utilize MusicXML as encoding format for groundtruth. Although the concept is quite similar to the evaluation framework of the authors, the innovative implementation detail is proposed furthermore. The authors also refer to the automatic evaluation counterparts (Kana et al., 1995; Stamatopoulos et al., 2008) of optical character recognition (OCR) as the equivalence of OMR in text domain. No matter how to make the evaluation, groundtruth annotation are necessary, labor-intensive and error-prone. For those reasons, musicological properties, time signature and pitch transition, are deployed to find those errors. Not only the groundtruth could have errors, but the scores of datasets also have due to handwriting input in the creation process.

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/applying-machine-learning-in-optical-music-recognition-of-numbered-music-notation/252118](http://www.igi-global.com/chapter/applying-machine-learning-in-optical-music-recognition-of-numbered-music-notation/252118)

## Related Content

---

### Rethinking Bloom's Taxonomy: Implicit Cognitive Vulnerability as an Impetus towards Higher Order Thinking Skills

Caroline M. Crawford and Marion S. Smith (2015). *Exploring Implicit Cognition: Learning, Memory, and Social Cognitive Processes* (pp. 86-103).

[www.irma-international.org/chapter/rethinking-blooms-taxonomy/120854](http://www.irma-international.org/chapter/rethinking-blooms-taxonomy/120854)

### Real-Time Streaming Data Analysis Using a Three-Way Classification Method for Sentimental Analysis

Srinidhi Hiriyannaiah, G.M. Siddesh and K.G. Srinivasa (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 1377-1390).

[www.irma-international.org/chapter/real-time-streaming-data-analysis-using-a-three-way-classification-method-for-sentimental-analysis/252086](http://www.irma-international.org/chapter/real-time-streaming-data-analysis-using-a-three-way-classification-method-for-sentimental-analysis/252086)

### Mammogram Classification Using Support Vector Machine

Youssef Ben Youssef, Elhassane Abdelmounim and Abdelaziz Belaguid (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 894-921).

[www.irma-international.org/chapter/mammogram-classification-using-support-vector-machine/252062](http://www.irma-international.org/chapter/mammogram-classification-using-support-vector-machine/252062)

### A Social Constructionist Model for Human-Machine Ecosystems

Neus Lorenzo Galés and Ray Gallon (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 1938-1961).

[www.irma-international.org/chapter/a-social-constructionist-model-for-human-machine-ecosystems/252119](http://www.irma-international.org/chapter/a-social-constructionist-model-for-human-machine-ecosystems/252119)

### Brave New Digital Tools for Action Research in Education: A Beginner's Guide

Reinhard Bauer, Klaus Himpsl-Gutermann, Martin Sankofi, Petra Szucsich and Ruth Petz (2020). *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications* (pp. 469-490).

[www.irma-international.org/chapter/brave-new-digital-tools-for-action-research-in-education/252039](http://www.irma-international.org/chapter/brave-new-digital-tools-for-action-research-in-education/252039)