

ALGORITMI PER IL MUSIC INFORMATION RETRIEVAL

Antonello Rizzi, Mario Antonelli

Dipartimento INFOCOM
Università di Roma “La Sapienza”
Via Eudossiana 18, 00184, Roma

Negli ultimi anni l'interesse dell'industria multimediale si è spostato verso nuove forme di vendita di prodotti musicali. La musica è uno dei contenuti multimediali più diffusi e disponibili online. Già da anni sono stati attivati centinaia di servizi in internet per lo streaming e il download di brani audio. I nuovi formati di compressione ed i sistemi di memorizzazione sempre più economici hanno permesso la creazione di database di dimensioni considerevoli, capaci di contenere centinaia di migliaia di brani musicali o addirittura milioni di brani, nel caso dei siti più autorevoli. La ricerca in database di tali dimensioni è un problema di non semplice soluzione. Attualmente le tecniche di ricerca si basano sull'informazione ancillare descritta in forma esplicita in appositi tags (metadata). Tali informazioni sono usualmente inserite direttamente dagli autori o dal gestore del database.

Un approccio emergente consiste nelle tecniche di ricerca basate sul contenuto, anche dette “Content-Based”. Secondo tale paradigma un utente può effettuare una ricerca all'interno di un database di grandi dimensioni semplicemente fornendo un esempio della melodia caratterizzante il brano desiderato. L'esempio può essere fornito da utenti esperti tramite una tastiera MIDI sulla quale suonare il motivo desiderato, o da utenti non esperti canticchiando un motivetto ad un microfono collegato ad un PC; quest'ultima tecnica di ricerca di brani musicali viene comunemente detta “Query-by-Humming”. Molti ricercatori hanno indirizzato la propria attenzione verso il “Music Information Retrieval” (MIR), un ambito di ricerca che richiede la padronanza di una vasta gamma di conoscenze (tecniche di elaborazione audio, tecniche Pattern recognition e di soft computing, algoritmi di ottimizzazione, software engineering, musicologia, ecc.).

Nel corso dell'ultimo anno, parte dell'attività di ricerca è stata dedicata alla creazione di un sistema di music retrieval del tipo Query-by-Humming. Una sequenza di note ritenuta significativa dall'utente (“query”) viene canticchiata e contemporaneamente acquisita tramite un microfono; un processo automatico di trascrizione da file PCM a sequenza MIDI può essere attivato al termine dell'acquisizione (elaborazione “batch” su sistemi monoprocesso) oppure lanciata dopo un opportuno time lag contemporaneamente al processo di acquisizione (elaborazione parallela multithread su sistemi multicore). Il file MIDI, che consiste in una rappresentazione simbolica della sequenza di note canticchiate dall'utente, viene utilizzato dal motore di inferenza per cercare all'interno del database il brano più simile alla query. Tale ricerca si basa usualmente su misure di dissimilarità definite tramite algoritmi di programmazione dinamica nello spazio delle sequenze MIDI (misure “di edit” tra sequenze di eventi). Recentemente è stato progettato ed implementato un algoritmo di trascrizione automatica della voce [2]. In tale algoritmo, la query dell'utente, acquisita sotto forma di segnale digitale PCM, viene inizialmente segmentata isolando gli istanti di inizio e di fine delle note canticchiate. La tecnica di segmentazione delle note si basa sull'analisi della derivata dell'involuppo del segnale. Una volta individuati i segmenti di segnale audio associati ad una nota, viene estratto il pitch (o frequenza fondamentale) di ognuno di questi. In letteratura sono state presentate tecniche per l'estrazione del pitch basate sulla funzione di

correlazione; l'algoritmo di trascrizione proposto si basa invece su di una nuova funzione di correlazione generalizzata, detta Correntropia [3].

Le prestazioni del sistema di trascrizione sono state valutate su un insieme di file di benchmark opportunamente selezionati per la misura delle prestazioni degli algoritmi di "Voice to MIDI"; i brani inclusi in tale repository consistono in sequenze di note canticchiate (pronunciando le eventuali parole del brano a cui si riferiscono) o riprodotte tramite sillabe standardizzate, oppure riproducendo i suoni a bocca chiusa ("humming"). Il repository ed i risultati di trascrizione dell'algoritmo proposto, sono disponibili on line [4].

Le registrazioni selezionate sono volutamente caratterizzate da una scarsa qualità; tale scelta è motivata dalla necessità di valutare le prestazioni dell'algoritmo di trascrizione in applicazioni reali, dove cioè la registrazione non può essere considerata ideale per la presenza di rumore di sottofondo di varia natura o di rumore introdotto da sistemi di acquisizione e da convertitori A/D di scarsa qualità. Inoltre, nel caso in cui la conversione Voice to MIDI venisse effettuata direttamente dal server remoto, la query verrebbe opportunamente compressa prima della trasmissione (codifica di sorgente Mpeg Layer III), introducendo rumore di compressione sul file originale. I test sono stati effettuati confrontando le trascrizioni con i file MIDI ricostruiti ispezionando visualmente le tracce audio. Tale tecnica, ampiamente utilizzata in letteratura, seppur noiosa e poco efficiente, consente di misurare rigorosamente le prestazioni di un algoritmo di trascrizione, consentendo il confronto di diversi algoritmi.

La funzione di Correntropia si è rivelata robusta ai diversi tipi di disturbo presenti nelle tracce audio. Inoltre, paragonando la funzione di Correntropia con le tradizionali tecniche di individuazione del pitch, è stata riscontrato che la tecnica proposta ha prestazioni migliori a parità di costo computazionale.

Nel prossimo futuro, tale sistema di trascrizione verrà integrato con un motore di inferenza in grado di effettuare il matching tra una query sottoposta in formato MIDI ed i brani MIDI monotraccia presenti in un archivio. Il popolamento di tale archivio verrà effettuato in modalità automatica ricorrendo a tecniche di "summarization" di brani musicali, basandosi sugli algoritmi di trascrizione automatica di musica polifonica [5] che sono stati messi a punto dal nostro gruppo di ricerca.

[1] Ning Hu and B. Dannenberg, "A Comparison of Melodic Database Retrieval Techniques Using Sung Queries", in *Joint Conference on Digital Libraries*, (2002), New York: ACM Press, (2002), pp. 301-307.

[2] Mario Antonelli and Antonello Rizzi, "A Correntropy-Based Voice to MIDI Transcription Algorithm", submitted to the *2008 International Workshop on Multimedia Signal Processing (MMSp 2008)*.

[3] J. W. Xu and J. C. Principe, "A novel pitch determination algorithm based on generalized correlation function," in *Proc. of Machine Learning for Signal Processing*, 2007.

[4] <http://infocom.uniroma1.it/rizzi/resfe.html>

[5] Antonello Rizzi, Mario Antonelli and Guido Del Vescovo, "A Two-fold Optimization Algorithm for Automatic Music Transcription", submitted to *IEEE Transaction On Speech, Signal and Audio Processing*.