

진화 하이퍼네트워크를 이용한 음악 학습 및 크로스오버 음악 생성

*김현우, 김병희, 장병탁
서울대학교 전기컴퓨터공학부
e-mail : hwkim, bhkim, btzhang @ bi.snu.ac.kr

Learning music and generation of crossover music using evolutionary hypernetworks

*Hyun-Woo Kim, Byoung-Hee Kim, Byoung-Tak Zhang
School of Computer Science and Engineering
Seoul National University

요 약

본 연구는 미디 형태의 음원에서 멜로디 패턴을 학습하고, 학습된 패턴을 기반으로 음악을 생성하는 문제를 다룬다. 언어의 학습 및 연상과정을 모사하는 데 유용한 하이퍼네트워크 프레임워크를 이용하여, 음악 장르별 특징적인 패턴을 학습하고 연상 작용을 통해 학습된 패턴을 인출 및 조합이 가능함을 보인다. 본 논문에서는 장르별로 독립적으로 학습된 여러 개의 하이퍼네트워크를 이용해 여러 장르의 지역적인 패턴을 혼합하여 크로스오버 음악을 생성하는 방법을 제시하고, 가요와 미국민요, 스코틀랜드 민요를 다양한 조합으로 혼합하여 크로스오버 음악을 생성한 결과를 소개한다.

1. 서 론

유전 알고리즘과 진화 컴퓨팅을 이용해서 컴퓨터과학 분야와 예술분야를 접목하려는 시도가 있어 왔다. 새로운 해결방안이나 새로운 패턴을 생성하는 특성을 가진 유전 알고리즘의 특성은 진부한 패턴을 벗어나 참신한 창작물을 만들려고 하는 예술의 특성에 적합할 수 있다. 하지만, 사람들이 예술적 창작물에 공감하기 위해서는 지나치게 생소한 것은 문제가 될 수 있다. 기존의 사람들의 사고방식과 다른 창작물에서 볼 수 있는 공통된 패턴 또한 필수적인 요소이다.

본 논문에서는 하이퍼네트워크를 이용하여 음악 생성에 적용하고자 한다. 하이퍼네트워크는 학습과정에 진화 연산적 요소를 내포하고 있다. 또한 학습 데이터를 통해서 만들어진 하이퍼네트워크와 학습 방법을 이용해 기존의 학습데이터에 존재하는 패턴을 학습할 수 있고 생성할 수 있다[1]. 이러한 두 가지 특성을 통해 하이퍼네트워크를 이용해 주어진 곡들의 패턴을 학습하고 이를 기반으로 생성된, 주어진 곡들의 전형적인 패턴과 유전 알고리즘적 특성을 이용해 만들어내는 새로운 패턴을 조합함으로써 하나의 창작물인 크로스 오버 음악을 만들어내는 방법을 제안해 보고자 한다.

2. 관련 연구

2.1 알고리즘을 이용한 음악 생성

모차르트는 미뉴엣을 완성할 때 미리 만들어 놓은 음악의 조각을 주사위를 던져서 주사위 눈의 합을 이용해 조각들을 연결해서 미뉴엣을 완성했다고 한다. 이와 유사하게 알고리즘을 이용한 음악 작곡에는 우연성을 이용한 방법이 있다. 비선형 동적 방정식을 이용하여 음표를 생성해 내는 방법도 하나의 접근 방법이다[2]. 프렉탈을 이용하여 음악을 생성하는 방법도 있다. Diaz-Jerez의 FractMus가 대표적이다[3]. 유전 알고리즘과 사용자와의 상호작용을 이용해서 사용자가 선호하는 곡을 시스템이 생성하도록 하는 연구가 있었다[4][5].

이와 구분되게 학습 과정을 포함시킨 알고리즘 이용한 음악 작곡 연구가 있으며, 관련 연구의 방향은 크게 세 종류로 구분할 수 있다. 첫째는 음악 반주 생성이다. 둘째는 학습을 통한 음원 복구이며, 셋째는 학습된 음악 스타일을 기반으로 생성해 내는 것이다. 음악 반주 생성의 기존 연구 중 하나인 Music Plus One은 독주자의 반주를 학습하여 생성해 내는 것이다[6]. 또한 MySong은 독창자의 반주를 생성해 내는 연구이었다 [7]. 학습을 통한 음원 복구 방법으로는 동적화음모델이 제안되었다[8]. 학습된 음악 스타일을 기반으로 음악을 생성하는 연구로는 recurrent neural networks을 이용한 방법과[9][10] Hidden Markov models을 이용한 방법[11] 등이 있다. 우리의 연구는 학습된 음악 스타일을 기반으로 음악을 생성하는 연구로 분류될 수 있지만, 세부적으로 여러 학습 데이터의 각 각의 스타일을 학습하여 융합하려는 점에서 차이가 있다. 즉 각기 다른 장르의 음악을 각각 학

습함으로써, 각 장르의 고유한 패턴을 학습하고 각 학습된 패턴을 적절히 조합함으로써 크로스 오버 음악을 생성하거나 새로운 스타일의 음악을 생성 할 수 있다.

2.2 하이퍼네트워크

하이퍼네트워크는 사람의 병렬적 연상 메모리를 모사하기 위해 제안된 알고리즘이다[12]. 적절하게 학습된 하이퍼네트워크는 연상 메모리로서의 역할을 잘 할 수 있게 된다. 이와 관련하여 언어 생성 문제가 연구된 바 있다[1]. 하이퍼네트워크 연구는 패턴 분류를 위한 알고리즘으로서 다양한 분야의 데이터 분류에 적용되었다[13][14]. 또한 주식 예측을 위해서 일정 기간의 값을 연속적으로 생성하여 예측하는 연구도 이루어진 바 있다[15].

하이퍼네트워크의 학습 방법은 주어진 데이터의 부분의 조합으로 이루어진 하이퍼에지의 가중치를 학습하는 과정으로 이루어진다. 하이퍼에지의 차수에 따라서 각 인스턴스에 대한 정보를 갖고 있는 정도가 결정된다. 하이퍼에지의 차수가 높음에 따라 하이퍼에지는 그 한 인스턴스를 잘 표현할 수 있다. 하지만, 다른 인스턴스를 설명할 수 없다. 즉 지역적인 정보(locality)를 갖고 있게 된다. 반대로 하이퍼에지의 차수가 낮아짐에 따라서 매치되는 인스턴스의 수는 많아져 많은 수의 다양한 인스턴스를 설명 할 수 있게 된다. 즉 전역적인 정보(globality)를 갖게 된다. 차수가 다양한 하이퍼에지를 적절하게 섞어 학습된 하이퍼네트워크는 전역적 정보와 지역적 정보가 조화된 glocality 속성을 가지게 된다. 전반적인 경향과 또한 세부적인 사항을 균형있게 기억하는 연상 메모리의 역할을 하게 된다.

본 논문에서는 하이퍼네트워크의 이러한 특성에 착안하여 다양한 차수를 포함하는 하이퍼네트워크를 이용하여 음악의 패턴을 학습 및 생성을 시도하여 그 결과를 정리한다.

3. 하이퍼네트워크를 이용한 음악 학습 및 생성기법

크로스오버 음악을 생성하기 위하여 크로스오버하기 위한 여러 장르의 음악을 선별 한다. 장르별 하나씩 하이퍼네트워크를 독립적으로 학습한다. 충분히 학습된 하이퍼네트워크는 각 장르의 패턴을 포함하고 있다. 크로스 오버 음악을 생성하기 위하여 각 하이퍼네트워크의 가중치를 설정한다. random 수를 생성하여, 가중치 비율의 확률만큼 하이퍼네트워크를 선택하여 주어진 단서를 기반으로 일정길이의 음악적 조각을 생성하게 된다. 이 과정을 그림으로 나타내면 그림 1과 같다.

생성된 음악은 단서를 기반으로 연상 메모리 역할을 하는 하이퍼네트워크에서 인출된 패턴으로 단서에서 연상될 수 있는 각 장르의 흐름을 이어가는 역할을 하게 된다. 생성된 음악적 조각을 이어가면서 단서를 확장해 가고, 확장된 단서를 이용해 다시 다음 음악적 조각을 인출을 반복한다. 단서를 기반으로 연관된 다음 조각을 찾아내는 특성은 비교적 자연스러운 곡을 만들게 한다.

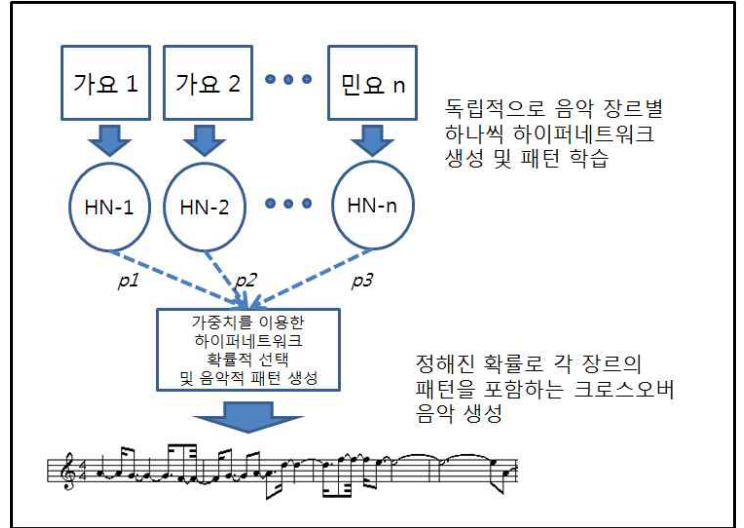


그림 1. 크로스오버 음악 생성과정

3.1 단선율 음악과 다선율 음악

음악은 선율이 중첩되는 것을 기준으로 단선율 음악과 다선율 음악으로 분류할 수 있다. 단선율(monophony) 음악은 한 시점에 하나의 음만이 연주되는 것을 말한다. 그 외 동시에 여러 음이 중첩되어 연주되는 것을 다선율 음악이라고 한다. 단선율의 여집합을 다선율 음악으로 볼 때, 다선율 음악에는 폴리포니(polyphony)와 호모포니(homophony)가 포함된다. 폴리포니는 여러 음이 중첩되고 각의 시작과 끝은 다른 음악을 말한다. 그에 반해 호모포니는 음이 중첩되지만 중첩되는 음들이 동시에 시작하여 동시에 끝나는 음악을 지칭한다.

본 연구에서는 다선율의 midi 파일을 전처리를 통하여 단선율 음원으로 바꾸었다. 단선율의 음악의 패턴을 학습 및 생성하였다.

3.2 크로스오버 음악과 실험 설계

크로스오버 음악이란 어떤 장르에 이질적인 다른 장르의 요소가 합쳐져서 만들어진 음악을 일컫는다. 하지만, 본 연구에서는 크로스오버 음악을 좀 더 넓은 의미로 서로 다른 스타일의 음악적 요소를 합쳐 만든 음악으로 해석하고 음악을 생성하였다. 한국 대중가요 중에서 우리 민요인 '아리랑'을 락 또는 대중가요의 입장에서 재구성한 '윤도현 밴드'의 '아리랑'과 '쾌지나 칭칭나네'를 삽입한 '1TYM'의 '쾌지나 칭칭'과 같은 곡 등은 한 장르의 음악에 이질적인 다른 장르의 요소를 합쳐서 만들어진 점을 생각할 때 크로스오버 음악으로 분류될 수 있다.

본 연구에서 넓게 해석한 크로스오버 음악을 생성하기 위하여 서로 다른 스타일의 대중가요를 혼합하여 음악을 생성하였으며, 또한 민요와 가요를 혼합하여 음악을 생성하였다.

3.3 하이퍼네트워크 학습 과정

표 1. 하이퍼네트워크 생성 및 학습 알고리즘

<p>초기화: 하이퍼에지의 각 차수별 동일한 비율로 무작위로 선택 라이브러리 하이퍼에지의 가중치 초기화</p> <p>학습: For 학습 데이터의 각 멜로디 Do sliding window를 설정하고 단서를 초기화한다. For 생성할 각 음표위치 - 현재 단서와 매치조건을 만족시키는 모든 하이퍼에지를 모은다. - 매치 조건을 만족시키는 하이퍼에지 중 가장 높은 가중치를 갖고 있는 하이퍼에지에 의해 다음 음표를 예측한다. - 매치 조건 : 하이퍼에지의 차수가 k일 때 하이퍼에지의 마지막 vertex를 제외한 나머지 k-1개의 vertices와 주어진 단서의 뒤에서부터 k-1개의 값과 일치 - ‘가중치 갱신정책’[15]을 따라 가중치를 갱신한다. - 현재 단서로 예측된 음표를 현재 단서에 붙여서 단서를 확장한다. End For sliding window를 한 음표만큼 이동한다. Until 멜로디의 끝 교체 정책[14]에 따라 하이퍼에지를 교체한다. End For</p>

하이퍼네트워크 생성 및 학습 과정은 표 1과 같다. 세부적으로는 연속된 sequence를 생성하기 위하여, 각 차수별 가중치 변경 정책을 차수의 크기에 맞게 업데이트 하였다. 차수의 크기를 이용한 것은 진화적 하이퍼네트워크를 이용한 주가 변동 예측을 한 연구에서 사용한 방법이다[15]. 또한 다양한 차수의 하이퍼에지를 포함하는 하이퍼네트워크를 학습시키기 위하여 생존한 하이퍼에지의 분포를 반영해 새로운 하이퍼에지를 선택하는 방법을 적용하였다[16]. 보다 상세한 알고리즘 및 설명은 [17]을 참고하기 바란다.

3.4 크로스오버 음악 생성과정

충분히 학습된 하이퍼네트워크는 각 장르의 부분적인 특성을 잘 반영하며, 주어진 단서에 대해서 예측을 가능한 상태에 이른다[17]. 각 장르의 패턴을 갖고 있는 하이퍼네트워크를 이용하여 다음 음표를 생성함으로써, 생성할

표 2. 학습 음악 구성

인덱스	표기	장르	설명	곡 수
1	Scot.	민요	스코틀랜드 민요	15
2	K-Kim.	가요	김건모 대표곡 모음	15
3	K-Shin.	가요	신승훈 대표곡 모음	15
4	Amer.	민요	미국 민요	15

1. 이어질 음을 예측하는 방법을 통해 하나의 하이퍼네트워크에 하나의 장르 음악을 학습한다. 2개 이상의 장르의 음악을 학습한다.
2. 장르별 가중치 결정(사람이 선택) 한 번에 생성될 음표 수 결정한다.
3. 가중치 비율을 따르는 확률로 하이퍼네트워크를 무작위 선택한다.
4. 선택된 하이퍼네트워크에서 단서를 이용하여 한 번에 생성될 음표 수만큼을 인출한다
5. 3-4를 반복한다.
6. 일정 길이의 음악을 만들면 종료

그림 2. 크로스오버 음악 생성 흐름도

때 마다 각 장르의 패턴을 반영 하게 된다. 그 결과 확률적으로 각 장르의 패턴들이 뒤섞이게 되고, 크로스오버 음악을 생성하게 된다.

4. 음악 학습 및 생성 실험 결과

크로스 오버음악을 생성하기 위해 대중가요 중 성격이 다른 두 가수 김건모와 신승훈의 곡과 외국 민요인 스코틀랜드 민요와 미국 민요를 각 15곡 씩 구성하였다 [표 2]. 외국 민요와 한국의 가요로 크게 구분 될 수 있으며, 각 작은 미국 민요와 스코틀랜드 민요, 김건모의 노래와 신승훈의 노래로 구분될 수 있다. 장르로 분류 했을 때 크게 이질적이며 같은 장르 내에서도 국가별 가수별 서로 이질적이다.

4.1 하이퍼네트워크 장르별 학습

충분히 학습한 하이퍼네트워크는 학습된 곡에서 주어진 단서에 대하여 제한된 범위의 이어질 음표들을 예측 할 수 있다. 학습에 따라 예측 정확도가 증가한다[17]. 그림 3는 주어진 단서에 대해서 연속적으로 6개의 음표를 생성할 때 각 자리의 음표의 예측 적중률을 보여준다. 본 연구에서는 약 35세대를 지나고 나면 학습이 안정되는 것을 확인 할 수 있다.

4.2 음악 학습을 위한 음원 전처리

음악의 패턴을 단순화하고 생성과정을 단순화하기 위해 음원을 단선율로 변화하여 사용하였다. 단선율의 음원을 만들기 위해서 실제 다선율 음원에서 주 멜로디를 포함하는 채널을 사람이 듣고 직접 추출하였다. 또한 추출된 채널을 사람이 직접 듣고 주 멜로디를 찾아서 단선율로 만들었다. 이런 전처리 과정을 거친 음원들은 연속된 음표들의 열거형으로 표현될 수 있다. 전처리 과정을 마친

음원은 쉽게 하이퍼에지에 표현될 수 있고 하이퍼네트워 크로부터 인출 가능하다.

4.3 크로스오버 음악 생성 실험 설계 및 결과

크로스오버 음악 생성 실험은 이음줄(slur)을 제외한 붙임줄(tie)로 연결된 음들을 하나의 음로 계산했을 때 5개의 음으로 이루어진 다양한 단서를 주고 30개의 음을 생성하였다.

4.3.1 한국 대중가요 혼합 생성

본 연구에서 넓은 의미로 해석한 크로스오버 음악의 생성을 위해 서로 다른 스타일의 대중가요를 각각 학습해 크로스오버 음악을 만들어 보았다. K-Kim과 K-Shin을 학습한 하이퍼네트워크의 선택될 확률을 1:4와 4:1로 설정하고 생성하였다[그림 4-A, B]. 그림 4-A, B에서 5음의 동일한 단서가 주어 졌다. 하지만 같은 단서에 대하여 서로 다른 음악을 생성하였다.

4.3.2 외국 민요 혼합 생성

미국 민요와 스코틀랜드 민요를 학습한 하이퍼네트워크의 선택될 확률을 1:4, 4:1로 설정하여 크로스오버 음악을 생성하였다[그림 4-C, D]. 그림 4-D에서 스코틀랜드 민요의 부분을 확인할 수 있으며 변화된 새로운 패턴이 생기는 것 또한 확인할 수 있다.

4.3.3 외국 민요와 한국 대중가요 혼합 생성

우리의 알고리즘은 두 가지 이상의 음악 장르를 혼합하는 것도 가능하다. 대한민국 가요를 바탕으로 외국 민요인 미국 민요와 스코틀랜드 민요를 혼합하고자 했다. 이를 위해서 네 가지의 음악 Scot., K-Kim, K-Shin, USA을 학습한 하이퍼네트워크의 선택될 확률을 1 : 5 : 5 : 1의 비율로 설정하여 생성하였다.

같은 세팅 값을 이용해도 곡을 생성하는데 random으로 하이퍼네트워크를 선택하여 음악을 생성하기 때문에 같은 큐가 주어 졌어도 생성할 때마다 다른 곡을 생성해 낸다.

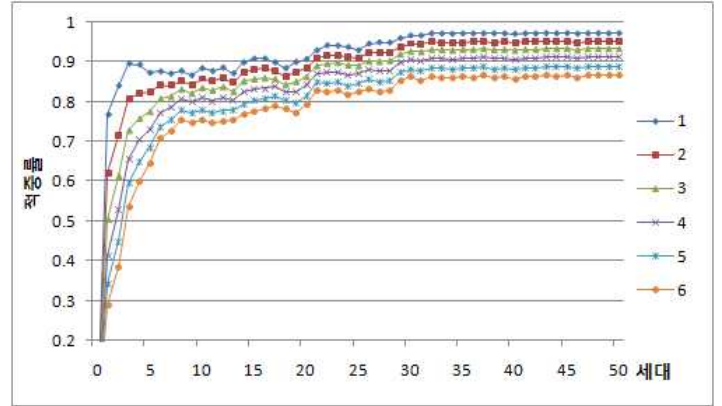


그림 3. 예측된 음표의 적중률

4.3.4 음악 생성의 최소 단위 설정

지금까지 언급된 실험결과는 음악을 생성할 때 한 개의 음표마다 인출될 하이퍼네트워크를 선택하도록 설정하였다. 하지만, 하이퍼네트워크에서 일정길이의 패턴을 인출할 수 있다. 즉 한 번에 하나의 하이퍼네트워크에서 일정 개수의 음표를 인출한 다음 하이퍼네트워크를 다시 선택하는 방법으로 음악 생성의 최소 단위를 변경할 수 있다. 네 가지 음원 집합을 같은 비율로 혼합하였을 때 상대적으로 불안정하고 생소한 패턴이 생성되었다.

하지만 두 개 이상의 음표로 이루어진 음악 조각을 하나의 하이퍼네트워크에서 인출할 경우 상대적으로 사람에게 익숙한 패턴이 생성 되었다. 하나의 하이퍼네트워크에서 인출되는 연속되는 음표의 수가 많아질수록 친숙한 패턴을 생성했다.

즉 생성되는 음악 조각의 크기와 하이퍼네트워크의 선택 확률의 변화를 통해 음악의 참신함과 친숙함의 정도를 조절할 수 있다. 실험적으로 주가 되는 하나의 장르를 선택하여 그 장르를 학습한 하이퍼네트워크의 선택될 확률을 다른 장르를 학습한 하이퍼네트워크의 선택될 확률의 4배 이상으로 설정하는 것이 안정적인 패턴을 생성할 수 있었다.

표 3에 본 논문의 4절에서 소개한 크로스오버 음악 생성 예제 파일 목록을 정리하였으며, MIDI 형태의 음악파일은 <http://bi.snu.ac.kr/~hwkim/kcc09/result.tar>에서 내려받을 수 있다.

A	
B	
C	
D	

그림 4. 크로스오버 음악 생성 결과

표 3. 실험결과 생성된 크로스오버 음악파일 목록. 혼합 비율은 표 2의 인덱스 순서대로 각 장르별 학습결과 하이퍼네트워크가 선택될 확률의 비율을 의미.

파일 이름	혼합 비율	생성 최소단위 (음표수)
4-A.mid	0:1:4:0	1
4-B.mid	0:4:1:0	1
4-C.mid	1:0:0:4	1
4-D.mid	4:0:0:1	1
4-3-3.mid	1:5:5:1	1
4-3-4-1.mid	1:1:1:1	1
4-3-4-2.mid	1:1:1:1	4

5. 결론

음악 학습과 크로스오버 음악 생성을 하기 위해서 하이퍼네트워크를 각 장르별로 하나씩 만들어서 학습하였다. 각 장르별로 학습된 하이퍼네트워크는 학습된 음악을 제한된 길이 내에서 약 85% 이상의 예측 적중률을 보였다. 이질적인 음악을 각 각 학습한 하이퍼네트워크를 확률적으로 선택하여 인출된 음악 조각을 이어가는 방식으로 크로스오버 음악을 생성했다.

하지만, 본 연구에서 단선율 음원을 사용하였으며, 생성된 음악 또한 단선율이다. 좀 더 자유도 높은 음악과 사람과 같은 크로스오버 음악을 생성하기 위하여, 다선율 음악 학습 및 생성 연구가 필요하다.

감사의 글

이 연구는 한국학술진흥재단(KRF-2008-314-D00377), 과학재단 미래유망 파이오니어 사업, 정보통신연구진흥원 IT산업원천기술개발사업(IITA-2009-A1100-0901-1639), BK21-IT에 의하여 지원되었음.

참고문헌

- [1] B. T. Zhang and C. H. Park, "Self-Assembling Hypernetworks for Cognitive Learning of Linguistic Memory," *CESSE2008*, vol. 27, pp. 134-138, 2008.
- [2] J. Pressing, "Nonlinear maps as generators of musical design," *Computer Music Journal*, pp. 35-46, 1988.
- [3] G. Diaz-Jerez, "Fractals and music: Patterns in nature can be a musical inspiration", *Electronic Musician*, 15(10), 108-113, 1999.
- [4] A. Horner and D. E. Goldberg, "Genetic algorithms and computer-assisted music composition," *Urbana*, vol. 51, pp. 61801.
- [5] N. Tokui and H. Iba, "Music composition with interactive evolutionary computation," 2000.
- [6] C. Raphael, "A probabilistic expert system for automatic musical accompaniment," *Journal of Computational and Graphical Statistics*, vol. 10, pp. 487-512, 2001.
- [7] I. Simon, D. Morris, and S. Basu, "MySong: automatic accompaniment generation for vocal melodies," 2008.
- [8] A. T. Cemgil and S. J. Godsill, "Probabilistic phase vocoder and its

- application to interpolation of missing values in audio signals," 2005.
- [9] M. C. Mozer, "Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multiscale processing," *Musical Networks: Parallel Distributed Perception and Performance*, pp. 227, 1999.
- [10] D. Eck and J. Schmidhuber, "Finding temporal structure in music: Blues improvisation with LSTM recurrent networks," 2002.
- [11] J. F. Paiement, Y. Grandvalet, and S. Bengio, "Predictive Models for Music."
- [12] B.-T. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory", *IEEE Computational Intelligence Magazine*, Vol. 3(3), pp. 49-63, 2008.
- [13] S. Kim, S.-J. Kim, and B.-T. Zhang, "Evolving hypernetwork classifiers for microRNA expression profile analysis", *IEEE Congress on Evolutionary Computation (CEC 2007)*, pp. 313-319, 2007.
- [14] J.-W. Ha, J.H. Jang, D.-H. Kang, W.H. Jung, J.S. Kwon, and B.-T. Zhang, Gender classification with cortical thickness measurement, *IEEE International Conference on Fuzzy Systems*, 2009 (accepted).
- [15] E. Bautu, S. Kim, A. Bautu, H. Luchian, and B.-T. Zhang, "Evolving hypernetwork models of binary time series for forecasting price movements on stock markets," *IEEE Congress on Evolutionary Computation (CEC 2009)*, 2009 (accepted).
- [16] J.-K. Kim, B.-T. Zhang, "Evolving hypernetworks for pattern classification," *IEEE Congress on Evolutionary Computation (CEC 2007)*, pp.1856-1862, 2007.
- [17] H.-W. Kim, B.-H. Kim, and B.-T. Zhang, "Evolutionary hypernetworks for learning to generate music from examples", *IEEE International Conference on Fuzzy Systems*, 2009 (accepted).