

음성을 이용한 사상체질 분류 보조 알고리즘

강재환*, 이해정*
한국한의학연구원*

Voice Classification Algorithm for Sasang Constitution

Jaehwan Kang*, Haejung Lee*
Korea Institute of Oriental Medicine*

Abstract - 본 연구에서는 기존의 특정 음성 변수에 대한 모수적 통계 접근 방법을 탈피하고 새로운 음성을 이용한 사상체질 분류 알고리즘을 개발하고자 먼저 5개의 모음과 2개의 문장으로 이루어진 총 120명의 여성 음성 데이터 수집하였다. 이후 다양한 음성 신호 분석 방법과 툴을 이용하여 총 134개의 음성 변수를 추출하였다. 각 변수에서는 체질별 최대값들의 최소값, 최소값들의 최대값을 이용해 4개의 조건 변수를 새로 생성하고 이를 관리하기 위한 메모리와 체질 점수 개념을 도입하여 비모수적인 통계 방법을 기반으로 한 분류 알고리즘을 개발하였다. 알고리즘 성능 테스트를 위해 10-fold cross 검증테스트를 실시하였으며 본 알고리즘은 최종적으로 이진 분류에서 진단률 41.5%와 정확률 79.5%를 가지는 것으로 확인되었다.

Boersma method[4]를 이용하여 피치 추출 알고리즘을 구현하였고 LPC (Linear Prediction Coding)분석과 root extraction 방법을 이용한 포먼트 추출 알고리즘을 구현하였다[5]. 두 번째로 13개의 MFCC 계열 변수를 추출 하였는데 이는 HTK (Hidden Markov Model Toolkit)[6]를 이용하여 12개의 MFCC (Mel-Frequency Cepstrum Coefficient)와 한 개의 Energy 변수를 추출하여 5모음 전체에서 125개의 변수를 구하였다. 마지막으로 문장에서는 음성 분석 프로그램인 PRAAT[7]을 이용하여 문장 발생시의 피치와 강도의 파형을 먼저 구하고 이 두 가지 신호의 10, 50, 90 percentile, 이들 간의 비와 상관도를 구하여 9개의 변수를 생성하여 최종적으로 입력된 음성 데이터는 134개의 음성 변수로 표현되도록 하였다.

1. 서 론

인간을 4가지 체질로 구분하여 병의 원인과 치료 방법을 고려하는 사상체질의학에서는 무엇보다도 정확한 체질 구분이 중요한 작업이 된다. 하지만 체질을 구분하는 방법은 한의사마다 그 기준이 다르고 여러 가지 체질 판별 요소 중에서도 개인의 취향에 따라 특정 요소를 비중 있게 취하기도 한다. 한의학에서 음성은 기본적인 진단 방법인 사진(四診) 즉, 망문문질(望聞問切)에도 포함되며[1] 중요한 체질 판별 요소 중 하나로 사용되기도 한다[2]. 일례로 사상체질의학 관련 문헌에 보면 “태양인은 호흡기가 크므로 소리가 높다. 태음인은 성량이 풍부하여 소리가 무겁다. 소양인은 호흡기가 작으므로 소리가 가볍고 낮다. 소음인은 성량이 넓으므로 소리가 활발하다[3].”로 나타났듯이 예전부터 사상체질과 음성과의 밀접한 관계를 가지고 있다. 하지만 대부분의 사상체질 전문의는 시각, 촉각보다도 더욱 추상적이고 모호한 음성의 감각은 주가 아닌 보조적 형태로 체질 판별 요소로 사용하고 있기에 인간의 음성이라는 하나의 요소만을 가지고 정확도와 진단률이 높은 음성 알고리즘을 구현하는 건 현실적으로 매우 어렵다. 본 연구에서는 이러한 한계를 수용하면서 진단률과 정확률간 성능을 조율하면서 음성을 이용하여 최대한 사상체질을 분류할 수 있으며 차후체형, 피부, 안면 등 모든 체질 판별 요소들로 구성되는 통합체질 알고리즘의 일부문으로 확장이 가능한 알고리즘 개발을 목표로 하였다.

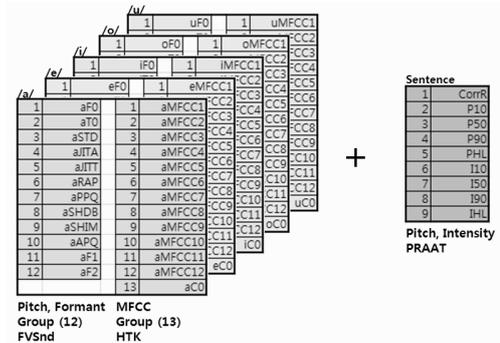
2. 본 론

2.1 실험 방법 및 음성 데이터 수집

본 연구에서는 5년 이상의 사상체질의학 경험이 있는 전문의 3명에게 동일한 체질을 진단 받은 피험자 중에서 건강하고 발음에서도 이상이 없는 소음, 소양, 태음 각각 40명씩 총 120명의 여성을 대상으로 음성 데이터를 수집하였다. 모든 음성 데이터는 동일하게 배경 잡음이 30dB 이하인 조용한 공간에서 실시되었으며 마이크는 Sennheiser e-835s를 마이크 스탠드로 고정하여 입과 마이크 거리가 4-5cm가 되도록 유지하였다. 피험자는 편안히 앉아 있는 상태에서 5개의 모음(/아, /에, /이, /오, /우/)을 각각 2초 이상 자연스럽게 발생하도록 했으며 각 모음 중간에는 약 1-2초 정도의 공백을 유지하도록 하였다. 모음 발생이 끝난 후에는 약 2-3초 공백을 두고 연속해서 하나의 문장(“우리는 높은 산에 올라가 맑은 공기를 마셨습니다.”)을 두 번 반복하도록 하였다. 음성 데이터는 모두 44,100Hz 샘플링 레이트와 샘플당 16 bits로 수집되었다.

2.2 음성 변수

수집된 120명의 음성데이터에서 그림 1에서 나타났듯이 3가지 그룹에 걸쳐 총 134개의 음성 변수를 추출하였다. 첫 번째로 각 모음에서 피치(F0)를 계산하고 계산된 피치로부터 MDVP(Multi-Dimensional Voice Program) 변수를 구하였다 그리고 1,2차 포먼트(Formant(F1, F2))를 추가하여 총 12개의 변수를 추출하였다. 피치와 MDVP 변수 추출을 위해 우리는 자기상관함수와 보간법을 이용하여 정확한 피치를 계산하는



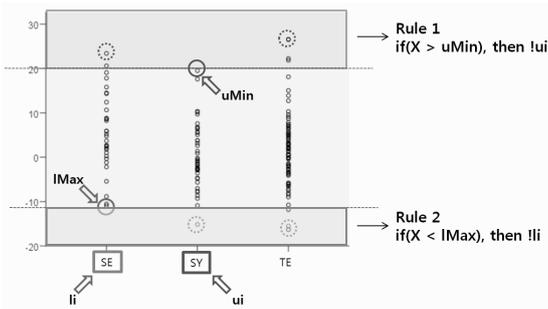
〈그림 1〉 Speech Features

2.3 분류 알고리즘

다양한 정보가 함축된 음성을 통한 사상체질 분류를 위해 비모수 통계 방법을 기본으로 한 규칙 기반 체질 분류 알고리즘을 구현하였다. 이를 위해 먼저 학습에 필요한 train set의 134개 전체 음성 변수를 추출하고 각각의 변수들에서 체질별 최대값과 최소값을 모두 구하였다. 이후 각 체질별 최대값들의 최소값을 $uMin$, 최소값들의 최대값을 $lMax$ 라 정의한다. 또한 $uMin$ 이 발생한 체질의 이름을 ui , $lMax$ 가 발생한 체질의 이름을 li 라는 변수에 저장한다. 즉 하나의 음성 변수에서 $uMin$, ui , $lMax$, li 라는 4개의 조건변수를 새로 생성하였다. 다음으로 생성된 4개의 조건 변수를 이용하면 다음과 같은 두 개의 논리 규칙을 유추해 낼 수 있다.

$$\begin{aligned} \text{If } (X > uMin) \text{ then NOT } ui & \quad \text{식 (1)} \\ \text{If } (X < lMax) \text{ then NOT } li & \quad \text{식 (2)} \end{aligned}$$

즉, 이는 그림 2에서 도식적으로 설명 되었듯이 먼저 구한 모든 변수의 4개의 조건 변수를 가지고 각 음성 변수별로 $uMin$ 을 상위 역치값, $lMax$ 를 하위 역치값으로 설정한다. 그 후에 임의의 음성 데이터의 특정 변수의 값(X)이 입력되고, 이 때 X가 $uMin$ 보다 큰 경우라면 “이 입력 데이터는 ui 체질은 아닐 확률이 그 반대인 경우 보다 더 크다.” 라고 할 수 있다. 동일한 조건으로, 발생한 값이 $lMax$ 보다 작다면 “입력된 데이터는 li 체질은 아닐 확률이 그 반대인 경우 보다 더 크다.”로 유추할 수 있다. 즉 정리하자면 총 134개의 음성 변수에서는 각 변수별로 4개의 조건 변수 ($uMin$, ui , $lMax$, li)를 생성할 수 있으며 이를 이용하여 각 변수별로 2개의 논리식, 총 268개의 체질 판별에 필요한 논리식을 유추해 낼 수 있다. 각 변수별로 4개의 조건 변수는 알고리즘 내에서는 134×4 행렬(이하 $gRule$ 로 명명)로 저장, 관리 할 수 있다.



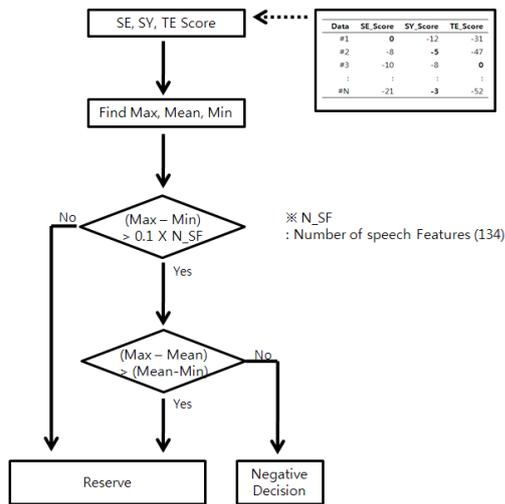
〈그림 2〉 gRule 및 2개의 논리식 생성

2.4 음성 변수와 나이, 키, 몸무게 상관 관계

사용된 134개의 음성변수 중에는 나이, 키, 몸무게와 같은 피험자 신체정보와도 관계가 있는 변수들이 존재하기에 이들의 상관성을 조사하고 이러한 요소를 다음과 같은 방법으로 본 알고리즘에 적용시켰다. 먼저 134개의 음성변수와 나이, 키, 몸무게, BMI (Body Mass Index) 간의 상관성을 조사하였더니 21개의 음성 변수(aF0, aT0, eF0, eT0, iF0, oF0, oT0, uF0, P10, P50, P90, aMFCC9, aMFCC12, eMFCC6, eMFCC11, eMFCC12, iMFCC6, iMFCC9, oMFCC9, uMFCC7, uMFCC9, uMFCC12)에서 나이와, 5개의 음성변수(eF0, P10, P50, P90, aMFCC12)에서 BMI와 상관성이 0.3 이상으로 나타났다. 키와 몸무게에서는 0.3 이상이 되는 상관성이 없었다. 음성 변수와 상관성이 있는 나이와 BMI를 알고리즘에 적용하기 위해서는 나이는 한의학적으로 의미가 있는 16세를 기준으로 5개의 그룹(~16세, 17~32세, 33~48세, 49~64세, 65세~)으로 나누었고, BMI는 4개의 그룹(~20, 20~25, 25~30, 30~)으로 나누어 위에서 언급된 상관성 있는 음성 변수에 대해서는 각 그룹들 끼리 다시 구한 sub-gRule을 생성하였으며 입력된 피험자 나이와 BMI를 확인하여 상위 gRule을 대신에 하위에 존재하는 적절한 sub-gRule을 사용하도록 하였다.

2.5 체질 점수와 판단 규칙

입력의 입력된 음성 신호의 체질 구분은 소음, 소양, 태음 점수가 모두 0으로 초기화 된 상태에서 생성된 gRule을 통과 시키면서 식 (1), (2)와 같은 논리식의 조건에 맞는 경우를 조사하여 이 규칙이 만족되는 경우 이에 맞는 ui, li 체질에 -1점의 벌점을 부과하게 된다. 결과적으로 입력된 음성 변수는 134x2개의 gRule규칙을 통과하면서 체질 점수를 부여 받게 된다.



〈그림 3〉 체질 판단 규칙

최종적으로 체질 진단은 “판단 보류”와 “특정 체질은 아니다.”라는 부정적 판단 두 가지를 내리도록 하였다. 이는 그림 3에서 나타나듯이 먼저 부여 받은 체질 점수의 최대, 중간, 최소값을 찾아서 최대값과 최소값의 차이가 사용된 체질 변수의 개수(N_SF) 134개의 10%인 13점이 되는 경우만을 대상으로 하였고 그 외의 경우는 판단을 보류 하였다. 최대값과 최소값이 13점 이상이 되는 경우에는 다시 최대값과 중간값의 차이와 중간값과 최소값이 차이의 관계를 조사하여, 전자의 경우는 판단을 보류하고 후자의 경우에는 최소값을 가졌던 그 체질에 대해 “이 체질은 아니다.”라는 판단을 내렸다.

2.6 알고리즘 성능 테스트 결과

구현한 알고리즘 성능 테스트를 위해 120명의 음성 데이터를 가지고 5번의 10-fold cross 검정 테스트를 실시하였다. 그 결과 전체 데이터에 대해서 약 58%에 대해서는 판단 보류, 42%에 대해서만 판단을 내렸다. 선택된 판단에 대해서는 약 80%의 정확률을 보이고 있었다. 체질별로는 진단률에서 정확률에서도 소음/비소음, 소양/비소양 보다 태음/비태음에서 진단률도 21.2%로 높고 정확률도 83.5%로 유용성과 구별력이 상대적으로 좋았다.

〈표 1〉 실험 결과

횟수	체질	판단률(%)	정확률(%)
1	소음	7.5	77.8
	소양	12.5	73.3
	태음	22.5	88.9
2	소음	5.0	66.7
	소양	14.2	82.4
	태음	18.3	88.4
3	소음	7.5	66.7
	소양	14.2	70.6
	태음	22.5	81.5
4	소음	5.0	66.7
	소양	15.8	78.9
	태음	20.0	75.0
5	소음	4.2	60.0
	소양	15.8	84.2
	태음	22.5	85.2
총	소음	5.8	68.5
	소양	14.5	78.2
	태음	21.2	83.5

3. 결론

음성을 통한 사상체질 분류를 목적으로 하였던 본 연구에서는 비모수 통계 방법을 기본으로 한 규칙 기반 알고리즘을 구현하였으며 이 알고리즘은 이진 방법으로 약 41.5%의 진단률과 79.5%의 정확률을 가진 것으로 확인이 되었다. 이는 분류 알고리즘이라는 기준에서 보면 그다지 좋지 못한 성능이지만 정성적인 방법으로 체형, 안면, 피부 등 통합적인 요소를 고려해야하는 사상체질 분류의 특수성에서 음성 신호 단독만을 가지고 행한 이번 실험의 한계를 고려하면 가능성만을 제시하였던 기존의 연구보다 좀 더 구체적이고 향상된 성능의 알고리즘 구현으로 사료된다. 이번 연구를 바탕으로 진단률도 중요하지만 그보다는 좀 더 정확률을 향상시키는 연구가 계속되어야 될 것이며, 여성뿐만 아니라 남성의 음성에 대한 연구도 함께 되어야 하며 음성 변수 각각에 대한 물리적 해석도 계속적으로 연구되어야 할 것으로 생각된다.

감사의 글

본 연구는 지식경제부 차세대기술개발사업 중 지능형 한방컨텐츠 개발(10028438)의 지원에 의해 이루어진 것임.

[참고 문헌]

- [1] 왕기 편저, “황제내경 소문금석”, 성보사, 서울, 26-38, 1983
- [2] 김달래, “동의수세보원초고”, 서울, 정담, p.134, 1999.
- [3] 김달래, “오음의 사상의학적 음성분석과 고찰”, 사상체질학회지, 50-59, 2003
- [4] Boersma, P. "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound.", *Proceeding Institute of Phonetic Sciences* 17, 97-110, 1993
- [5] 강재환, “정상 한국인 성별 체형정보와 MDVP 변수간의 상관관계 연구”, *음성과학* 제 15권 제 4호, 107-119, 2008
- [6] (Program) HMM Tool Kit <http://htk.eng.cam.ac.uk>
- [7] (Program) Praat <http://www.praat.org>