

의사결정나무를 통한 걸음걸이 분석 및 질병 예측에 관한 연구

김 영재, 류 관희, 나스리디노프 아지즈
충북대학교 소프트웨어학과

e-mail : {young3609, khyoo, aziz}@chungbuk.ac.kr

A Study on Walking Analysis and Disease Prediction with Decision Tree

Young-Jae Kim, Kwan-Hee Yoo, Aziz Nasridinov
Dept of Computer Science, Chung Buk National University

요 약

본 연구는 키넥트를 통해 사람의 걸음걸이를 측정하고 의사결정트리(Decision Tree)를 통해 분석함으로써 현재의 걸음걸이를 통해 측정자의 허리 또는 무릎에서 발생할 가능성이 높은 문제 또는 질병들을 예측하고 해당결과를 측정자에게 알린다. 본 연구를 진행하며 첫 번째 단계에서는 관련 논문이나 병원 자료 결과들을 통해 판별할 속성들을 정하였다. 두 번째 단계에서는 키넥트를 통해 측정된 실제 데이터를 적용하기에 앞서 첫 번째 단계에서 정한 속성들이 측정자의 문제 또는 질병들을 판단해내는 연관 정도가 높은지 테스트 데이터를 이용하였고 의사결정나무를 통해 분석하였다. 그 결과 7개의 속성 중 6개로 약 85.7% 정도의 연관이 있었다. 마지막 세 번째 단계에서는 판별식을 세우고 실제 데이터들을 쌓아나가며 69명의 측정된 데이터를 분석한 결과 6개의 속성 중 5개의 속성이 허리와 연관 정도가 높았고 이는 두 번째 단계에서 나왔던 결과인 약85.7%에 가까운 약83%의 결과가 도출되었다. 이를 기반으로 시스템을 개발해 나가며 판별 정확도를 향상시키기 위해 계속 측정해 데이터를 쌓아가고 관련된 식들의 문제점을 보완하며 또한 어떤 환경에서 키넥트의 측정값의 정확도가 올라가는지 연구할 예정이다.

1. 서론

기술의 발전은 사람들로 하여금 생활의 편의를 가져다 주었다. 이로 인해 차를 타거나 대중교통을 이용하면서 옛날보다 활동량이 줄어들었으며 취미 활동조차도 활동 범위가 실외에서 실내로 좁아져 실내에서 여러 활동들을 하거나 컴퓨터를 하는 학생들이 많아 졌다. 또한 대부분 사람들의 걸음걸이의 형태가 올바르지 않거나 잘못된 습관으로 인해 2000년대 이후 퇴행성관절염이나 허리 디스크와 같은 고 연령층에서 발생하는 질환들의 발병 연령대가 점점 낮아지고 있는 실정이다. 이러한 결과가 나타난 근본적인 문제 중 하나는 걸음걸이 자세에 대한 무관심에 있다. 이를 해결하기 위해서는 평소 걸음걸이에 대해 관심을 갖고 자신의 걸음걸이가 잘못됐다면 자세 교정을 받아야 하지만 보통 자세가 잘못됐다는 것을 알지 못한다. 실제로 척추치료 전문병원인 한국 자생한방병원의 통계에 따르면 허리디스크 환자 106명을 대상으로 걸음걸이에 대한 조사를 한 결과 환자의 58%(62명)는 안짱걸음, 22%(24명)가 팔자걸음과 같이 올바르지 않은 걸음걸이로 걷고 있었다. 결과적으로 걸음걸이에 대한 낮은 관심도와 바쁜 스케줄 속에서 병원에서 진단받고 교정을 받기란 쉽지 않기 때문

에 본 연구는 걸음걸이의 중요성을 인지하고 위와 같은 문제들을 해결하기 위해 진행되었으며 다음과 같은 3가지의 연구 목표를 선정하였다.

- 걸음걸이로부터 유발될 수 있는 질병과 연관 있는 속성들을 의사결정트리를 통해 분석하고 속성간의 연관 정도를 알아낸다.
- 의사결정트리를 통한 분석결과를 바탕으로 하여 개발된 프로그램의 분석 및 예측 정확도를 높인다.
- 키넥트의 측정 환경에 대해 연구하여 어떤 환경에서 각 관절의 정확한 값을 측정하는지를 분석한다.

본 연구에서는 먼저 현재 걸음걸이를 통해 앞으로 허리에 질병 발생 가능성을 예측하는 자세 연구를 진행하였다. 이를 위해 판별하는 속성들에 대한 연구 분석은 관련 논문과 병원자료를 토대로 데이터 마이닝의 의사결정나무를 통해 분석했으며, 이를 토대로 문제가 있는지에 대해 판별하는 각 속성의 수치 기준과 식을 보완해 나가는 과정을 진행하였다.

2. 관련 연구

Person Identification Using Anthropometric and Gait Data from Kinect Sensor(2015)는 센서의 인체 측정 및 보행 정보를 이용하여 사람을 식별하는 연구를 진행 하였다. 해당 연구의 목표는 정확하게 사람의 보행 상태를 측정하여 개별적인 보행 습관을 통해 개인을 구별 하는 것이었으며 동시에 키넥트센서를 통해 정확한 값을 측정할 수 있는지 확인하는 것이다. 이 연구 논문은 140명 을 키넥트를 통해 측정하였으며 머신 러닝 알고리즘인 KNN (K-Nearest Neighbor), SVM(Support Vector Mac hine), MLP(Multi - Layer Perception) 알고리즘을 통해 걸음걸이를 분석하여 어떤 알고리즘이 걸음걸이 분석에 효율적인지를 분석하였으며 키넥트를 통해서도 어느 정도 정확하게 사람을 식별할 수 있음을 보였다.

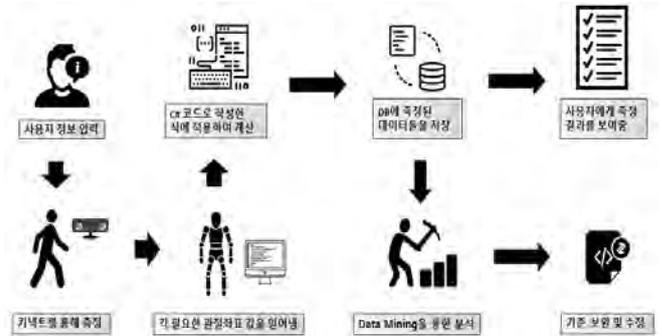
정상인들의 걸음형태에 따른 요통정도와 장애지수와의 관련성 조사(2004)에서는 정상인의 걸음걸이를 분석하여 그에 따른 허리의 통증과 관련된 연구를 진행하였다. 허리와 연관된 문제를 판별하는 속성들을 찾아내고 어떤 식으로 걸을 때 허리에 문제가 되는지 알기 위하여 해당 논문을 참고하였다. 해당 논문에서는 보폭의 넓이, 양발의 넓이, 족궁, 앞으로 나아가는 직선방향에서 발이 벗어나는 각도, 몸무게 등을 속성 값으로 설정하여 연구를 진행하였으며 연구 결과, 환자들은 해당 속성 의 기준 값에서 어느 정도 벗어나는 경우에 허리와 관련된 문제가 발생하는 것을 확인하였다.

걸음걸이 관리를 위한 시각적 피드백 시스템 디자인 (2016)에서는 걸음걸이 관리를 위한 시각적 피드백 시스템 디자인을 연구했다. 이 시각적 피드백 시스템의 구조는 웨어러블 디바이스를 사용자가 직접 착용하여 걸음으로써 나온 결과 값을 스마트폰 어플리케이션을 통해 분석하도록 하였다. 이를 위해 걸음걸이를 분석하는 속성은 발끝이 떨어진 정도와, 팔자걸음으로 걸은 횟수, 안장 걸음으로 걸은 횟수, 총 걸음 횟수 등을 선정하여 사용하였다.

A Kinect-Based Walking Assessment for Disease Prediction(2016)에서는 걸음걸이에 따른 질병 예측에 대한 연구를 진행하였으며 중점적으로 엉덩이 관절을 키넥트로 측정해 현재 걸음걸이로 인해 엉덩이 부분에 문제가 있는지를 분석하고 예측하는 연구를 진행하였다. 키넥트로 양쪽 엉덩이 관절좌표를 이용해 기울기를 계산하였고 약 30명을 측정한 결과 93%의 정확도로 측정자가 문제가 있는지를 판별하였다. 해당 논문과 달리 현재 연구하는 본 논문에서는 허리나 무릎에 대한 문제를 판별하는 것이다. 이를 통해 나온 결과 값들을 의사결정나무를 이용하여 분석하고 그 결과들을 반영하여 판별하는 기준치와 식을 보완해 나가고 키넥트의 측정 환경에 대한 연구를 진행하며 측정 정확도를 향상시키는 연구를 진행한다는 점에서 차이가 있다.

3. 제안 방법

이 장에서는 사용자가 어떤 과정을 통해 측정을 수행하고 분석된 결과를 얻는지 각 단계별로 설명을 한다. 또한 측정된 데이터를 어떤 식에 근거하여 판별하는지 각 속성별로 정의된 식에 대해 설명한다.



(그림 1) 시스템 흐름도

제시된 (그림 1)을 통해 전체적인 시스템 흐름을 보인다. 첫 번째로 먼저 한 사용자가 측정을 하고자 할 때 측정할 이력이 없고 처음 측정하는 경우라면 회원가입을 통해 사용자의 정보를 입력 한 후 고객 정보 검색 페이지에서 새롭게 추가된 자신의 정보를 선택하여 측정단계를 진행한다. 만약 측정할 이력이 있다면 고객 정보 검색 페이지에서 가입한 자신의 정보를 선택한 후 측정단계를 진행하게 된다. 고객 정보 검색 페이지는 여러 가지 기능이 포함되어 있는데 먼저 첫 번째로 해당 사용자의 이름을 검색하여 측정단계로 넘어가기 위해 자신의 정보를 선택할 수 있다. 두 번째로는 자신의 주소나 전화번호를 수정할 수 있고 자신의 정보를 삭제 할 수도 있다. 그리고 마지막으로 세 번째로는 자신의 정보를 선택하여 좀 더 자세한 측정 기록을 볼 수 있도록 하는 기록조회기능을 포함한다.

위의 첫 번째 과정을 수행 한 후 두 번째인 측정 단계에서는 사용자는 키넥트와 약 3.6m 거리에서 측정을 시작하여 자신의 평소 걸음걸이대로 키넥트를 향해 정 방향으로 걷도록 한다.

그런 뒤에 세 번째 단계에서는 진행하여 개발된 프로그램을 통해 측정 된 각 관절 좌표들을 각 공식에 적용하여 값을 계산하고 세워진 기준을 비교하여 현재의 걸음걸이로 견제 되면 앞으로 어디에 문제가 될지 판별하여 관련 데이터들을 도출한다.

마지막 단계에서는 도출된 데이터를 DB에 저장하고 사용자에게 관련 결과를 개발된 프로그램을 통해 보여준다.

추가적으로 DB에 저장된 각 사용자의 데이터들을 의사결정나무로 분석하여 연관된 속성과 연관정도가 낮은 속성 값들을 알아내고 세워진 기준 값을 보완해 나가는 과정을 반복적으로 진행하여 측정을 통한 분석 및 예측 정확도를 높인다.

다음 (그림 2)은 보폭 차이와 보행간격에 대해 혼동될 수 있으므로 용어의 정의를 하였다.



(그림 2) 보폭 차이와 보행 간격 용어 설명

보행 간격 : 보행 시 양 발 사이의 간격

보폭 차이 : 걸음을 걸을 때 앞발 뒤축에서 뒷발 뒤축까지의 거리

또한 분석을 위해 각 속성들을 계산하는 식을 정의했다. 설명하기에 앞서 각 식에서 사용될 count라는 변수는 키넥트가 처음부터 실행되어 각 관절의 좌표를 한번 측정할 때 마다 1씩 증가하는 변수이다. 사용자의 걸음속도를 측정하기 위해 사용자가 이동한 걸음걸이에 측정시간으로 나누었으며 정의한 식(3.1)은 다음과 같다.

$$Speed = |(a-b)/(끝나는시간 - 시작시간)| \quad (3.1)$$

a = 측정시작점 spinebase.z
b = 측정끝점 spinebase.z

속성	의미	약어
Stride_diffDeviResult	보폭 차이	Stride
BalanceBodyResult	몸의 균형	SB
StepWidthResult	보행 간격	SW
SpeedResult	걸음 속도	Speed
Shoulder_SlopeAVG	어깨 기울기	SHD
O-leg	O자형 다리	Oleg
Arms_moving	팔의 움직임	AM

(그림 3) 연구에 최종적으로 적용된 속성 및 약어

아래의 식(3.2)은 선정한 7가지의 속성에 관한 식이며 측정을 통해 얻은 필요한 관절 좌표를 정의한 식에 적용하여 계산한다. 사용한 속성에 대한 약어는 (그림 3)에 있는 용어를 사용하였다.

$$SHD = \left(\sum (|(Shoulderdiff.y) \div (Shoulderdiff.x)|) \right) \div count$$

$$Stride = \left(\left(\sum |anklediff.z| \right) \div count \right) \div Speed$$

$$SB = \left(\sum |SHD - initSHD| \right) \div count$$

$$SW = \left(\sum |anklediff.x| \right) \div count$$

$$Oleg = \left(\sum \left(\left(|a \times knee.x - knee.y + b| \right) \div \left(\sqrt{a^2 + 1} \right) \right) \right) \div count$$

$$a = (Hip.y - ankle.y) / (Hip.x - ankle.x)$$

$$b = Hip.y - (a \times Hip.x)$$

$$AM = \left(\left(\sum (|Wristleft.y| + |Wristright.y|) \right) \div count \right) \div Speed \quad (3.2)$$

각 식(3.2)에는 관절의 오른쪽과 왼쪽좌표 차이, 기울기 등을 이용하여 식을 정의하였으며 보폭 차이나 팔의 움직임 같은 경우 걸음 속도에 영향을 받아 그 값으로 나누어 주어 좀 더 측정값의 정확도를 높이고자 하였다. 또한 O자형 다리 같은 경우는 엉덩이와 발목좌표를 이용하여 직선의 방정식을 통해 식을 세우고 무릎 좌표와의 거리를 계산하였다. 이러한 식(3.2)을 토대로 본 논문에서는 각 속성의 정상 범위에 만족하는지 판별하는 과정을 거쳤고 2개 이상의 속성에 대해 만족하지 못할 경우 허리에 문제가 발생할 가능성이 높은 것으로 판단하도록 구현 하였다.

4. 연구 결과

본 연구의 목표는 특정한 병을 판단하는 환자가 대상이 아닌 보통의 사람들이 키넥트를 통해 자신의 걸음걸이를 측정하며 앞으로 허리 부분에 문제가 발생할 가능성이 높은지를 측정을 통해 알려주는 것으로써 각 부분 중 어느 부분에 문제가 발생할지 구분하는 속성 값들을 설정하여 다음과 같은 결과들을 도출하였다.

각 속성 값들은 걸음걸이와 관련된 논문과 병원 자료들을 참고하여 걸음 속도, 보폭 차이, O자형 다리, 걸음걸이 습관, 팔의 움직임, 어깨의 수평정도, 몸의 안정도 7개로 설정하였다. 그리고 실제 측정 데이터를 이용한 연구에 앞서 7개의 속성과 허리 문제와의 연관 정도를 알기위해 50명 정도의 테스트 데이터(Test data)들을 만들어 진행하였다. 의사결정 나무를 수행한 결과 걸음걸이 습관이 가장 연관 정도가 높았고 팔의 움직임과 걸음 속도, 어깨의 수평 정도와 O자형 다리, 몸의 안정도 순으로 연관 정도가 높은 것으로 나왔으며 보폭 차이 속성은 상대적으로 연관 정도가 낮게 결과가 도출되었다. 이를 통해 위의 7가지 속성 중 6가지가 허리의 문제를 판별하는데 관계가 높다고 판단하였다. 하지만 이 중 걸음걸이 습관 속성은 측정자의 걸음걸이가 팔자걸음, 안짱걸음 등의 좋지 않은 걸음걸이를 가지고 있는지 판단하는 것으로 연관 정도가 높게 나왔지만 이를 판별해 내는 방법은 여러 가지 조건을 결합하여 측정 정확도를 높이고 조금 더 보완해야 할 것으로 보여 실제 측정을 통해 결과를 도출하는 과정에서는 해당 속성을 제외하였다. 또한 무릎과 연관 정도가 높은 몸의 안정도 대신에 허리와 연관 정도가 높은 보행 간격으로 대체하여 6가지 속성으로 결정하여 진행하였다.

다음으로는 키넥트로 측정하여 실제 데이터를 적용하여 연구를 진행하였다. 이어지는 실험에서는 키넥트를 통해 측정된 데이터와 데이터마이닝을 통해 분석된 결과를 토대로 세운식이 올바른지에 대한 검토를 먼저 진행하였다. 측정을 진행하면서 보폭 차이와 팔의 움직임은 사용자의 걸음속도에 영향을 받는다는 것을 알았고 걸음속도 값으로 나누어 사용자의 보폭 차이와 팔의 움직임의 측정값 정확도를 향상 시킬 수 있었다. 또한 키넥트는 발쪽의 데

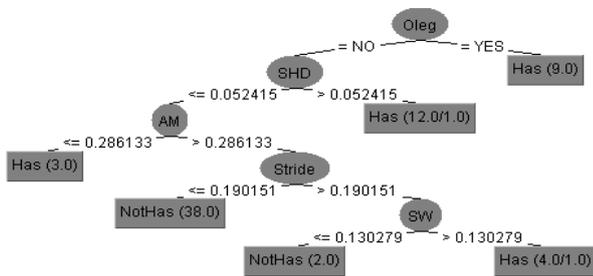
이더가 부정확한 현상이 있는 문제가 발견되었는데 그 원인은 키넥트는 적외선을 쏘아서 반사된 적외선을 인식하는 방식으로 구성 되어있어 발 쪽 부분은 바닥에서 난반사가 일어남으로써 제대로 반사되어 돌아오지 못하거나 경로가 틀려져 떨림 현상이 발생하게 된다는 것으로 결론이 도출되었다. 그래서 보행 간격 값이 비정상적으로 나올 경우, 무릎좌표를 이용하여 계산을 하는 식(3.3)을 추가하여 다음과 같이 정의하고 연구를 진행하였다.

$$SW = \left(\sum |kneedi\text{ff}.x| \right) \div \text{count} \quad (3.3)$$

User id	Stride	SW	Speed	SHD	Oleg	AM	Problem
1	0.179039	0.144674	0.139095	0.028095	NO	0.297553	NotHas
2	0.173387	0.200671	0.039057	0.034269	YES	0.295512	Has
3	0.1844	0.149252	0.100776	0.025982	NO	0.378528	NotHas
.....							
67	0.186091	0.170821	0.454499	0.045921	NO	0.361919	NotHas
68	0.184973	0.149303	0.489125	0.082164	NO	0.353913	NotHas
69	0.251688	0.233988	0.17295	0.092831	NO	1.437215	Has

(그림 4) 측정된 값들의 Data Table 표

69명의 걸음걸이를 키넥트로 측정하였고 실제 데이터를 통해 세운 식을 거쳐 나온 값들을 정리하였다. (그림 4)는 데이터베이스에 저장된 사용자의 측정된 값을 나타내며 User id는 측정된 사용자의 고유한 번호를 의미한다.



(그림 5) 측정된 값을 의사결정나무로 나타낸 사진

위의 (그림 4)의 측정된 자료들을 토대로 데이터 마이닝 한 결과 (그림 5)의 결과가 나왔으며 O자형 다리는 첫 번째로 허리에 문제와 연관 정도가 가장 높은 것으로 측정되었다. 두 번째로는 어깨 기울기와의 연관 정도가 높았으며 측정된 값이 0.052415를 초과할 경우 문제가 생길 수 있음이 도출되었고 세 번째 순서로 연관이 있는 팔의 움직임은 0.286133 이하 값인 경우에 문제가 생길 수 있다는 것으로 도출되었다. 네 번째로 연관이 높은 것은 보폭 차이였으며 보폭은 0.190151 이하의 값일 경우 문제가 생기지 않음이 도출되었다. 마지막 다섯 번째로는 보행 간격 값이 약 0.130279를 초과 할 경우 앞으로 허리에 문제가 될 수 있음이 도출되었다. 걸음 속도 속성은 허리과 관련된 문제에서 다른 속성들에 비해 관련된 정도가 상대적으로 낮아 위 (그림 5)에 나타나지 않았다.

5. 결론

본 연구에서는 키넥트를 통해 사용자의 걸음걸이를 측정하여 얻은 데이터를 바탕으로 생성한 공식을 통해 분석을 수행하고 그 결과를 기준 값과 비교하여 현재의 걸음걸이가 앞으로 허리에서 질병을 유발할 가능성이 있는지를 예측하는 연구를 진행하였다.

이를 위해 의사결정나무를 통해 설정한 속성 값들의 문제를 판별하기 위해 어느 정도 연관이 있는지를 분석한 결과 속성 값들 약 85.7%가 허리와 무릎의 문제를 판별하는데 연관이 있음을 확인하였다. 본 연구에서는 허리 부분을 우선적으로 하여 실제로 측정된 데이터들을 바탕으로 실험을 진행한 결과 6개의 속성 중 걸음 속도를 제외한 5개의 속성, 즉, 약 83%의 속성이 허리와 연관이 있음을 알 수 있었다.

이러한 연구를 실제 활용하기 위해 보완해야할 점은 팔의 움직임에서 팔의 전체적인 움직임을 감지할 수 있도록 회전, 3차원적인 움직임 등의 개념을 적용하는 식을 새롭게 도출하거나 보완해야 한다. 또한 상대적으로 허리와의 연관정도가 낮게 나온 걸음 속도 속성의 값이 음수 값으로 측정되는 현상을 보완 할 예정이며 발 부분의 좌표 측정 정확성을 향상시키기 위해 키넥트의 발 부분에서의 난반사의 문제를 최소화해야하므로 어떤 환경에서 난반사를 최소화 할 수 있을지에 대한 추가적인 연구가 진행될 필요가 있다. 마지막으로 현재 69명의 데이터를 통해 의사결정트리로 분석을 하고 연구를 진행하였지만 좀 더 정확한 분석과 식의 기준을 도출하기 위해서는 데이터가 적다고 판단되어 계속해서 측정해나가 사용자 데이터를 축적할 예정이다.

Acknowledgement

본 논문은 교육부가 지원하고 충북대학교가 수행하는 지역선도대학육성사업의 지원을 받아서 수행되었습니다.

참고문헌

[1] Virginia O. Andersson and Ricardo M. Araujo, Person Identification Using Anthropometric and Gait Data from Kinect Sensor, Twenty-Ninth AAAI Conference, 2015
 [2] 권혁수, 정상인들의 걸음형태에 따른 요통정도와 장애 지수와의 관련성 조사, 대한정형도수치료학회지, 2004
 [3] 신주영, 이해연, 김성연, 배병철, 조준동, 걸음걸이 관리를 위한 시각적 피드백 시스템 디자인, HCI 2015, 2014
 [4] Young-Jae Kim, Seung-Joo Ra, Ha-Yeon Jin, Se-In Jang, Chang-Soo Kim, Kwan-Hee Yoo, Nasridinov Aziz, A Kinect-Based Walking Assessment for Disease Prediction ,ICIARE 2016 ,2016