

위한 인공지능 교육 - 인공지능과 인간 -

LESSON 1. 튜링 테스트

LESSON 2. 규칙 기반과 학습 기반 시스템



1 튜링 테스트

1. 튜링 테스트의 이해

1) 튜링 테스트의 기원과 의의

인공지능은 기계가 인간과 같은 지능을 가지기 위해 모방하는 기술의 총집약체라 할 수 있습니다. 그렇다면 '지능'은 무엇일까요?

현대 컴퓨터의 아버지라 불리는 튜링이 1950년도에 '컴퓨팅 기계와 지능'(=Computing Machinery And Intelligence)이라는 논문을 발표하였습니다. 처음으로 튜링 테스트라는 단어가 등장한 논문이기도 합니다. 튜링 테스트는 인간의 지능과 동등하거나 구별할 수 없는 지능적인 행동을 기계가 보여줄 수 있는지 테스트하는 것입니다. 튜링 테스트를 이미테이션 게임이라고도 부릅니다.

① 튜링 테스트의 기원: 튜링은 **기계가 생각할 수 있는가?** 라는 질문을 합니다. 질문에서 '생각'이라는 단어가 굉장히 모호하기 때문에, '생각'이라는 단어를 바꾸어 다음과 같이 질문을 합니다.

'기계가 지능을 측정하는 행동 테스트를 통과할 수 있는가'

이 질문을 통해 명확한 단어로 기계가 인간의 지적능력을 가지고 있는지에 대한 판단기준을 제시하고자 했습니다. 50년이 지난 이 시점에서 바라보면 이론의 빈약함과 모순이 발견되지만 '지능'에 대해 심도 있는 논의를 시작하게 된 출발점이라는 데서 큰 의의를 찾을 수 있습니다. 튜링 테스트의 매력은 단순함에서 비롯됩니다. 심리학, 철학, 신경과학과 같은 복잡한 학문의 이론을 알지 않더라도, 단순한 질문과 응답을 바탕으로 지능에 대해 측정할 수 있는 기준을 제시하려고 노력하였습니다.

② 튜링 테스트 통과자

사실 튜링 테스트의 통과 기준은 없습니다. 튜링은 2000년대가 되면, 기계들이 5분간의 시험에서 인간 판사의 30%를 속일 수 있을 것이라고 주장합니다. 이러한 주장이 변형되어, 어느덧 튜링 테스트의 통과기준이 30%가 되어버렸다. 비판적인 시각을 가진 학자들은 튜링 테스트는 통과기준이 없는 주관적인 테스트라며 평가절하하기도 합니다.

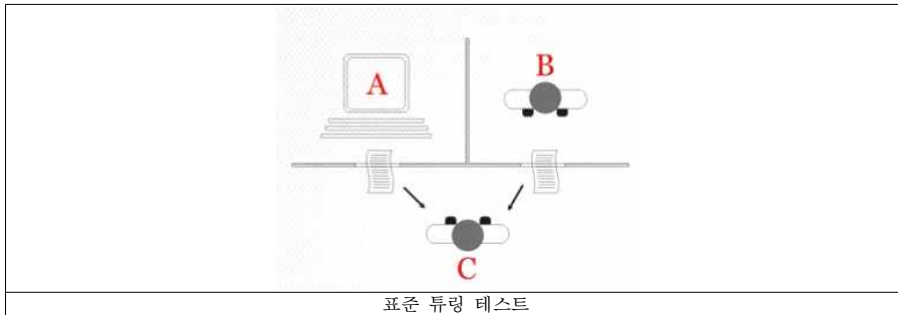
유진 구스트만은 튜링 테스트의 첫 통과자로 알려져 있습니다. 영국의 한 대학에서 개발한 유진 구스트만은 13세의 우크라이나 소년으로 묘사됩니다. 2012년에 첫 도전에서 실패했고, 2년 뒤 2014년 6월경 재도전을 통해 튜링 테스트 통과 기준인 30%를 넘어선 것으로 알려져 있습니다. 그러나 인공지능 전문가들은 유진 구스트만이 실제 튜링테스트에 통과했다고 인정하지 않는 분위기입니다. 테스트 결과 우크라이나에 산다고 대답해 놓고, 우크라이나에 가본 적이 있냐는 질문에는 가본 적이 없다고 대답하는 등 인간만큼의 지능을 가진 것으로 보기 어렵기 때문입니다.

2) 튜링 테스트의 진행과정

① 전체 진행과정(표준 튜링 테스트): 튜링 테스트의 진행방법은 인간과 기계를 다른 공간에 배치한 다음, 질문자(=판단자)가 그들과 텍스트로 대화를 나눕니다. 음성을 사용할 경우 기계나, 성별 등의 영향을 받을 수 있기에 텍스트를 활용합니다. 질문자는 어떤 공간에 누가 있는지 알 수 없습니다. 몇 번의 질문과 응답을 통해 질문자가 기계와 인간을 명확하게 구분하지 못한다면 기계가 지능을 가진 것으로 판단합니다. 통상 5번 정도씩 텍스트를 활용해 대화를 나누고, 약 5분 정도 진행한다고 되어 알려져 있

습니다.

아래 그림에서 C가 질문자(=판단자)가 되고, A가 인간이 되고 싶은 기계, B가 실제 인간입니다. 여기까지가 널리 알려진 기본적인 튜링 테스트 진행과정입니다.



② 세부 진행과정(모방게임)

모방게임은 튜링의 논문에 있는 진행방식입니다. 튜링은 이를 **튜링 테스트라 부르지 않고 '모방 게임(imitation game)'**이라고 불렀습니다. 모방게임은 총 3단계로 구성되어 있습니다.

1단계는 남자 1명, 여자 1명, 질문자 1명이 필요합니다. 테스트가 시작되기 전 한 공간에는 남자를 다른 공간에는 여자를 배치합니다. 질문자는 둘의 존재를 알 수 없는 상황에서 필기한 질문을 둘에게 전달하고 답변을 받습니다. 질문자는 둘 중 누가 남자인지 가려야 합니다. 단, 남자는 질문자를 속이기 위해 여자인 척을 하게 됩니다.

2단계에서 남자 대신 기계가 배치됩니다. 이번에는 기계가 여자인 것처럼 대화를 해야 합니다. 인간을 모방하는 것에서 더 나아가 성별을 모방할 수 있는지 테스트 하는 것입니다. 1단계와 동일하게 질문자는 둘의 존재를 알 수 없는 상황에서 필기한 질문을 둘에게 전달하고 답변을 받습니다. 질문자는 둘 중 누가 기계이고, 여자인지를 가려야 합니다. 기계는 단순히 사람을 모방하는 것에서 멈추지 않고, 사람처럼 실수도 할 수 있어야 하고, 애매한 답변이나 답변 유보 등을 할 수 있어야 합니다.

기계는 인간과 동일한 지능을 가지고 있다는 것을 증명하기 위해 질문자의 질문에 숨어있는 진짜 의미를 파악해야 합니다. 튜링 테스트의 질문과 답변 샘플에는 이런 내용들이 반영되어 있습니다. 너무 복잡한 수학적 문제를 빠른 시간 안에 풀지 못하는 인간적인 모습, 동시에 간단한 계산을 몇 초안에 대답을 할 수 있는 연산능력, 작품을 감상하거나 자신의 의견을 말하는 감정까지 모방합니다.

3단계는 여자를 다시 남자로 대체하고, 남자에게 여성인 척 흉내를 내게 합니다. 기계는 여성을 흉내 내고 있는 남성과 경쟁하여 더 여성인 척을 합니다. 이때 질문자는 어느 쪽이 여성인척 하는 남자이고, 어느 쪽이 여성인척 하는 기계인지를 맞춰야 합니다. 여성인척 하는 기계와 여성인척 하는 남성 사이에서 테스트가 진행됩니다.

2. 튜링 테스트의 한계와 활용

1) 튜링 테스트의 한계

① 진짜 인공지능인가?

튜링 테스트를 통과했다고 인간의 지능과 비슷한 지능을 가진 진짜 인공지능이라고 말할 수 있을까? 튜링 테스트는 인공지능 판단기준으로 사람과 자연스럽게 채팅할 수 있는 능력만을 꼽았기 때문입니다. 사람의 지능으로 할 수 있는 다양한 능력 중 단 하나인 채팅과 의미파악만을 기준으로 정합니다. 1950년대와 2020년도에 바라본 인공지능의 개념은 다를 수밖에 없다. 최근에는 진짜 인공지능이라면 종합적인 판단력을 가지고 있어야 한다고 이야기합니다. 기술이 발달함에 따라 인공지능에 대한 기대감이 올라가고 동시에 익숙해진 것들을 인공지능이라고 생각하지 않게 된다는 것입니다. 이를 AI효과라고도 합니다.

'AI effect'로 인해 1950년도의 튜링 테스트가 지금은 유용하지 않습니다. 단순 챗봇이나 기존의 SW방식으로도 튜링 테스트 통과를 위한 프로그램은 만들 수 있습니다. 게리 마커스(Gary Marcus) 뉴욕대 인지과학자는 기술의 발전과 AI의 활용확산에 발맞춰 튜링 테스트를 업그레이드 해야 한다고 주장합니다.

② 존설의 중국어방 논증: 캘리포니아 버클리 대학의 철학과 명예교수인 존설은 튜링 테스트를 비판하기 위해 다음과 같은 사고 실험을 제안합니다.

- 방안에 중국어를 전혀 할 줄 모르는 사람 A가 들어갑니다. 방 안에 한자 매뉴얼이 있어서 A는 중국어로 된 질문이 들어오면 한자 매뉴얼을 보고 적절한 답변을 해서 내보냅니다. 이런 방식으로 중국어를 전혀 모르는 사람도 매뉴얼을 참고하기만 하면 중국인과 질의응답을 할 수 있게 됩니다. 물론 A는 대화의 내용을 전혀 모릅니다. 여기서 중국어를 정말로 이해하고 있는 사람은 질문을 하는 중국인과 매뉴얼을 제작한 사람뿐입니다.

- 존설 교수는 다시 질문을 합니다.

"이 방 안에는 중국어를 할 줄 아는 사람이 없는데도 중국어 대화가 가능합니다. 즉, 중국어를 몰라도 중국어를 이해하는 것처럼 대화를 할 수 있는 것입니다. 이것을 튜링 테스트에 적용시켜봅시다. 단순히 사람과 대화를 할 수 있다고 해서 그 기계가 지능이 있다고 판단할 수 있습니까?"

- 존설의 중국어방 논증에 대한 비판: 인간의 뇌는 뉴런이라는 신경세포로 이루어져 있습니다. 인간이 언어를 학습할 때, 인간의 뉴런에서만 학습이 일어나기에 인간도 언어를 이해하는 것이 아니라 뉴런만 언어를 이해한다는 결론이 나옵니다.

③ 인간이 기계로 오해되는 경우: 튜링 테스트의 또 다른 한계는 바로 인간이 기계로 오해받는 경우입니다. 흥미로운 특징은 고전 튜링 테스트에서는 보통 기계와 인간을 1개체씩 놓고 테스트를 진행하는데 질문자(=판단자)에 의해 잘못 판단되는 경우입니다. 질문자들이 기대하는 응답이 반드시 전형적인 인간의 대답은 아닐 수 있습니다.

예를 들어 "핵무기에 대해 어떻게 생각하십니까?"라는 질문에 "핵무기는 매우 위험하므로 사용해서는 안된다"라고 대답한다면 기계로 오해받을 수 있다는 것입니다.

기계는 어떠한 질문에 대해 응답을 하지 않고, 계속적인 침묵이나 '음'과 같은 용어만 응답으로 한다면, 질문자가 오직 추측에 의해서만 인간/기계를 구분해야 한다는 문제점도 있습니다. 의도적으로 응답을 회피하는 기계와 실제로 정답을 말할 수 없어서 침묵하는 인간을 판단자가 정확히 구분할 수 있을까?

이렇게 컴퓨터의 지능이 아니라 질문자의 태도, 추측, 기술 등에 의해 결과가 달라질 수 있습니다.

2) 튜링 테스트의 활용

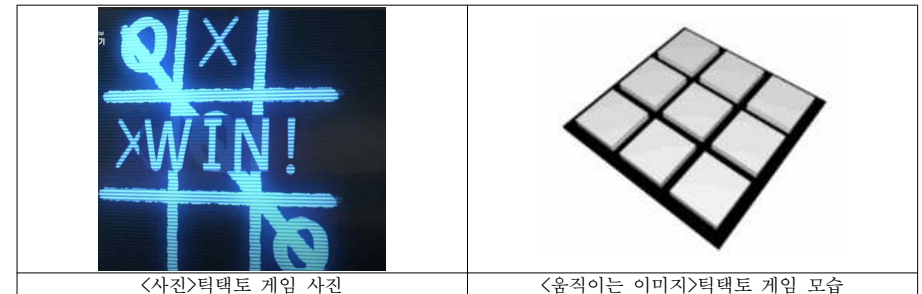
- CAPTCHA: 인간이 기계와 인간을 구분하는 것이 튜링 테스트라면, 역튜링 테스트는 반대로 기계가 인간과 기계를 구분하는 형태입니다. 튜링 테스트가 변형된 것입니다. 웹사이트에서 특정 작업(예시-홈페이지 가입, 비밀번호 변경 등)을 수행하기 전 사용자에게 왜곡된 그래픽 이미지를 제시되고, 해당 문자나 숫자를 입력하라고 요청합니다. 사람인지 기계인지를 컴퓨터가 판단하는 것입니다. 아래 그림은 학교 현장에서 종종 발생하는 문제로, 구글 클래스룸에 가입 후 비밀번호를 변경할 것을 요구받는 CAPTCHA 사진입니다. 이미지 왜곡이 심하여 인간인 학생들도 종종 이 테스트를 한 번에 통과하지 못하곤 합니다.

구글 클래스룸 가입 후 비밀번호 변경 요구 사진

2 규칙기반과 학습기반 시스템

1. 틱택토(=TIC TAC TOE) 게임

틱택토 게임은 같은 모양으로 세 줄을 먼저 채우면 이기는 게임으로 최초로 컴퓨터에 구현된 게임으로 알려져 있습니다. 서양에서 하던 삼목으로 두 명이 번갈아가며 O와 X를 3×3 판에 표시하며 같은 글자를 가로, 세로, 혹은 대각선상에 놓이도록 하는 놀이입니다. 틱택토 게임에서는 **선공을 잡은 사람의 승리 확률이 매우 높습니다.** 그렇지만 선공만 이기는 것은 아닙니다. 틱택토와 같이 특정 크기의 바둑판같은 모양에서 두 명이 번갈아 가면서 게임을 하는 경우 한 가지 특징이 있습니다. 자신이 놓은 돌이 적어도 게임의 승률을 떨어뜨리지 않는 다는 것입니다. 그렇기에 선공과 후공이 각각의 상황에 맞는 최선의 선택만을 한다고 가정할 경우, 선공이 이기거나 비기는 결과만 존재하게 됩니다.



1) 틱택토 게임이란?

① 게임 방식 및 게임결과: 기본규칙은 다음과 같습니다.

첫째, 상대방이 놓은 돌 위에는 자신의 돌을 놓을 수 없습니다.

둘째, 상대방과 번갈아 가면서 돌을 놓게 됩니다.

셋째, 게임의 결과는 크게 3가지로 구분합니다. 승리, 패배, 무승부입니다.

② 게임의 진행: 게임을 진행하는 칸이 3×3칸이기 때문에 선공을 잡은 사람은 최대 5번까지 돌을 놓을 수 있고, 후공을 잡은 사람은 최대 4번까지만 돌을 놓습니다. 연속된 3개의 돌을 놓은 사람이 승리하게 되고, 후공을 잡은 사람이 마지막 9번째(선공 5번, 후공 4번의 합) 돌을 두었을 때도 연속된 3개의 돌이 나오지 않으면 무승부로 끝나게 됩니다.

③ 틱택토 게임을 온라인으로 진행하려면 검색어로“틱택토”를 입력한 다음, 게임을 진행합니다.

틱택토 게임의 난이도를 변경해볼 수 있습니다. 초급, 중급, 고급 단계로 1번씩 게임을 진행합니다. 게임을 진행하면서 게임의 필수전략을 생각해 봅니다.

필수전략을 빨리 파악하기 위해 난이도를 고급으로 변경한 다음, 컴퓨터가 선공일 경우에 놓는 첫수를 잘 살펴보고 바랍시다. 일정한 패턴을 발견할 수 있을 것입니다. 컴퓨터가 선공일 때, 항상 모서리에 첫수를 놓습니다. 고급 버전에서는 100번, 1000번을 진행해도 컴퓨터가 놓는 첫 수가 일정합니다. 반대로,

고급버전에서 플레이어가 선공이 되어 모서리에 돌을 놓는 경우에도, 후공인 컴퓨터가 돌을 놓는 위치가 항상 동일하다는 것을 발견할 수 있습니다. 고급버전의 컴퓨터는 잠시 뒤 살펴볼 규칙기반에 의해 만들어진 필승전략을 사용합니다.

2) 선공이 되어 진행하기

① 선공이 되어 틱택토 게임을 진행합니다. **틱택토 게임에서는 'X'가 선공으로 표시** 됩니다.

아래표에서 왼쪽 상단을 기준으로, 숫자를 지정하여 첫 수를 놓는 상황을 제한합니다.

1	2	3
4	5	6
7	8	9

② **진행1:** 선공이 되어서 **첫 수를 반드시 중앙인 5번 자리에 두고 난 다음** 게임을 진행합니다. 최소 2차례 이상 게임을 진행해야 합니다. 첫수를 5번 위치에 두고 나면 컴퓨터는 대부분 3번의 자리에 돌을 놓습니다. 그 다음, X돌을 둘 수 있는 경우의 수는 크게 2가지 경우입니다. 모서리 또는 모서리가 아닌 나머지 자리입니다. 이런 방식으로 3번째 수, 4번째 수, 5번째 수를 모두 골고루 체험해 봅니다.

③ **진행2:** 선공이 되어서 **첫 수를 반드시 모서리에 두고 난 다음** 게임을 진행합니다. 표에서 모서리의 자리는 1, 3, 7, 9번입니다. 예를 들어 7번 자리에 첫수를 두었다면, 컴퓨터는 대부분 5의 자리에 돌을 놓습니다. 그 다음, X돌을 둘 수 있는 경우의 수는 크게 2가지 경우입니다. 모서리 또는 모서리가 아닌 나머지 자리입니다. 이런방식으로 3번째 수, 4번째 수, 5번째 수를 모두 골고루 체험해 봅니다.

④ **진행3:** 선공이 되어서 **첫 수를 반드시 모서리나 중앙이 아닌 나머지 부분에 두고** 난 다음 게임을 진행합니다. 표에서 2, 4, 6, 8번이다. 예를 들어 8번 자리에 첫수를 두었다면, 컴퓨터는 대부분 2의 자리에 돌을 놓습니다. 그 다음, 내가 둘 수 있는 수를 구분해보면 크게 3가지 경우입니다. 중앙, 모서리 또는 모서리가 아닌 나머지 자리입니다. 이런 방식으로 3번째 수, 4번째 수, 5번째 수를 모두 골고루 체험해 봅니다.

3) 후공이 되어 진행하기

① 후공일 경우는 선공인 경우보다 경우의 수가 더 적습니다. 대부분 선공이 중앙 또는 모서리에 돌을 놓기 때문입니다.

② 선공이 중앙에 돌을 놓는 경우 대응해봅시다. 이 때, 후공은 모서리에 돌을 놓는 경우와 모서리가 아닌 자리에 돌을 놓는 경우로 나뉩니다. 이런 방식으로 3번째, 4번째 수를 골고루 봅니다.

③ 선공이 모서리에 돌을 놓는 경우 대응해봅시다. 모서리에 돌을 놓는 경우 내가 중앙에 돌을 놓는 경우, 다른 나머지 3가지 모서리 중 하나에 돌을 놓는 경우, 중앙에 돌을 놓는 경우로 나뉩니다. 이런 방식으로 3번째, 4번째 수를 골고루 체험해 봅니다.

2. 규칙기반과 학습기반을 통한 틱택토의 필승전략

1) 규칙기반을 통한 필승 전략 수립

필승전략 수립을 위해서는 **돌을 둘 때 우선순위를** 정해야 합니다. 1순위는 즉시 경기를 승리할 수 있는

위치에 돌을 두는 것이 됩니다. 2순위는 지금 막지 않으면 바로 패배하는 경우입니다.

① 선공의 경우(규칙기반의 알고리즘 생성)

가. 첫 수로 가장 유리한 수는 모서리에 두는 것입니다. 수비하는 상대방이 모서리나, 귀(=모서리와 모서리 사이)에 두면 무조건 이기게 됩니다. 수비하는 상대방이 말을 놓는 경우가 8가지인데, 중앙을 제외한 7가지의 경우는 필승입니다. 따라서 7/8의 확률을 얻습니다.

(보충 설명: 수비하는 상대방이 첫수로 중앙에 돌을 놓지 않으면 필승입니다.)

선공이 모서리에 놓았을 경우 승리 확률을 모두 더하면 7/8보다 높습니다. 이유는 수비하는 상대방에 중앙에 놓았을 경우에도 승리하는 경우가 있기 때문입니다.

만약, 상대방이 중앙에 놓는 경우 벌어지는 상황을 살펴볼까요?

선공이 이어지는 대각선에 있는 모서리에 놓으면 최소 1/3의 확률로 이기게 됩니다. 선공이 승리할 확률은 7/8 + 1/8*3/1이 됩니다.

$$\text{리 확률} = \frac{7}{8} + \frac{1}{8} * \frac{1}{3} = \frac{11}{12}$$

나. 선공이 첫 수를 중앙에 놓으면 대략 1/2확률로 이기고, 1/2확률로 비기게 됩니다. 확률의 계산은 선공이 중앙에 놓았을 경우 상대방이 모서리에 놓지 않으면 이기게 된다고 볼 수 있습니다.

다. 선공일 경우의 규칙을 정리하면 다음과 같습니다.

1. 모서리에 돌을 놓습니다. 2. if: 상대방이 중앙에 돌을 놓지 않았으면 대각 반대쪽 모서리로 이동합니다. else: 상대방이 중앙에 돌을 놓았다면, 아무 모서리로 이동합니다. 3. if: 두 개의 X와 줄에 공백이 있으면 그 공간으로 이동합니다. else if: 두 개의 O가 있고, 그 사이에 공백이 있으면, 그 공백으로 이동합니다. else: 아무 모서리로 이동합니다. 4. if: 두 개의 X와 줄에 공백이 있으면 그 공간으로 이동합니다. else if: 두 개의 O가 있고, 그 사이에 공백이 있으면, 그 공백으로 이동합니다. else: 아무 모서리로 이동합니다. 5. 빈 공간에 둡니다.
<틱택토 게임 선공일 경우 필승 전략>

② 후공의 경우: 후공의 경우는 선공이 최선의 선택만 한다는 가정하에는 이기는 것이 불가능합니다. 따라서 비기는 것을 최우선으로 전략을 수립해야 합니다. 상대방이 최선의 선택을 하지 않는다면 승리할 수도 있습니다.

2) 학습기반을 통한 필승 전략 수립

규칙기반을 통한 틱택토 게임의 필승 전략을 인공지능이라 할 수 있을까? **인공지능은 규칙기반의 전문가 시스템에서 발전**해 왔다. 규칙기반은 전문가들이 만들어놓은 알고리즘에 따라 정교하게 작동하기 때문에 인간이 보기에 인공지능처럼 보일 수도 있습니다.

학습기반이기 때문에 틱택토게임에서 승리한 상황(=경우)을 학습데이터로 사용해야 합니다. **규칙기반과의 가장 큰 차이점**이라고도 할 수 있습니다. 머신러닝 방식은 사람이 필승 전략을 따로 알려주지 않습니다.

① 컴퓨터와 사람이 게임을 진행하면서 학습합니다. 사람은 게임 방법, 필승 알고리즘을 컴퓨터에게

명령어로 알려주지 않습니다. 대신, 게임을 플레이하는 사례를 보여주고, 승리한 경우의 데이터를 활용해 컴퓨터를 학습시킵니다. 학습데이터가 늘어날수록 컴퓨터의 틱택토 승리확률이 올라가게 됩니다.

그림은 머신러닝 포 키즈 인공지능 틱택토 프로그램을 활용한 사진입니다. 학습을 시작하기 전 학습 공간(레이블 속의 학습 데이터)이 비어 있는 것을 확인할 수 있습니다.

② 프로그램을 실행시켜 컴퓨터와 사람 간 틱택토 게임을 진행합니다. 이 프로그램에서 컴퓨터에게 틱택토 게임 규칙을 설명하지 않습니다. 컴퓨터에게 연속된 3개의 줄을 이어야 한다고 가르치지 않습니다. 프로그램 속에서 게임 규칙으로만 설정되어 있습니다. 컴퓨터는 프로그램을 진행하면서 플레이어(=사람)가 이겼을 경우의 돌의 위치, 컴퓨터가 이겼을 경우의 돌의 위치만을 학습합니다.

③ 학습이 진행되면 학습데이터가 점차적으로 수집하게 됩니다. 학습데이터를 수집한 후, 학습을 진행하게 되면 모델이 생성됩니다. 생성된 모델은 규칙을 가르쳐주지 않더라도 게임에서 승리할 수 있게 됩니다.

④ 학습이 완료된 인공지능 모델을 개략적으로 확인해보겠습니다. 모델 생성을 위해 의사결정트리가 사용됩니다. 의사결정트리는 데이터를 분석하여 이들 사이에 존재하는 패턴을 예측 가능한 규칙들의 조합으로 나타내며, 그 모양이 '나무'와 같다고 해서 의사결정나무라 불립니다. 질문을 던져서 대상을 좁혀나가는 '스무고개' 놀이와 비슷한 개념입니다.

이상으로 인공지능을 활용해 틱택토 필승 전략을 만드는 과정을 살펴보았습니다. 의사결정 트리의 구조를 하나씩 살펴보기 힘들뿐더러, 인간이 이해하기 쉽지 않습니다. 인공지능이 이렇게 하나의 모델을 생성하게 되면, 해당 내용을 인간이 잘 이해할 수 없기에 인공지능이 만든 알고리즘이나 모델구조를 블랙박스라고 부르기도 합니다.