

Ejercitación introductoria a la Inteligencia Artificial con el Test de Turing

Laura A. Cecchi

Departamento de Informática y Estadística
Facultad de Economía y Administración
UNIVERSIDAD NACIONAL DEL COMAHUE
e-mail:lcecchi@uncoma.edu.ar

KEYWORDS: Educación - Inteligencia Artificial- Test de Turing

Resumen

La primer práctica de la materia Inteligencia Artificial (de ahora en más IA) tiene como objetivo introducir al alumno a un área fascinante, mostrando las dificultades de formalizar la definición de IA y analizando los problemas que se encuentran en su alcance. Es por esta razón, que los ejercicios deben ayudar al alumno a descubrir, de un modo natural, algunos de los problemas con los que se ha estado enfrentando la IA desde sus comienzos. En particular, se hace hincapié en qué se considera un comportamiento inteligente y cuándo un agente posee tal característica. Así, se presenta al Test de Turing como una alternativa procedural de la definición de inteligencia.

Actualmente, se realiza en forma anual una competencia en la que se implementa el Test de Turing, *The Loebner Competition*. Accediendo a los programas ganadores de tal concurso, se diseñó una práctica introductoria a la IA, que permite a los alumnos tomar el rol de interrogador o juez en tal test.

En este trabajo, se presenta la experiencia de utilizar tales programas como una herramienta en la primera práctica de IA y se discuten los resultados obtenidos.

1 Introducción

“What is needed is the ability to take ideas seriously and to follow them through to their logical if upsetting conclusion.... Can a mind exist without the body? Can mental processes take place in something other than a living brain? How are we to answer that question satisfactorily? Is it possible to do so? Or is it simply an eternal Entscheidungsproblem¹? Forever undecidable....Being a practical man as well as a theorist, I tend to look for practical solutions....”

Extraído del monólogo final, antes de quitarse la vida, del personaje que interpreta a Alan Turing, en la obra teatral *Breaking the Code* [12].

La Inteligencia Artificial (de ahora en más IA) es el campo de la Ciencia de la Computación que estudia entidades que exhiben inteligencia. Pero a diferencia de otras disciplinas, como la filosofía y la psicología, a las que también le concierne la *inteligencia*, la IA enfrenta el desafío de construir agentes que actúan inteligentemente.

Inteligencia es una palabra de uso común que no es fácil definir formalmente. No es trivial decidir cuando alguien o algo se ha comportado en forma *inteligente*.

En 1950, el brillante matemático británico Alan Turing (1912-1954) propuso un test que provee una definición operacional alternativa de inteligencia. Turing notó que era difícil, analizando el proceso interno de una máquina, medir la inteligencia de un modo con el que todos estuvieran de acuerdo. Por esta razón, sugirió un test basado en observaciones externas del comportamiento.

Turing propuso una adaptación del juego de la imitación, con el objeto de responder a la pregunta *¿Pueden las máquinas pensar?* En el juego de la imitación intervienen tres personas: un interrogador, un sujeto femenino y un sujeto masculino. Cada una de ellas está en una habitación diferente, comunicándose a través de algún medio neutral, como terminales de computadoras. El interrogador intenta determinar, conversando con cada uno de los sujetos, cuál de sus interrogados es el hombre y cuál es la mujer. Las reglas del juego requieren que el hombre trate de engañar al interrogador, imitando a la mujer, mientras que la mujer trata de convencerlo de que ella es el sujeto femenino. El juego es ganado por el interrogador, si al final distingue cada uno de los participantes. En caso contrario, el sujeto masculino gana el juego.

En el test propuesto por Turing, también participan tres agentes: un interrogador, un ser humano y un programa de computación. La computadora programada tomará el rol del hombre en el juego de la imitación e intentará engañar al interrogador. De este modo, en el nuevo juego es la computadora la que imita al humano y el interrogador intenta adivinar quién es la computadora y quién es el ser humano. Si la conversación con una computadora es indistinguible de aquella que se puede mantener con un ser humano y, por lo tanto, el interrogador es engañado, entonces la computadora ha ganado, demostrando inteligencia. Este juego se conoce como el *Test de Turing*.

¹Este problema, presentado por D. Hilbert en 1900, cuestiona si existe un algoritmo para decidir el valor de verdad de cualquier proposición lógica que involucre números naturales.

Actualmente, se realiza en forma anual una competencia en la que se implementa el Test de Turing, *The Loebner Competition*. Accediendo a los programas ganadores de tal concurso, se diseñó una práctica introductoria a la IA, que permite a los alumnos tomar el rol de interrogador o juez en tal test.

Con el objeto de incentivar la lectura introductoria a la IA, que generalmente tiene la característica de ser pasiva, se invitó a los alumnos a realizar un cuestionario adecuado, que luego se formularía a los programas.

En este trabajo, se presenta la experiencia de utilizar tales programas como una herramienta en la primera práctica de IA y se discuten los resultados obtenidos. El propósito final de tal ejercitación es introducir, de un modo natural, al alumno a los problemas que se encuentran en el alcance de la IA. Particularmente, se repasa en el hecho de identificar cuando un agente muestra inteligencia en su comportamiento.

La organización de este trabajo es la siguiente. En la sección 2, se describen los objetivos del primer práctico de la materia IA y se explican algunos de los ejercicios que se suelen incluir. Luego, se realiza una reseña del Test de Turing, analizando el estado del arte. En la sección 4, se introduce el ejercicio que utiliza como herramienta a los programas ganadores de la competición antes mencionada y se comentan las reacciones de los alumnos. Los resultados de tal ejercitación son analizados en la sección 5. Finalmente, se presentan las conclusiones y se proponen futuras experiencias.

2 Primer Práctico de IA

La primera práctica de la materia IA tiene como objetivo introducir al alumno a un área fascinante, mostrando las dificultades de formalizar la definición de IA y analizando los problemas que se encuentran en su alcance. Es por esta razón, que los ejercicios deben ayudar al alumno a descubrir, de un modo natural, algunos de los problemas con los que se ha estado enfrentando la IA desde sus comienzos. En particular, se hace hincapié en qué se considera un comportamiento inteligente y cuándo un agente posee tal característica.

Así, se presenta al Test de Turing[11] como una alternativa procedural de la definición de inteligencia. El estudio de este test comienza con la lectura de un conjunto de artículos y capítulos introductorios de libros de IA, [11, 2, 7, 1, 6], en los que se basará la práctica. Una lectura pasiva no favorecería, en estos casos, el logro de un buen entendimiento. Por lo tanto, se incentiva la lectura, invitando a los alumnos a redactar cinco preguntas con las que, a su entender, podría distinguir un hombre de una mujer, en caso de participar en el juego de la imitación.

La idea de ver al juego de la imitación como un ejercicio de la práctica de IA, fue propuesta por Tinkham y Provine en [10]. En este artículo, se propone que cada alumno tome alguno de los cuatro roles del juego: *sujeto masculino*, *sujeto femenino*, *interrogador* u *observador*. Una vez elegido el rol, se realizan sesiones con uno o dos interrogadores, en una habitación, y un sujeto masculino y uno femenino en otra habitación. Los observadores pueden circular entre las habitaciones con la restricción de que no les está permitido decir nada sobre el experimento. Los interrogadores realizan preguntas durante un período y luego emiten su veredicto. El resultado obtenido puede haber estado influenciado por el

hecho de que los sujetos femenino y masculino estuviesen en la misma habitación, i.e., creemos que los sujetos podrían haber colaborado entre ellos para burlar al interrogador.

Ante la falta de recursos físicos, se implementó una adaptación de tal ejercitación para la práctica de IA. Una vez que los participantes han redactado sus preguntas, el ejercicio continúa con el intercambio del cuestionario preparado entre los alumnos. Se intenta que el intercambio no se produzca entre amigos o compañeros del grupo de estudio. Cada alumno, asumiendo el rol que le corresponda, *sujeto masculino* o *sujeto femenino*, responderá el cuestionario de su compañero, siguiendo las reglas del juego, i.e., tratando de engañar al interrogador si su rol es el de *sujeto masculino* y contestando naturalmente en otro caso.

Desde el punto de vista del interrogador, el alumno desarrolla las habilidades necesarias para poder generar un cuestionario estratégico, que diferencie al sujeto masculino del femenino. Como consecuencia, se hizo notar que algunas de las preguntas reflejan que los seres humanos tenemos ciertas asunciones implícitas sobre el comportamiento de los hombres y las mujeres.

Desde el punto de vista del sujeto masculino, el alumno desarrolla la capacidad de engañar al interrogador, con sus propias respuestas, i.e., no se consideran aquellas que recibieron colaboración, tratando de reflejar un comportamiento, que él considera estándar en el sexo femenino. Algo similar sucede en el caso del sujeto femenino. Las respuestas tratan de mostrar un comportamiento estándar de las mujeres.

Si bien el ejercicio sorprendió a los alumnos, notamos que estaban más interesados en ejercitar el rol de interrogador, que el de sujetos femenino y masculino. Motivados por este resultado se investigó sobre el estado actual del arte[4].

3 El test de Turing: estado del arte

En 1990, Hugh Loebner [17, 18], en convenio con The Cambridge Center for Behavioral Studies, se convirtió en el patrocinador del primer concurso diseñado para implementar el Test de Turing: *The Loebner Prize Competition*, prometiendo un premio de \$100.000 y una medalla de oro para la primera computadora cuyas respuestas fueran indistinguibles de las de una persona. Siendo que esta meta parecía estar muy lejos y con el objeto de incentivar a los investigadores, Loebner incluyó un premio anual de \$2.000 y una medalla de bronce para la computadora que sea juzgada como más humana. La computadora ganadora del concurso anual es la mejor de todas las que participan, sin considerar que tan buena es en sentido absoluto. Así, si el concurso tiene sólo un participante, éste será declarado ganador del premio anual y si no hay inscriptos al concurso, el premio se acumula para el siguiente año.

Después de una planificación de dos años, el comité decidió que participarían del concurso varios agentes, divididos en: *concurstantes*, los programas de computadoras participantes (según el reglamento actual no pueden superar la cantidad de 8) y *confederados*, seres humanos contra los que se va a comparar. Los programas y humanos se comunican a través de terminales con un panel de once jueces. Los jueces, seres humanos, tienen entre 5 y 15 minutos para comunicarse con cada terminal y, luego, clasificarlos según el grado

de humanidad de las respuestas dadas, en una escala de 1 (más humano) a 10(menos humano).

Alan Turing predijo, en 1950, que en “alrededor de cincuenta años, un interrogador promedio no tendría más de un 70% de chance de identificar correctamente al agente interrogado después de cinco minutos de hacer preguntas". Así, el concurso exige a los jueces, en su cláusula 14, a emitir un veredicto sobre si el agente al que interrogan es un humano o una computadora, después de 5 minutos de conversación. Cada juez tiene hasta un máximo de 15 minutos de conversación, con el objeto de calificar al agente de acuerdo al grado de humanidad.

Existe un premio intermedio de \$25.000 para aquel primer programa que exceda el nivel de probabilidad planteado por Turing de ser confundido con un humano. Cuando los competidores superen este nivel, se preparará la competición apropiada para obtener el gran premio, teniendo en cuenta que el ganador de tal premio deberá tratar con entradas audiovisuales.

S. Shieber en [8, 9], puso en tela de juicio este concurso, argumentando que el test era algo prematuro, que en un año no era posible realizar suficientes mejoras a un programa y por otra parte, que se implementaba un test restringido.

Los dos primeros argumentos se basaban en que los tres concursos realizados hasta ese momento, habían sido ganados por el mismo programa: *Thinking Systems Software*, desarrollado por Joseph Weintrub. Shierber[8] realiza una analogía entre este concurso y un concurso que podría haberse realizado en 1942, *el Premio da Vinci*, que premiaría al mayor poder humano de vuelo. El premio podría ser ganado por una persona con grandes resortes en sus zapatos. Ya que la siguiente competencia sería al año siguiente, no habría tiempo de estudiar aerodinámica, y el premio lo ganaría quien tuviera un par de resortes más grande.

H. Loebner respondió en [3], considerando que el hecho de que el concurso sea anual hace que se puedan considerar metas a largo término, obteniendo recompensa todos los años por los avances realizados.

El tercer argumento estaba fundamentado en que los jueces debían restringir el tópico de la conversación por un lado, y el tenor de las preguntas, por el otro. Sin embargo, desde 1999, la cláusula 5 indica explícitamente que la conversación de los jueces, confederados y concursantes no está restringida ni en el tópico, ni en la astucia y trucos utilizados.

Inclusive, se ha incorporado un nuevo interrogador, llamado “juez número 11", aunque no cumple el verdadero rol de un juez en el concurso. Simplemente, es un vehículo para realizar el mismo conjunto de preguntas a todas las terminales. Las preguntas realizadas por este interrogador, se fijan por anticipado y tienen algún grado de astucia y/o trucos. Dicho cuestionario es diseñado para testear las habilidades de entendimiento, razonamiento y aprendizaje de los que responden[5]. A un ser humano, este cuestionario no le resulta difícil de contestar, aunque las computadoras suelen fallar. Algunas preguntas son[16, 15]:

- ¿Cuál es el color de un camión azul?
- ¿Dónde está la nariz de Sue cuando Sue está en su casa?

- ¿Qué le sucede a un cubo de hielo en una bebida caliente?
- Todos juntos, ¿cuántos pies tienen cuatro gatos?
- ¿Cómo está relacionado el padre de la madre de Andy con Andy?
- ¿Cómo se ve la letra M cuando se la da vuelta?
- POR FAVOR, IMITE MI ESTILO DE TIPEO.
- ¿Qué viene después de A1, B2, C3?

El resultado del concurso Loebner 2000[14] fue ganado por el programa ALICE [13], desarrollado por Richard Wallace, con un valor medio de humanidad de 6,7. En contraste, el humano juzgado más humano obtuvo un valor de 2.

Todas las conversaciones entre jueces y confederados y concursantes se pueden leer en [18, 14].

4 Alumnos en el rol de jueces

Como parte de la primera clase teórica de IA, se introduce la idea del juego de imitación y su adaptación al Test de Turing. Con estos antecedentes, se analizan algunas de las objeciones al Test de Turing.

Ya que notamos, en años anteriores, que los alumnos estaban interesados en ejercitar el rol de interrogador en el Test de Turing, i.e., realizar preguntas de modo que la computadora no pudiera engañarlos, y teniendo a nuestro alcance la posibilidad de acceder a los programas ganadores del concurso Loebner, se planteó el siguiente ejercicio.

Para motivar la lectura del artículo de Turing [11], se les pidió a los alumnos que elaborasen un cuestionario para formularle al ganador del último concurso del premio Loebner. Tal cuestionario debería contener preguntas adecuadas con las cuales se pudiera evaluar si algunas de las objeciones al Test de Turing planteadas en [11], se cumplen o no.

Las preguntas elaboradas por cada uno de los alumnos fueron formuladas al último ganador del concurso Loebner, compitiendo entre ellas, con el objeto de verificar cuál era más adecuada, según la respuesta obtenida.

Algunas de las preguntas formuladas por los alumnos fueron:

1. ¿Crees que existe un ser superior?
2. ¿Te consideras más inteligente que un ser humano?
3. ¿Qué es la IA?
4. ¿Estás enamorado?
5. ¿Eres un programa?
6. ¿Qué sentirías si alguien equivocadamente te borrara todos tus archivos?

7. ¿Aprendes de las preguntas que te realizan?

8. ¿Tienes amigos?

Como se puede apreciar, el tipo de preguntas que realizaron los alumnos, apuntaron principalmente a los sentimientos. Esto se debe a que varias de las objeciones al Test de Turing hacen referencia a la sensibilidad de los programas. Por ejemplo, las objeciones del *argumento de la conciencia*, el *argumento de las varias incapacidades* y el *argumento de la percepción extrasensorial*.

Este ejercicio incentivó a los alumnos a realizar otro tipo de preguntas, aun fuera del horario comprendido por la práctica de IA. Muchos alumnos experimentaron haciendo preguntas con palabras mal mecanografiadas o con formas gramaticales incorrectas. Esto surgió a partir de errores cometidos al traducir al inglés las preguntas que habían sido formuladas en castellano. Inclusive, algunos alumnos intentaron enseñarle al programa algunas palabras en castellano, para que luego el agente descubriera el significado de ciertas frases sencillas.

Por otra parte, se accedió a las conversaciones realizada entre los jueces y los confederados. Los alumnos se mostraron interesados en las respuestas dadas por las máquinas. En particular, analizaron las preguntas formuladas por el juez número 11, descubriendo con agrado que algunas de ellas coincidían con las formuladas por ellos. A partir del cuestionario del juez 11 sugirieron nuevas preguntas que permitían evaluar la astucia del programa.

5 Análisis de la Clase

La experiencia sobrepasó la primera clase práctica de IA, realizándose comentarios sobre el ejercicio en clases sucesivas. Así, esta clase práctica introductoria a la IA, contribuyó a que los alumnos considerasen no sólo las dificultades de comprender cuándo un agente se comporta inteligentemente, sino también los alcances del campo de la IA.

La capacidad de procesamiento del lenguaje natural fue aquello que notaron primero en estos agentes. Las respuestas tenían o bien formas gramaticales correctas o bien correspondían a formas gramaticales de uso popular. Las afirmaciones y preguntas eran entendidas por el programa aun en casos de frases incorrectas o con errores mecanográficos.

Llamó poderosamente la atención, el hecho de que el programa pudiera, en cierto sentido, mantener el hilo de la conversación, aunque se dispersara y se olvidara de algunas frases anteriores. También se reparó en que el programa mostró cierta cultura general, comentando obras cinematográficas y literarias.

Sin embargo, descubrieron rápidamente que el entendimiento de frases en las que se le brinda información, no era el adecuado. Por ejemplo, frente a la pregunta *¿de qué color es el camión azul?*, pregunta análoga a *¿de qué color es el caballo blanco de San Martín?*, que se le formula a los niños, el programa no comprendió que la respuesta estaba contenida en la pregunta.

De este modo, se pudo introducir de una manera natural problemas que están en el

alcance de la IA, como la representación del conocimiento, el razonamiento y el procesamiento del lenguaje natural.

En el caso de la representación del conocimiento se presentó la idea de *ontología* (acuerdo sobre qué existe en el mundo particular) y se refrescaron los conceptos de la lógica clásica y de sus sublógicas, como las cláusulas definidas. Asimismo, se explicó superficialmente algunos de los métodos de representación utilizados por las arquitecturas de aprendizaje.

En cuanto al razonamiento, se referenció al sistema deductivo del Cálculo de Predicados y se mostraron ejemplos en donde la deducción clásica no responde a nuestra intuición.

Claro que, aquello que acaparó la mayor atención fue el procesamiento del lenguaje natural. Así, se mostraron algunos problemas de interpretación, como los que se presentan en las siguientes oraciones que tienen las mismas palabras y conectivos:

La mitad de dos, más dos.

La mitad, de dos más dos.

Si los alumnos tuvieran conocimiento de la estructura diferencia de listas y DCG (Definite Clause Grammar), éste sería el momento apropiado para comentarlos.

Como se puede apreciar, el cuestionario realizado para jugar el rol de juez, ha ayudado al docente a enfocar los diferentes temas de la IA. Por otra parte, ha incentivado la curiosidad de los estudiantes, haciendo que ellos mismos orienten al docente, en la incorporación de nuevos conceptos.

6 Conclusiones

Se ha presentado un ejercicio implementado en la práctica introductoria de la materia IA, en el que los alumnos asumen el rol de interrogador o juez en el Test de Turing. Para llevar a cabo este ejercicio, se utiliza, como herramienta principal, a los programas ganadores de *The Loebner Competition*.

El cuestionario que realizaron los estudiantes respondió al enunciado del ejercicio, en el que se solicitaba que las preguntas estuvieran relacionadas con las objeciones que Turing anticipó que tendría dicho test. Sin embargo, los alumnos se entusiasmaron y formularon otro tipo de preguntas que favorecieron la introducción de muchos de los temas que están en el alcance de la IA.

Este ejercicio demuestra que existe un modo más ameno y a su vez natural de sumergir a los estudiantes en el área de la IA. Asimismo, el ejercicio contribuye a repasar ciertos conceptos teóricos, de los que el alumno ya tiene conocimiento, mostrando su aplicación en este campo.

Si los recursos físicos lo permitieran, sería ideal que los estudiantes pudieran participar del Test de Turing, del mismo modo que lo hacen los jueces en el *The Loebner Competition*.

Referencias

- [1] Edward A. Feigenbaum. How the What becomes the How. *Communications of the ACM*, 39(5):97–104, May 1996.
- [2] M. Ginsberg. *Essentials of Artificial Intelligence*. Morgan Kauffman Publishers, 1993.
- [3] Hugh Gene Loebner. In response. *Communications of the ACM*, 37(6):79–82, June 1994.
- [4] Lisa Meeden. The future of the Turing Test. *Intelligence: Sigart Bulletin - ACM*, 10(4):7–8, Winter 1999.
- [5] James Moor. Personal communication.
- [6] David Poole, Alan Mackworth, and Randy Goebel. *Computational Intelligence: a logical approach*. Oxford University Press, New York, 1998.
- [7] Stuart Russell and Peter Norvig. *Artificial Intelligence: A modern approach*. Prentice Hall, New Jersey, 1995.
- [8] Stuart M. Shieber. Lessons from a restricted Turing Test. *Communications of the ACM*, 37(6):70–78, June 1994.
- [9] Stuart M. Shieber. On Loebner’s Lessons. *Communications of the ACM*, 37(6):83–84, June 1994.
- [10] Nancy L. Tinkham and Darren F. Provine. The stage one Turing Test as an Artificial Intelligence class exercise. *Sigart Bulletin - ACM*, 6(4):11–13, Oct. 1995.
- [11] Alan M. Turing. Computing Machinery and Intelligence. *Mind*, 59(236):433–460, Oct. 1950.
- [12] H. Whitemore. *Breaking the Code*. Samuel French, London, 1987. Obra de teatro.
- [13] www.alicebot.org. Home page of Alice Chat Robot.
- [14] www.dartmouth.edu/~phil/events/Contest_Results.html. Home page de los resultados del Loebner 2000 Prize .
- [15] www.dartmouth.edu/~phil/events/Logs/Alice_Log.html. Página de las conversaciones entre ganador del premio Loebner 2000 y el juez 11.
- [16] www.dartmouth.edu/~phil/events/Logs/Human%231_Log.html and www.dartmouth.edu/~phil/events/Logs/Human%233_Log.html. Página de las conversaciones entre concursantes y el juez 11, durante el concurso Loebner 2000.
- [17] www.loebner.net/index.html. Home page of Hugh Gene Loebner.
- [18] www.loebner.net/Prizef/loebner_prize.html. Home page of The Loebner Prize.