

# *Propuesta de Evaluación basada en Big Data para facilitar la integración de Agentes Conversacionales Pedagógicos en las aulas*

**Silvia Tamayo-Moreno, Diana Pérez-Marín**

Escuela Técnica Superior de Ingeniería Informática, Universidad Rey Juan Carlos  
Avda. Tulipán s/n, Móstoles, Madrid, España  
diana.perez@urjc.es

**Resumen:** Se propone una evaluación basada en Big Data de la interacción entre agentes conversacionales pedagógicos y estudiantes para facilitar la integración de este recurso TIC en las aulas. Así, se espera conseguir que la sociedad pueda aprovecharse de sus beneficios.

**Palabras clave:** Agente Conversacional Pedagógico, Big Data, Interacción Persona-Ordenador, Educación Infantil.

**Abstract:** An evaluation based on Big Data of the interaction between Conversational Pedagogic Agents and students is proposed to make the integration of the agents into the classroom. The goal is to provide teachers with another ICT resource to take advantage of its advantages.

**Keywords:** Pedagogic Conversational Agent, Big Data, Human-Computer Interaction, Pre-Primary Education

## **1. Introducción**

La tecnología se encuentra presente en diversos ámbitos de la sociedad, aumentando su importancia y su uso en las diferentes áreas, entre ellas la educación [1]. En este contexto, es de vital importancia la confluencia de sistemas y técnicas informáticas y su aplicación a las prácticas y metodologías educativas [2].

Los Agentes Pedagógicos Conversacionales son sistemas interactivos que permiten a los estudiantes repasar de una forma entretenida y amigable según un diálogo en lenguaje natural [3]. En la actualidad, existen gran cantidad de agentes distintos [4] aplicados en dominios muy diversos desde la enseñanza de Sistemas Operativos en la Universidad [5], Ciencias Naturales en Educación Primaria [6] o competencias como la capacidad de contar cuentos [7] o establecer empatía con otras culturas [8].

Los resultados conseguidos en niveles universitarios son prometedores, con mejoras de hasta 0.8 en la puntuación final del examen de los estudiantes que repasaron con un agente según los experimentos realizados [5]. En el caso de Educación Secundaria y Primaria aunque no siempre se reportan mejoran significativas en la puntuación de la evaluación final, se reportan beneficios como el efecto Persona [6] según el cual la mera presencia del agente en el entorno educativo puede tener un efecto positivo en la percepción de la experiencia educativa por parte del estudiante; el efecto Proteo [9] según el cual los estudiantes pueden aprender motivados por conseguir las características de sus agentes y parecerse a ellos; y, el efecto Protégé [10] según el cual los estudiantes pueden hacer un esfuerzo mayor para enseñar a su agente que para aprender ellos mismos.

Además, también está adquiriendo especial relevancia en el ámbito tecnológico el Big Data y las posibilidades que ofrece, entre ellas, capacidades analíticas. Una de las definiciones de Big Data es la propuesta por Beyer & Laney [11] que la define como “un alto volumen, alta velocidad y/o amplia variedad de activos de información que requieren nuevas formas de procesamiento para permitir la toma de decisiones mejorada, el descubrimiento de conocimiento y la optimización de procesos”. El panorama de la tecnología está cambiando rápidamente en los últimos años, tanto es así que como resultado de ello, algunos límites entre áreas y prácticas se están empezando a difuminar, a lo que se añade la aparición de nuevas categorías. En este sentido, es importante el conocimiento de las mismas y cómo se relacionan entre sí. De esta manera, Business Intelligent (BI), utilizado por el investigador de IBM Hans Peter Luhn en 1958, definiéndolo como “La capacidad de comprender las interrelaciones de los hechos presentados de tal forma como para orientar la acción hacia una meta deseada”, se podría considerar como la toma de decisiones basada en datos, abarcando los procesos y procedimientos de recolección de datos, intercambio, presentación de informes, aportación de información... incluyendo la generación, agregación, análisis y visualización de datos para facilitar la toma de decisiones, gestión y estrategia. Como analítica, se entienden las diferentes maneras de desglosar los datos, evaluación de tendencias en el tiempo, comparativas, las diferentes formas en las que el dato se puede visualizar para detectar las tendencias y relaciones de un vistazo. El Big Data es la tecnología que almacena y procesa datos (en general gran cantidad) de fuentes internas y externas, que requieren un gran poder de computación para recopilar y analizar.

La minería de datos o data mining hace referencia a la búsqueda de información relevante y apropiada en grandes conjuntos de datos y su transformación en una estructura comprensible. La minería de datos ha ganado popularidad en el campo de las bases de datos, donde destaca el concepto de “knowledge discovery in databases” (KDD), acuñado en 1989 por Piatetsky-Shapiro [12], que se refiere al proceso global de descubrimiento de conocimiento útil a partir de datos. Otra disciplina relevante en el área del análisis de datos es el Machine learning, aprendizaje

automático o de máquinas, que consiste en el desarrollo de técnicas que permitan aprender a las máquinas [13].

En el ámbito educativo cada vez es más patente la necesidad de aprovechar las posibilidades que ofrece la tecnología, y por tanto, de nuevos métodos sustentados en ella para poder hacer el seguimiento de los alumnos, mejorar tutorías, métodos de enseñanza, obtener datos objetivos de sus evaluaciones, aprendizaje y comprensión de los alumnos o predecir los riesgos académicos. En este sentido, el uso de Big Data en la educación y la analítica que ofrecen juega un papel fundamental, traducándose en mejoras de gestión educativa, desarrollo de nuevos métodos para la enseñanza y el aprendizaje, creación de nuevas carreras y opciones profesionales, o explotación y aprovechamiento de recursos generados. Además, en la actualidad, a partir de la analítica Big Data y su integración con dispositivos inteligentes, tecnologías, entre ellas web, se derivan una serie de métodos que ya se están aplicando en el ámbito educativo como son el aprendizaje adaptativo [14], educación basada en competencias [15] o aula invertida y aprendizaje combinado (Flipped Classroom y Blended Learning) [16].

Aunque el uso de big data y las posibilidades analíticas que ofrece son ampliamente utilizadas y en constante crecimiento, en la revisión de literatura realizada no se encuentran ejemplos del empleo de las mismas aplicadas a agentes conversacionales pedagógicos que aún no están integrados en las aulas, a pesar de los beneficios reportados. Por lo tanto, se pretende investigar la viabilidad de la aplicación de este tipo de técnicas de análisis de datos a la interacción entre los estudiantes y los agentes para facilitar su integración en las aulas tanto de Primaria como de Infantil. El objetivo, por lo tanto, es doble: 1) realizar una propuesta de uso de las técnicas de Big Data para analizar la interacción entre los estudiantes y los agentes, y 2) facilitar la integración de los agentes en las aulas de Educación Primaria e Infantil.

El artículo está organizado en cinco secciones: la Sección 2 contiene la revisión del estado del arte; la Sección 3 centra la propuesta de evaluación; y, la Sección 4 concluye el artículo con las principales conclusiones obtenidas.

## 2. Estado del arte

### 2.1. Agentes Conversacionales

Los agentes conversacionales se definen como guías, maestros, ayudantes, la personalidad de la máquina o de un programa específico [3]. En interfaces conversacionales, el uso que se le da es la búsqueda de una interacción más personalizada entre la máquina y el usuario [17]. Entre las características de los agentes está la autonomía, ya que deben ser capaces de responder acorde a los conocimientos que les permitan conseguir el objetivo planteado; proactividad, tomando la iniciativa mediante sugerencias o acciones que les permitan alcanzar su objetivo; sociabilidad puesto que deben ser comunicativos, cooperativos y negociación; y tienen que tener una personalidad, un comportamiento de acuerdo al objetivo que persiguen; reactividad ya que el contexto en el que operan es percibido por ellos, y reaccionan en consecuencia; y persistencia, ejecutándose el código continuamente y decidiendo en qué momento debería realizar alguna acción.

Entre las ventajas del modelado de agentes se encuentra la cercanía a nuestra forma de ver el mundo, ya que muchas organizaciones están caracterizadas por un conjunto de roles y relaciones entre ellos, no obstante, se necesitan métodos y herramientas para el desarrollo de sistemas que se basan en agentes. Más concretamente, los agentes animados o conversacionales, son entidades que se pueden representar con personas, animales, cosas que pueden hablar con sonido o texto, e incluso tratan de mostrar y gestionar emociones [18].

Existen infinidad de agentes, de diversos tipos y utilidades [19, 5, 20]. Se reportan aquí algunos ejemplos para ilustrar el concepto, por su relevancia en el sector educativo, seleccionados según su rol de profesor cuyo objetivo es enseñar a los estudiantes; otros roles que no se contemplan en este estudio son el rol de estudiante que aprende; o rol de compañero que proporcionan soporte efectivo.

Los ejemplos de agentes seleccionados con el rol de profesor y destacables por alguna característica especial son: Autotutor, Laura, y Dr. Roland. Se describen en los siguientes párrafos.

- Autotutor [5], se basa en teorías constructivistas, teniendo iniciativa mixta de dialogo y animación 3D siendo un referente desde los años 90. En su funcionamiento utiliza Análisis de la Semántica Oculta combinada con técnicas de procesamiento de lenguaje natural [21]. Destacando que, supone una mejora en los resultados finales de los exámenes de quienes lo utilizan para repasar. Actualmente se trabaja en la mejora de la capacidad del agente para la detección de emociones y responder empáticamente al estudiante [22]. La Figura 1 muestra un ejemplo de su interfaz.



**Figura 1.** Ejemplo de pantalla de Autotutor

- Laura [19], se trata de un agente Web inteligente, se ha empleado para alumnos de un curso de aprendizaje de lengua española y fue diseñado para hacer conversaciones con estudiantes mediante texto y diálogos para proporcionar una herramienta que sirva para aprender nuevo vocabulario y practicar aspectos de conversación en español. Además, cuenta con imágenes animadas de lo que representa y voz para pronunciar palabras. La Figura 2 muestra un ejemplo de su interfaz.

- El agente Dr. Roland [23], se desarrolló con el objetivo de ayudar a los estudiantes a comprender, y con ello, mejorar su capacidad de resolver problemas de matemáticas. Desde el principio se desarrolló teniendo en cuenta las necesidades y requisitos que pedían profesores y estudiantes. El diálogo del agente sigue un algoritmo que escoge cada ejercicio según su tipo y su nivel de dificultad, adecuándose a cada estudiante, y registrando los avances de cada estudiante, comprobando además si el estudiante

comprende el ejercicio y prestando ayuda a medida que la va necesitando. La Figura 3 muestra un ejemplo de su interfaz.



Figura 2. Ejemplo de pantalla de Laura



Figura 3. Ejemplo de pantalla de Dr. Roland para Educación Infantil

## 2.2. Análisis Big Data

El término Big Data originalmente hace referencia al “volumen, velocidad y variedad de los conjuntos de datos” según Doug Laney en un informe de investigación de META Group, actualmente Gartner [24]. Con el tiempo, el término ha ido ampliándose existiendo diversas definiciones para el mismo. Cabe destacar las publicadas en el blog del departamento de Data Science de Berkeley con el objetivo de dar respuesta a la pregunta “¿Qué es big data?” [25]. En 2012, Gartner modificó su definición inicial de Big Data como “alto volumen, alta velocidad y/o amplia variedad de activos de información que requieren nuevas formas de procesamiento para permitir la toma de decisiones mejorada, el descubrimiento de conocimiento y la optimización de procesos” [11]. Se puede considerar que el Big Data es la tecnología que almacena y procesa datos (en general gran cantidad)

de fuentes internas y externas, que requieren un gran poder de computación para recopilar y analizar. Los datos, debido a su diversa procedencia, en muchas ocasiones requieren un proceso de limpieza para poder extraer información útil de ellos.

Business Intelligent (BI) en el glosario IT de Gartner es definido como “un paraguas que incluye aplicaciones, infraestructura y herramientas y las mejores prácticas que hacen posible el acceso y el análisis de información para mejorar y optimizar decisiones y rendimiento” [26]. De tal manera, que puede entenderse como una categoría amplia que, al menos en el contexto de negocios IT, abarca los conceptos de analítica, Big data y Data Mining. Se podría considerar como la toma de decisiones basada en datos, abarcando los procesos y procedimientos de recolección de datos, intercambio, presentación de informes, aportación de información... incluyendo la generación, agregación, análisis y visualización de datos para facilitar la toma de decisiones, gestión y estrategia.

Como analítica, se entienden las diferentes maneras de desglosar los datos, evaluación de tendencias en el tiempo, comparativas, las diferentes formas en las que el dato se puede visualizar para detectar las tendencias y relaciones de un vistazo. Se pueden plantear además cuestiones acerca del futuro con sistemas que realizan análisis predictivo. Hay quienes tratan el análisis y el BI como sinónimos, sin embargo, el análisis hace referencia a la explotación de los datos, a la fase de preguntas y respuestas que conduce a la fase de toma de decisiones en el proceso global de BI.

La minería de datos (data mining) hace referencia a la búsqueda de información relevante y apropiada en grandes conjuntos de datos y su transformación en una estructura comprensible. Es el proceso que trata de identificar patrones en grandes conjuntos de datos, siendo un campo de la estadística y las ciencias de la computación [27]. La minería de datos ha ganado popularidad en el campo de las bases de datos, y es en este punto donde cabe destacar el concepto de “knowledge discovery in databases” (KDD) que, acuñado en 1989 [12], hace hincapié en que el conocimiento es el producto final de un descubrimiento impulsado por los datos. De esta manera, KDD hace referencia al proceso global de

descubrimiento de conocimiento útil a partir de datos, mientras que minería de datos se refiere a un determinado paso en este proceso, más concretamente, es la aplicación de algoritmos específicos para la extracción de los patrones de datos.

Los pasos adicionales en el proceso de KDD son esenciales para asegurar el conocimiento útil derivado de los datos, ya que la aplicación a ciegas de métodos de minería de datos puede ser una actividad peligrosa, que desemboque en el descubrimiento de patrones sin sentido y no válidos [28]. En 1996, Brachman & Anand proporcionaron una visión práctica del proceso KDD, enfatizando la naturaleza interactiva del mismo, que en términos generales es el siguiente:

- 1) Comprensión del dominio de aplicación, el conocimiento previo relevante e identificación del objetivo del proceso.
- 2) Selección del conjunto de datos.
- 3) Limpieza de datos y preprocesamiento.
- 4) Reducción y proyección de los datos.
- 5) Establecer una sintonía en los objetivos del proceso KDD (paso 1) con un particular método de data-mining.
- 6) Análisis explicativo y selección de hipótesis y modelo.
- 7) Data mining.
- 8) Interpretación de patrones, y quizá sea necesaria una vuelta a los pasos 1 a 7 para una iteración completa.
- 9) Actuar sobre el conocimiento.

Las técnicas de minería de datos son algoritmos con diferente nivel de sofisticación que se aplican sobre un conjunto de datos para obtener resultados, y proceden de la inteligencia artificial y de la estadística. Algunas técnicas importantes, que se clasifican en función del objetivo de análisis de los datos en supervisados (o predictivos, a partir de datos conocidos predicen un dato o un conjunto de datos desconocidos previamente) y no supervisados (o del descubrimiento del conocimiento, descubren tendencias y patrones en los datos) son las siguientes [29]: redes neuronales [30], árboles de decisión [31], y clustering [32]. Se puede afirmar que data mining se alimenta de la estadística cogiendo técnicas como análisis de varianza, regresión, prueba chi-cuadrado, clustering, análisis discriminante, o series de tiempo.

Así como de la informática, con algoritmos genéticos, inteligencia artificial, sistemas inteligentes y expertos, o redes neuronales.

Otra disciplina relevante en el área del análisis de datos es Machine learning, aprendizaje automático o de máquinas [13], rama de la inteligencia artificial que consiste en el desarrollo de técnicas que permitan aprender a las máquinas, construyendo programas con la capacidad de generalizar comportamientos a partir de información no estructurada proporcionada en forma de ejemplos, es un proceso de inducción al conocimiento.

Aunque a veces el campo de actuación del machine learning y la estadística coinciden al basarse ambas en el análisis de datos, el primero está más centrado en la complejidad computacional de los problemas, pudiendo considerarse como un intento de automatización de partes del método científico mediante modelos matemáticos. El resultado de machine learning es un modelo para la resolución de una determinada tarea, entre los modelos, cabe destacar: geométricos, probabilísticos y lógicos [33]. Los modelos también se agrupan como modelos de agrupamiento (intentan dividir el espacio de instancias en grupos) y de gradiente (representan un gradiente en el que se puede diferenciar cada instancia). Existen diferentes tipos de algoritmos que se agrupan en una taxonomía teniendo en cuenta la salida de los mismos [33], algunos son: aprendizaje supervisado, no supervisado, semisupervisado, por refuerzo, transducción y multitarea, explicados en el siguiente apartado.

Existen diferentes enfoques en el aprendizaje automático [29]: árboles de decisión, reglas de asociación, algoritmos genéticos, redes neuronales artificiales, máquinas de vectores de soporte, el análisis por agrupamiento o clustering, o redes bayesianas. Con el aprendizaje automático se pueden obtener tres tipos de conocimiento: crecimiento (adquirido de lo que nos rodea), reestructuración (con la interpretación de los conocimientos, el individuo razona y genera nuevo conocimiento al cual se le llama de reestructuración), y ajuste (obtenido de generalizar varios conceptos o generando los propios).

En cuanto a la aplicación del Big data a la educación, es un recurso clave para analizar, visualizar, entender y mejorar la educación. El método “tradicional” de observación en el aula va perdiendo valor en cuanto a la forma más efectiva para mejorar el proceso educativo, entrando en juego otras opciones como por ejemplo la analítica del big data, siendo un recurso importante para entender y mejorar el proceso educativo [34]. Su uso ha empezado a dar beneficios en diferentes aspectos, desde mejora de gestión educativa, desarrollo de nuevos métodos para la enseñanza y el aprendizaje, creación de nuevas carreras y opciones profesionales, o explotación y aprovechamiento de recursos generados. Además, en la actualidad, en el ámbito educativo se hace cada vez más patente la necesidad de aprovechar las posibilidades que ofrece la tecnología, y por tanto, de nuevos métodos sustentados en ella para poder hacer el seguimiento de los alumnos, mejorar sus tutorías, métodos de enseñanza, obtener datos objetivos de sus evaluaciones, aprendizaje y comprensión de los alumnos o predecir los riesgos académicos. En definitiva, facilitar la analítica del aprendizaje con la tecnología de big data e identificar nuevas áreas de oportunidad para mejorar la educación. A partir de la analítica Big Data y su integración con dispositivos inteligentes y tecnologías, se derivan una serie de métodos que ya se están aplicando en el ámbito educativo: aprendizaje adaptativo, educación basada en competencias, aula invertida y aprendizaje combinado, gamificación y aprendizaje móvil [35].

La analítica del aprendizaje en el ámbito educativo consiste en la medición, recopilación, análisis y presentación de datos sobre los estudiantes, sus contextos y las interacciones generadas, con el fin de comprender el proceso de aprendizaje que se está desarrollando y optimizar los entornos en los que se produce. En este caso, el lago de datos para el análisis se forma a partir de acciones explícitas de los estudiantes y otras actividades que no son parte del trabajo del alumno y que pueden englobar los sitios web que visitan, su ubicación, las obras que consultan, su actividad social en línea, etcétera. Una vez extraídos los datos, éstos se analizan, relacionan y evalúan con el objetivo de poder enfocar la enseñanza a las capacidades específicas de cada estudiante.

### **3. Propuesta de evaluación basada en Big Data**

Se recomienda basar la evaluación de cualquier software interactivo educativo en técnicas de indagación y test, siempre con usuarios, puesto que en este caso las heurísticas no suelen ser suficiente [36]. En particular, se aconseja el uso combinado de técnicas de observación directa, con el uso de encuestas tanto a estudiantes como a profesores, registro de comportamiento del usuario, la interacción, las respuestas y el diálogo que ha tenido lugar durante la interacción, para su posterior análisis con técnicas como KDD [12], y en su paso 7 emplear el algoritmo que mejor se ajuste a aquello que se quiera obtener para extraer patrones de los datos (data mining). Los pasos, siguiendo la aproximación práctica del proceso propuesta por Brachman & Anand [37]:

- 1) Comprensión del dominio de aplicación, el conocimiento previo relevante e identificación del objetivo del proceso. En este punto es fundamental contar o haber contado con personas con experiencia en el área de aplicación del agente conversacional, y de todas las áreas implicadas en el desarrollo, y si es posible involucrar en el proceso (formando parte del mismo), con el objetivo de una mejor comprensión del contexto, de los factores que pueden afectarle, y mejor interpretación de los resultados que pretenden obtenerse.
- 2) Selección del conjunto de datos, identificación de variables objetivo que se quiere predecir, calcular o inferir y de variables independientes útiles para hacer el cálculo o proceso, muestreo de los registros disponibles,...teniendo en cuenta el agente conversacional pedagógico y sus características, así como las de las personas con las que interactúa y el contexto en el que lo hace.
- 3) Limpieza de datos y pre procesamiento. Esta etapa es muy importante, aunque a veces no se le otorgue la consideración que debiera, especialmente si el proceso de recolección de los datos no es controlado. Es importante que los datos tengan valores con sentido, y tomar decisiones respecto al ruido (valores faltantes, atípicos, incorrectos,...) pues de lo

contrario su introducción en un algoritmo de data mining lleva a dificultar el proceso de aprendizaje o que los resultados se alejen del comportamiento real.

- 4) Reducción y proyección de los datos. El objetivo es localizar las características más significativas para representar los datos, en función del objetivo del proceso. Para ello, pueden utilizarse procesos de transformación con el objetivo de reducir el número efectivo de variables o encontrar otras representaciones de los datos.
- 5) Establecer una sintonía en los objetivos del proceso KDD (paso 1) con un particular método de data-mining.
- 6) Análisis explicativo y selección de hipótesis y modelo.
- 7) Data mining, mediante técnicas de minería de datos, en general cada una obliga a un preprocesado diferente de los datos, se obtienen modelos de conocimiento, que representan patrones de comportamiento observados en los valores de las variables del problema o relaciones entre variables. En este paso se realiza la búsqueda de patrones en una determinada forma de representación o sobre un conjunto de representaciones.
- 8) Interpretación de patrones, y quizá una vuelta a los pasos 1 a 7 para una iteración completa si se considera necesario, el proceso se puede retroalimentar, repitiéndose desde el principio si el modelo final no supera la evaluación, o a partir de cualquiera de los pasos si un experto lo considera oportuno todas las veces que se crea necesario hasta la obtención de un modelo válido.
- 9) Actuar sobre el conocimiento descubierto, el modelo está listo para su explotación cuando se valida considerándose aceptable con salidas adecuadas y/o márgenes de error admisibles.

La Figura 4 muestra un ejemplo gráfico de este proceso. Un aspecto fundamental que puede ser incorporado y de gran utilidad para los agentes conversacionales pedagógicos, es el desarrollo de técnicas que les permitan aprender, siendo capaces de generalizar comportamientos (machine learning). Para ello, un proceso que puede seguirse es el que a continuación se presenta, es importante destacar que

se parte de la base de que ya se cuenta con un primer conjunto de datos sobre los que hacer el análisis resultado de la interacción del agente conversacional pedagógico con los estudiantes de su área de aplicación.

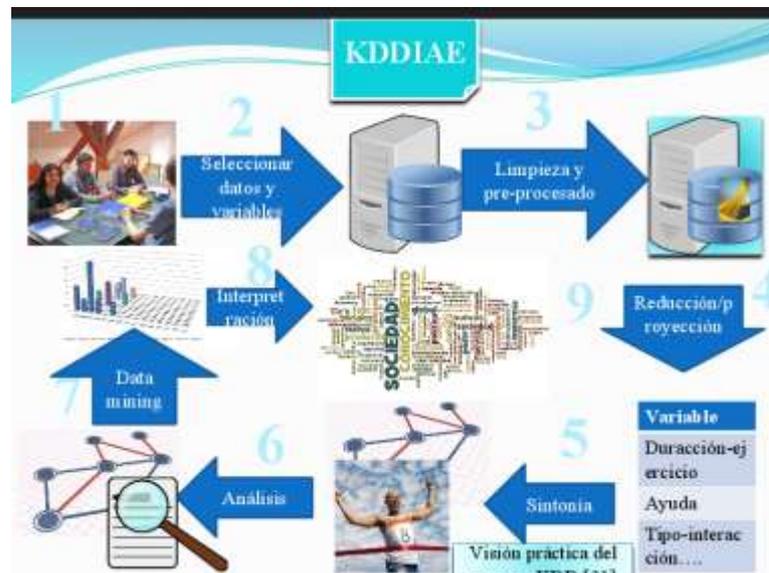
- 1) Identificación de la tipología del agente conversacional de que se trata, con el objetivo de determinar qué técnicas pueden ser mejores para los resultados que quieren obtenerse. Se puede usar cualquier taxonomía existente para ello, como puede ser la propuesta en Pérez-Marín [38] que los clasifica en función de 10 criterios. O según el rol que tenga el agente conversacional pedagógico, como profesores que enseñan, estudiantes que aprenden o compañeros que proporcionan soporte efectivo

- 2) Establecer qué tipo de salida quiere obtenerse, para seleccionar el grupo de algoritmos que mejor se adapte. Se puede seleccionar uno de los siguientes tipos de algoritmos [33] que se agrupan en una taxonomía teniendo en cuenta la salida de los mismos:

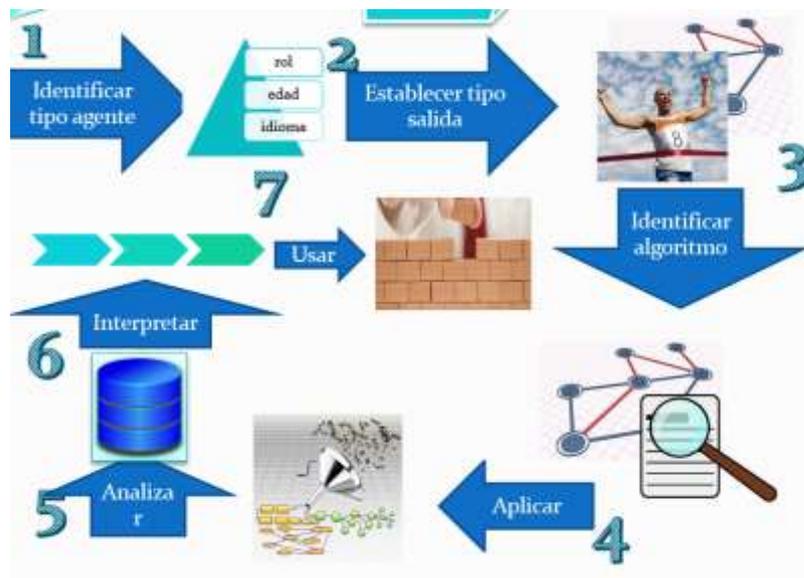
- i. Aprendizaje supervisado, si el algoritmo produce una función que establece correspondencias entre las entradas y salidas del sistema
- ii. Aprendizaje no supervisado, si el proceso de modelado se realiza sobre el conjunto de ejemplos formado por las entradas del sistema, sin información de las categorías de esos ejemplos, de tal manera que el sistema tiene que tener la capacidad de reconocer patrones para etiquetar las nuevas entradas.
- iii. Aprendizaje semisupervisado, combina los aprendizajes anteriores para clasificar adecuadamente, se consideran datos marcados y no marcados
- iv. Aprendizaje por refuerzo, se aprende observando al mundo que le rodea, siendo la información de entrada el feedback del mundo exterior como respuesta a sus acciones, aprendiendo del ensayo-error
- v. Aprendizaje transducción, que es similar al supervisado pero no construye explícitamente una función,

- basándose en los ejemplos de entrada, intenta predecir las categorías de los futuro
- vi. Aprendizaje multitarea, usa el conocimiento aprendido previamente por el sistema para enfrentarse a problemas parecidos a los ya vistos.
- 3) Identificación del algoritmo o algoritmos a aplicar, pertenecientes al grupo de algoritmos identificado en la fase previa.
  - 4) Aplicación de los algoritmos a los datos de que se disponen, en función del algoritmo de que se trate, se aplicará el procedimiento correspondiente para su aplicación.
  - 5) Análisis de los resultados obtenidos
  - 6) Si los resultados son los esperados, seguir trabajando en cómo seguir potenciado y usando esa información, retroalimentando al agente. Si no son los esperados, analizar qué puede estar fallando centrándose en las siguientes líneas:
    - i. Si los datos que tengo son suficientes y adecuados para obtener el resultado que se pretende
    - ii. Si los datos sobre los que se hace el análisis no son correctos o son insuficientes, analizar de nuevo qué datos necesito para obtener el resultado que pretendo.
    - iii. Identificar los datos a capturar
    - iv. Analizar qué modificaciones hay que hacer en el agente conversacional para capturar esos datos
    - v. Modificar el agente para corregir la captura de los datos erróneos y/o capturar los nuevos que son necesarios.
    - vi. Si se está aplicando el conjunto de algoritmos adecuado
  - 7) Si no es así, analizar qué conjunto de algoritmos, por sus características y las del agente conversacional se adapta mejor
  - 8) ii. Seleccionar ese conjunto de algoritmos para realizar el análisis
    - i. Si el algoritmo o algoritmos concretos dentro de ese grupo seleccionado previamente es el adecuado
      - ii. Si no es así, analizar qué algoritmo podría ser el adecuado de acuerdo a sus características
      - iii. Seleccionar un algoritmo
      - iv. Aplicar dicho algoritmo
      - v. Si se está aplicando adecuadamente el algoritmo
  - 9) Si no es así, estudiar cómo se aplica el algoritmo
  - 10) ii. Repetir el proceso y aplicar el algoritmo.
  - 11) iii. Aplicar dicho algoritmo
  - 12) Uso de resultado
    - i. Si los resultados son los esperados y se quiere continuar mejorando el proceso de aprendizaje, se puede seguir trabajando de la siguiente forma.
    - ii. Analizar qué conocimiento adicional o nuevo quiere incluir
    - iii. Identificar qué datos se necesitan para ello
    - iv. Analizar el agente y sus características para identificar qué modificaciones se le podrían hacer para adaptarlo al objetivo que se pretende
    - v. Realizar las modificaciones en el agente
    - vi. Repetir la fase experimental para la recogida de datos
    - vii. Cuando se disponga de datos suficientes, volver al paso uno del proceso
    - viii. Si no eran los esperados, y habiendo realizado el proceso indicado en el paso 6, volver al paso 4.

La Figura 5 muestra un ejemplo gráfico de este proceso.



**Figura 4.** Visión global del proceso basado en KDD propuesto para analizar la interacción entre los estudiantes y los agentes (KDDIAE)



**Figura 4.** Visión global del proceso BIDAIE propuesto para analizar la interacción entre los estudiantes y los agentes

#### 4. Conclusiones

Estamos en una sociedad en la que abunda la cantidad de datos al alcance, y que por tanto, pueden ser recolectados, y con ello, la necesidad de sacar información útil de los mismos.

Es en este punto donde adquieren especial relevancia las técnicas de análisis de Big Data, aplicables a casi cualquier área, entre ellas la educación.

En este trabajo se ha realizado una propuesta de evaluación para el estudio de datos recolectados por agentes pedagógicos conversacionales para facilitar su integración en el aula.

## Agradecimientos

Este trabajo se ha financiado con los proyectos de investigación TIN2015-66731-C2-1-R del Ministerio de Economía y Competitividad y S2013/ICE-2715 de la Comunidad Autónoma de Madrid.

## Referencias

- [1] Russell, M., Bebell, D., O'Dwyer, L., & O'Connor, K. (2003). Examining teacher technology use implications for preservice and inservice teacher preparation. *Journal of Teacher Education*, 54(4), 297-310.
- [2] Herold, B. (2016). Technology in Education: An overview, *Education Week*, February, 5. [En línea], [Fecha de consulta: 28 Mayo 2016]. Disponible en: <http://www.edweek.org/ew/issues/technology-in-education/>
- [3] Johnson, W., Rickel, J., & Lester, J. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments, *Journal of Artificial Intelligence in Education* 11, 47-78.
- [4] Pérez-Marín, D., Pascual-Nieto, I. (2011). *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*. IGI Global.
- [5] Graesser, A., Person, N., & Harter, D. (2001). Teaching tactics and dialog in AutoTutor, *International Journal of Artificial Intelligence in Education* 12(3), 23-29.
- [6] Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B. & Bhogal, R. (1997). The person effect: affective impact of animated pedagogical agents, *Proceedings of the SIGCHI conference on Human factors in computing systems*.
- [7] Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning, *Journal of computer assisted learning*, 19(2), 195-208.
- [8] Hays, M., Lane, C., Auerbach, D., Core, M., Gomboc, D. & Rosenberg, M. (2009). Feedback Specificity and the Learning of Intercultural Communication Skills, *Artificial Intelligence in Education*.
- [9] Yee, N., & Bailenson, J. (2007). The Proteus effect: The effect of transformed self-representation on behavior, *Human Communication Research* 33, 3.
- [10] Chase, C., Chin, D., Opezzo, M., & Schwartz, D. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning, *Journal of Science Education and Technology* 18, 334-337.
- [11] Beyer, M. & Laney, D. (2012): *The Importance of 'Big Data': A Definition*, Gartner Research Report
- [12] Piatetsky-Shapiro, G. 1991. Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Work-shop. *AI Magazine* 11(5): 68-70.
- [13] Mitchell, T. (1997). *Machine Learning*, McGraw Hill.
- [14] Fleming, B. (2014) *Adaptive Learning Technology: What It Is, Why It Matters*, Eduventures, April. [En línea], [Fecha de consulta: 28 Mayo 2016]. Disponible en: <http://www.eduventures.com/2014/04/adaptive-learning-technology-matters/>
- [15] Vázquez, Y. (2001). Educación basada en competencias. *Educación: revista de educación/nueva época*, 16, 1-29.
- [16] Strayer, J. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, 15(2), 171-193.
- [17] Mas, A. (2015). *Agentes software y sistemas multiagente: conceptos, arquitecturas y aplicaciones*. Pearson Educación.
- [18] Alegsa, L. (2016). Definición de agente de software, [En línea], [Fecha de consulta: 28 Mayo 2016]. Disponible en: <http://www.alegsa.com.ar/Dic/agente%20de%20software.php>
- [19] Theodoridou, K. & Yerasimou, T (2008). Learning Spanish with "Laura": The role of an intelligent agent in a Spanish language course, in *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, vol. 2008, no. 1, 2008, pp. 4907-4912.
- [20] Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces, *AI Magazine*, vol. 22, no. 4, pp. 67-83
- [21] Graesser, A., Penumatsa, P., Ventura, M, Cai, Z & Hu, X (2007). Using Isa in autotutor:

- Learning through mixed initiative dialogue in natural language, Handbook of latent semantic analysis, pp. 243-262
- [22] D'mello, S. & Graesser, A. (2013). Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back, *ACM Trans. Interact. Intell. Syst.*, vol. 2, no. 4, pp. 23:1-23:39, Jan. ). [En línea], [Fecha de consulta: 28 Mayo 2016]. Disponible en: <http://doi.acm.org/10.1145/2395123.2395128>
- [23] Tamayo, S. (2012) Propuesta de desarrollo centrado en el usuario de un agente conversacional pedagógico para la comprensión lectora de ejercicios de matemáticas a nivel escolar, Master's thesis, Universidad Rey Juan Carlos
- [24] Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety, Technical report, META Group [En línea], [Fecha de consulta: 28 Mayo 2016]. Disponible en: <http://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- [25] Dutcher, J. (2014, 03 de Septiembre). What Is Big Data?, Blog, Departamento Data Science Universidad de Berkeley. ). [En línea], [Fecha de consulta: 28 Mayo 2016]. Disponible en: <https://datascience.berkeley.edu/what-is-big-data/>
- [26] Gartner (2016). Business Intelligence. IT Glossary. [En línea], [Fecha de consulta: 28 Mayo 2016]. Disponible en: <http://www.gartner.com/it-glossary/business-intelligence-bi/>
- [27] Maimon, O & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. Springer, New York. ISBN 978-0-387-09823-4.
- [28] Fayyad, U., Piatetsky-Shapiro, G. & Padhraic, S. (1996). From data mining to knowledge discovery: an overview. In *Advances in knowledge discovery and data mining*, Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy (Eds.). American Association for Artificial Intelligence, Menlo Park, CA, USA 1-34.
- [29] Witten, I. & Frank, E. (2011). *Data Mining: Practical machine learning tools and techniques* Morgan Kaufmann.
- [30] Shiffman, D. (2012). "The Natural of Code: Simulating Natural Systems with Processing". Chapter 10.
- [31] Rokach, L. & Maimon, O. (2008). *Data mining with decision trees: theory and applications*. World Scientific.
- [32] Rousseeuw, P.J.; Kaufman, L. (1990). *Finding Groups in Data: An Introduction to Clúster Analysis*. Wiley.
- [33] Flach, P. (2012). *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge University Press.
- [34] Salazar, J. (2016). "Big Data en la educación", *Revista Digital Universitaria*, 17(1).
- [35] Tamayo, S, Pérez-Marín, D. 2016a. Analizando la interacción de estudiantes de Educación Infantil y Primaria con un Agente Conversacional Pedagógico. XI International GUIDE Conference and IX Ikasnabar EDTECH Conference. In press.
- [36] Pérez-Marín, D., Salmerón, L. (2014). When Human-Computer Interaction meets Blended Learning and Heuristics are not Enough: Evaluating the usability with Field Observation and Interviews. In Press, Nova Publishing.
- [37] Brachman, R., and Anand, T. (1996). *The Process of Knowledge Discovery in Databases: A Human-Centered Approach*. In *Advances in Knowledge Discover yand Data Mining*, 37–58, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Menlo Park, Calif.: AAAI Press.
- [38] Pérez-Marín, D. (2010). Uso de agentes conversacionales pedagógicos en sistemas de aprendizaje híbrido (b-learning).