Proyecto Integrador
Inteligencia Artificial, Compilación, Simulación

Bryan Machín García, José Alejandro Solís Fernández y Adrianna Alvarez Lorenzo

Universidad de La Habana,
San Lázaro y L. Plaza de la Revolución, La Habana, Cuba
{bryan.machin,jose.solis,adrianna.alvarez}@estudiantes.matcom.uh.cu
http://www.uh.cu

Resumen En el presente informe se discute una propuesta de diseño e implementación de un sistema para un estudiante en particular, que permita simular diferentes ámbitos de aprendizaje a partir de un entorno inicial, con el propósito de obtener una estrategia de aprendizaje según las materias que este desea aprender.

Palabras Claves: aprendizaje, entorno, estrategia, estudiante, simulación

Abstract. This report discusses a proposal for the design and implementation of a system for a particular student, that allows the simulation of differents learning environments from an initial one, with the purpose of obtain a learning strategy according to the subjects he wants to learn.

Keywords: environment, learning, simulation, strategy, student

1. Introducción

El aprendizaje de un curso supone a un estudiante frente a un conjunto de contenidos de los cuales, a priori, puede desconocer absolutamente de su naturaleza. Por esta razón, la optimización de un proceso de aprendizaje es un trabajo que le presentaría altas dificultades. Este trabajo propone una alternativa de solución a dicha problemática.

2. Dominio del Problema

2.1. Elemento

Definición 1 *Es una materia o disciplina que se centra en un área de conocimiento diferenciada.*

Un elemento puede tener un conjunto de dependencias a otros elementos. Es decir, no se puede aprender el elemento e sin aprender el elemento e', si este último es dependencia de e.

2.2. Actividad

Definición 1 *Es una acción que posibilita el desarrollo del proceso de aprendizaje de ciertos elementos.*

Una actividad se constituye de la siguiente manera:

- conjunto de elementos que intervienen en la actividad
- cantidad de puntos de conocimiento que brinda por cada uno de los elementos
- tiempo estimado de duración

2.3. Entorno de Aprendizaje

Un entorno se compone de recursos que determinan cierto ámbito de aprendizaje:

- un estudiante
- un conjunto de elementos
- un conjunto de actividades
- un conjunto de reglas

El ente principal que caracteriza a un entorno de aprendizaje es un *estudiante*, el cual tiene como propósito aprender un subconjunto de elementos del conjunto inicial(*objetivos*), según las condiciones definidas para él en dicho entorno. Estas condiciones están dadas por el estado en que se encuentra el estudiante en cada uno de los elementos que conforma el entorno, denominadas *categorías*.

2.4. Categoría

Definición 1 *Especifica el nivel de aprendizaje en el que se encuentra un estudiante en cierto elemento.*

Las categorías existentes son:

- aprendido
- aprendible
- no aprendido
- olvidado

La transición de una categoría a otra se determina por el cumplimiento de reglas:
Ejemplo: Para que un elemento pase de la categoría *no aprendido* a la categoría *aprendible* es necesario que todas las dependencias de ese elemento estén en *aprendido*.

Además, la probabilidad de dicho tránsito puede estar condicionada por varios factores que se definan.
Ejemplo Un elemento puede pasar de la categoría *olvidado* a la categoría *aprendido* de manera totalmente aleatoria o pudiera ser que la probabilidad dependa de cuántas dependencias de ese elemento se encuentren en la categoría *olvidado* o *aprendido*.

3. Problema

Se desea diseñar e implementar un sistema que permita simular diferentes entornos de aprendizaje para un mismo estudiante a partir de un entorno inicial, basándose en el siguiente criterio:

Un entorno x difiere de un entorno x' en el nivel de aprendizaje de dicho estudiante en los elementos que lo constituyen.

Con lo antes mencionado, se propone obtener una estrategia de aprendizaje de elementos, en la que el estudiante en cuestión logre aprender la cantidad máxima de objetivos posibles, según las condiciones predefinidas de su entorno.

4. Modelación del Problema

Sea E un entorno de aprendizaje y un estudiante S. Se tiene que:

El conjunto de elementos de E se puede representar como un grafo dirigido $G = (V, A)$, donde:

- V es el conjunto de elementos de E
- $A = \{< e_1, e_2 > \in V \times V : e_2 \in D(e_1)\}$, siendo D el conjunto de dependencias de e_1.

G es un grafo acíclico dirigido(DAG). En efecto:
Como un arco $< e_1, e_2 >$ indica que e_1 depende de e_2, entonces sin pérdida de generalidad, la existencia de un ciclo $c = \{e_1, e_2, e_3, e_1\}$ implicaría que e_1 es dependencia de sí mismo, lo cual no tiene sentido en este modelo.

Luego, para determinar el nivel de aprendizaje del estudiante S en cada uno de los vértices de G se verifica el cumplimiento de las reglas definidas. Como resultado se obtiene la ubicación de estos en cada una de las categorías predeterminadas del sistema, al adaptar S a las condiciones impuestas por E.

Dado que G es un DAG, el conjunto de posibles estrategias de aprendizaje para satisfacer los objetivos de S son ordenaciones topológicas de G, pues la construcción de estas se basan en un orden de elementos dado por prioridades, de manera tal que el elemento e sea aprendible cuando se validen las restricciones de esa regla según sus dependencias.

Se considera como solución del problema una secuencia ordenada de los objetivos que deben ser aprendidos. Algunas soluciones pueden ser mejores que otras y el valor por el cuál se comparan es el porciento de aprendizaje obtenido al realizarse.

Para cada objetivo de la solución se buscan varios caminos en los que se aprenden

contenidos que a través del sistema de dependencias permitan aprenderlo. Estos caminos se encuentran siguiendo distintas métricas: buscando dependencias cuyos puntos disponibles a obtener sean mayores o que la cantidad de elementos por aprender para que su categoría mejore sea menor, dado que estos tienen un mayor grado de probabilidad de éxito. También se busca un camino aleatorio para incrementar la exploración del grafo. Además, estos caminos son simulados varias veces, devolviendo su porcentaje de aprendizaje promedio y tiempo que toman en completarse.

Una solución inicial factible se puede encontrar creando un orden topológico con los objetivos. Esta solución se somete a un proceso en el cual se buscan soluciones cercanas a ellas, que pertenezcan a una misma vecindad variable. Se realiza un solo intercambio en el orden de un par de objetivos y se evalúa el resultado de esta nueva solución. El par de soluciones es, entonces, comparado para obtener la mejor y repetir el proceso.

A pesar que de esta manera se busca una mejora de solución, es probable que se caiga en un óptimo local. Para contrarrestar esto, se utiliza la idea de una población de soluciones, aumentando la exploración y abarcando mayor terreno en el espacio de soluciones.

Para que las simulaciones se puedan lleavar a cabo satisfactoriamente y el resultado de evaluar las soluciones sea certero, es imprescindible un correcto y apropiado manejo de las categorías en las cuales se encuentran los elementos. Esta tarea es llevada a cabo por un agente que interactúa en y con el ambiente. Cuando este agente toma acción, decide si un elemento cambia de categoría o se mantiene en la misma.

También es importante no mantenerse en una simulación que persigue un objetivo que no puede ser alcanzado. Para evitar esta situación se toma auxilio de otro agente, que observa la repetición de un mismo contenido en la simulación y decide cuando es momento de dejar de intentar aprenderlo para el estado actual del ambiente

Para aprender un elemento es necesario realizar un conjunto de actividades, por lo que es importante elegir una sucesión de actividades apropiada a realizar. Distintas de estas sucesiones son revisadas, para explorar los escenarios posibles y más probables en los que el estudiante se puede ver envuelto: escogiendo las actividades qué más puntos de aprendizaje le pueda aportar, la que menos tiempo tome o mediante una elección aleatoria.

5. Compilación

Para el manejo del sistema se emplea el lenguaje de dominio específico *learnPro*. Este lenguaje se modela a partir de un conjunto de símbolos, cuya

implementación se encuentra en la clase `Symbol`. Esto permitirá posteriormente la definición símbolos terminales y símbolos no terminales.

Mediante el uso de símbolos se facilita la formación de oraciones al agruparse con el operador +, permitiendo el reconocimiento de la cadena especial **epsilon** a través de la propiedad `is_epsilon`. Además, posibilita el acceso a la gramática en la que se definió mediante el campo `Grammar` que contiene cada instancia, así como la consulta de su notación a través del campo `Name`.
En el caso de los símbolos no terminales, su modelación se encuentra en la clase `NonTerminal`, la cual extiende a la clase `Symbol` para permitir reconocer las producciones que tienen a este símbolo como cabecera, mediante el campo `productions` de cada instancia; añadir producciones para ese no terminal a través del operador %= e incluir las propiedades `is_non_terminal` e `is_terminal` que devolverán `True` o `False` respectivamente. Esto último se añade de igual manera para los símbolos terminales, cuya modelación se encuentra en la clase `Terminal`.

La clase `EOF` modela el símbolo de fin de cadena, cuyo comportamiento se hereda al extender la clase `Terminal`.
Las oraciones y formas oracionales de este lenguaje se modelarán con la clase `Sentence`, siendo una colección de terminales y no terminales. Con esta se conoce a priori la longitud de la oración, además de que se accede a los símbolos que componen la oración a través del campo `symbols` de cada instancia, y se puede conocer si dicha oración está completamente vacía a través de la propiedad `is_epsilon`. Mediante el operador + se puede obtener la concatenación con un símbolo u otra oración.

Para logar definir las producciones que tengan la misma cabecera en una única sentencia, se emplea el agrupamiento de oraciones usando el operador |, lo cual se maneja con la clase `SentenceList`.

En la clase `Epsilon` se modelará tanto la cadena vacía como el símbolo que la representa: ϵ. Dicha clase extiende las clases `Terminal` y `Sentence` por lo que adopta el comportamiento de ambas, sobreescribiendo la implementación del método `is_epsilon` para indicar que toda instancia de la clase reprensenta epsilon.

La clase `Production` modela las producciones, con la cual se puede acceder a la cabecera y cuerpo de cada producción mediante los campos `left` y `right` respectivamente, así como consultar si la producción es de la forma $X \to \epsilon$ haciendo uso de la propiedad `is_epsilon` y bifurcar la producción en cabecera y cuerpo haciendo uso de las asignaciones: `left, right = production`.

- Para definir una producción de la forma $E \to E + T$:

$$E \,\%= E + \textbf{plus} + T$$

- Para definir múltiples producciones de la misma cabecera en una única sentencia ($E \to E + T|E - T|T$):

$$E\,\%\, = E\ \mathbf{plus} + T|E + \mathbf{minus} + T|T$$

- Para usar ϵ en una producción, por ejemplo, $S \longrightarrow aS|\epsilon$ se procederá de la siguiente manera:

$$S\,\%\, = S + a|\mathbf{G.Epsilon}$$

La modelación de las gramáticas se realiza en la clase `Grammar`. Sus funcionalidades básicas son definir de la gramática los símbolos terminales, a través de los métodos `terminal` y `terminals` y los no terminales mediante `non_terminal` y `non_terminals` y denotar las producciones de la gramática a partir de la aplicación del operador $\%$ = entre no terminales y oraciones. A su vez, se puede acceder a todas las producciones a través del campo `Productions` de cada instancia, a los terminales y no terminales mediante los campos `Terminals` y `NonTerminals` respectivamente, y al símbolo inicial, _epsilon y fin de cadena(EOF) a través de los campos StartSymbol, Epsilon y EOF respectivamente.

Para el manejo de la pertenencia o no de *epsilon* a un conjunto se emplea la clase `ContainerSet`, la cual funciona como un conjunto de símbolos, posibilitando consultar la pertenencia de epsilon al conjunto. Las operaciones que modifican el conjunto devuelven si hubo cambio o no. Dicho conjunto puede ser actualizado con la adición de elementos individuales, con el método `add`, o a partir de otro conjunto, mediante `update` y `hard_update`.

Por otra parte, el conjunto First de una forma oracional se define como:

- $\mathbf{First}(w) = \{x \in V_t | w \Rightarrow^* x\alpha, \alpha \in (V_t \bigcup V_n)^*\}$
- $\bigcup\{\epsilon\}$, si $w \to *\epsilon$
- $\bigcup\{\}$, en otro caso.

Este es posible computarlo para los símbolos terminales, no terminales y producciones haciendo uso de un método de punto fijo. Para ello los *firsts* se inicializan vacíos y mediante las siguientes reglas se van se van actualizando con la aplicación de forma incremental:

- Si $X \longrightarrow W1|W2|...|Wn$ entonces por definición: First(X) = $\bigcup_i$ First(W_i)
- Si $X\epsilon$ entonces $\epsilon \in$ First(X)
- Si $W = xZ$ donde x es un símbolo terminal, entonces First(W) = {x}
- Si $W = YZ$ donde Y es un símbolo no terminal y Z una forma oracional, entonces First(Y) $\subseteq$ First(W)
- Si $W = YZ$ y $\epsilon \in$ First(Y) entonces First(Z) $\subseteq$ First(W)

El cálculo de los `firsts` se da por terminado cuando finalice una iteración sin que se produzcan cambios.

Para la implementación de dicho algoritmo se tiene el método `compute_local_first`,

que calcula el `First(alpha)`, siendo `alpha` una forma oracional. Con el método `compute_firsts` se calculan todos los conjuntos firsts actualizando a los conjuntos iniciales según los resultados de aplicar `compute_local_first` en cada producción.

Una *gramática atributada* es una tupla $< G, A, R >$ donde:

- G $= < S, P, N, T >$ es una gramática libre del contexto
- A es un conjunto de atributos de la forma X $\cdot\alpha$ donde $X \in N \bigcup T$ y α es un identificador único entre todos los atributos del mismo símbolo y,
- R es un conjunto de reglas de la forma $< p_i, r_i >$ donde $p_i \in P$ es una producción $X \longrightarrow Y_1, ..., Y_n$ y r_i es una regla de la forma:
  1. X $\cdot\alpha = f(Y_1 \cdot \alpha_1, ..., Y_n \cdot \alpha_n)$, o
  2. $Y_i \cdot \alpha = f(X \cdot \alpha_0, Y_1 \cdot \alpha_1, ..., Y_n \cdot \alpha_n)$

Los atributos se dividen en dos conjuntos disjuntos: *atributos heredados* y *atributos sintetizados*, como es el caso de α en (1) y en (2) respectivamente.
Las condiciones suficientes para que una gramática sea evaluable son:

- Una gramática atributada es *s-atributada* si y solo si, para toda regla r_i asociada a una producción $X \to Y_1, \ldots, Y_n$, se cumple que r_i es de la forma $X \cdot a = f(Y_1 \cdot a_1, \ldots, Y_n \cdot a_n)$.
- Una gramática atributada es *l-atributada* si y solo si toda regla r_i asociada a una producción $X \to Y_1, \ldots, Y_n$ es de una de las siguientes formas:
  1. $X \cdot a = f(Y_1 \cdot a_1, \ldots, Y_n \cdot a_n)$, ó
  2. $Y_i \cdot a_i = f(X \cdot a, Y_1 \cdot a_1, \ldots, Y_{i-1} \cdot a_{i-1})$

A la *API* de gramáticas se añade una nueva clase: `AttributeProduction`. Con esta clase se modela las producciones de las gramáticas atributadas. Cada una de estas producciones se compone por un símbolo no terminal como cabecera, accesible a través del campo `Left`, una oración como cuerpo, a través del campo `Right` y un conjunto de reglas para evaluar los atributos, accesible a través del campo `atributes`.

Se implementó la clase `Item` para modelar los items del parser LR(1), cuyos `lookaheads` se almacenarán haciendo uso del parámetro `lookaheads`.

Cada item tiene definido una función `Preview`, la cual devuelve todas las posibles cadenas que resultan de concatenar lo que queda por leer del item tras saltarse x símbolos con los posibles `lookaheads`, que resultan de calcular el `first` de estas cadenas.

Para calcular la clausura se implementó la función `expand`, que recibe un item LR(1) y retorna un conjunto de items que sugiere incorporar debido a la presencia de un $\cdot$ delante de un no terminal.

$$expand(\text{"}Y \to \alpha.X\delta, c\text{"}) = \text{"}X \to .\beta, b\text{"} | b \in First(\delta c)$$

Luego se implementó la función `compress`, que recibe un conjunto de items LR(1) y devuelve dicho conjunto pero combinando los `lookaheads` de los items con mismo centro.

Teniendo en cuenta ambas funciones, se implementó la función de clausura utilizando la técnica de punto fijo, basándose en lo siguiente:

$CL(I) = I \bigcup \{X \longrightarrow \cdot\beta, b\}$ tales que:

- $Y \longrightarrow \alpha \cdot X\delta, c \in CL(I)$

- $b \in$ **First(**δc) Por otro lado se tiene la implementación de la función `goto(Ii, s)`, que cumple que:

$$Goto(I, X) = CL(\{Y \longleftrightarrow 0\alpha X \cdot \beta, c | Y \longleftrightarrow \alpha \cdot X\beta, c \in I\}$$

Esta función recibe como parámetro un conjunto de items y un símbolo, y retorna el conjunto `goto(items, symbol)`. Este método permite darle valor al parámentro `just_kernel=True` para calcular el conjunto de items kernels. En caso contrario, se requiere el conjunto con los `firsts` de la gramática para entonces calcular la clausura.

A continuación se muestra la implementación del algoritmo para construir el automáta LR(1):

Este parser LR(1) llena la tabla Acción-Goto de la siguiente manera:

- Sea "$X \to \alpha \cdot c\omega, s$" un item del estado I_i y `Goto(`I_i`, c) =` I_j, entonces $ACTION[I_i, c] = 'S'_j$ item Sea "$X \to \alpha., s$ün item del estado I_i, entonces $ACTION[I_i, s] = 'R'_k$ (producción `k` es $X \to \alpha$).
- Sea I_i el estado que contiene el item "$S' \to S., \$$"(S' símbolo inicial), entonces $ACTION[I_i, \$] =$ 'OK'
- Sea "$X \to \alpha.Y\omega, s$ïtem del estado I_i Y $Goto(I_i, Y) = I_j$, entonces $Goto[I_i, Y] = j$