

Análisis y detección de patrones en un grafo conceptual construido a partir de respuestas escritas en forma textual a preguntas sobre un tema específico.

María Alejandra Paz Menvielle, Cynthia Lorena Corso, Analía Guzmán, Martín Gustavo Casatti, Karina Ligorria

*Departamento de Ingeniería en Sistemas de Información
CIDS – Centro de Investigación Desarrollo y Transferencia
de Sistemas de Información
Facultad Regional Córdoba – Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina – Córdoba
0351 – 4686385*

pazmalejandra@gmail.com, corso.cynthia@gmail.com, analia.guzman@the-group.com.ar,
mcasatti@gmail.com, karinaligorria@hotmail.com

Resumen

En este trabajo se presenta un proyecto que busca descubrir patrones asociados a una base de conocimiento representada como una base de datos orientada a grafos. La misma contiene respuestas de exámenes en formato de texto de redacción libre relacionadas a un dominio específico, utilizada para realizar el análisis de texto en respuestas a preguntas de exámenes en la cátedra de Paradigmas de Programación, con el objetivo de detectar el grado de acierto de las respuestas de los alumnos, a la forma de representación de las preguntas de los docentes, entre otras.

Hasta el momento se ha realizado el estudio teórico de diferentes patrones, el estudio y la elección de la forma en que se implementa el diseño de un sistema de reconocimiento de los mismos, finalizando en el análisis e identificación de diferentes patrones que puedan servir para mejorar el dictado de la materia y mejorar el aprendizaje, objetivo este central en cualquier proceso educativo.

Palabras clave: evaluación - grafos – patrones.

1. Identificación

Código del PID: PIDEIUTNCO0004812.

Tema en que se inserta: Investigación aplicada en Sistemas de Información e informática, en el campo de aplicación del desarrollo de la educación y de la promoción general del conocimiento.

Fecha de Inicio: 01/01/2018

Fecha de Fin: 31/12/2019

2. Introducción

El proyecto utiliza como caso testigo a la cátedra de Paradigmas de Programación, perteneciente a la carrera Ingeniería en Sistemas de Información, dictada en la Facultad Regional Córdoba, de la Universidad Tecnológica Nacional.

El trabajo que aquí se presenta es la continuación de los trabajos realizados durante el desarrollo del PID EIUTNCO0003592 “Metodología para determinar la exactitud de una respuesta, escrita en forma textual, a un interrogatorio sobre un tema específico”. Durante el transcurso de dicho proyecto se generó una base de conocimiento modelada como grafo

dirigido, que posee preguntas y respuestas de exámenes escritos en forma textual, así como cualquier otro concepto contenido en el programa de estudios de la materia Paradigmas de Programación. Dicho grafo es una base de conocimientos amplia y extensible, generada inicialmente por los docentes de la cátedra, utilizada para evaluar a los alumnos y que pueda ser alimentada con las respuestas que no hayan sido consideradas aún y que aporten mayor variedad a la base de conocimiento.

En el presente proyecto se propone complementar la funcionalidad del proyecto anteriormente mencionado, creando un segundo grafo, obtenido a partir de las respuestas de todos los alumnos que realizan los exámenes y que almacene el historial de las respuestas de los alumnos, para poder cotejarlo después con el grafo inicial y así poder realizar la búsqueda, el análisis y la propuesta de patrones frecuentes en el grafo conceptual.

Con el análisis de patrones se pretende descubrir algunas características importantes que se relacionen con las respuestas de los alumnos y que puedan reflejar la evaluación y el aprendizaje de los mismos. También que se pueda determinar el nivel de profundidad con el que se evalúa cada tema, si los distintos exámenes son consistentes en alcance y profundidad de evaluación de conceptos, entre otros. En relación a los contenidos de la materia, se pretende identificar si hay áreas de conocimiento evaluadas con mayor frecuencia, si hay temas que no se están evaluando, qué conceptos de la materia son mejor conocidos por los alumnos y cuáles lo son en menor medida. Esto permitirá, en última instancia, mejorar los materiales didácticos empleados como así también ajustar los instrumentos de evaluación.

2.1 Patrones y grafos

Un patrón es una entidad a la que se le puede dar un nombre y que está representada por un conjunto de propiedades medidas y las relaciones entre ellas (vector de características) (Watanabe,1985).

En el dominio utilizado como caso testigo, por ejemplo, un patrón puede ser la ruta resultante de una respuesta de un alumno, de las cuales se extrae el vector de características formado por un conjunto de valores numéricos que pueden representar nivel de exactitud de la respuesta, la puntuación de la misma, la cantidad utilizada de conceptos y de relaciones, etc.

El reconocimiento automático, descripción, clasificación y agrupamiento de patrones son actividades importantes en una gran variedad de disciplinas científicas, como la biología, psicología, medicina, visión por computadora, inteligencia artificial, teledetección, etc.

Lo importante de detectar patrones en los datos es que se pueden inferir causas para la agrupación de los mismos (en el caso de que se estén detectando patrones ya conocidos y ya estudiados).

Este trabajo, tiene paralelos con el llamado SNA (Social Network Analysis, Análisis de Redes Sociales) que es una disciplina cuyo objetivo es “Analizar la estructura de una red social para ‘inferir conocimiento’ de un individuo, un grupo, o las relaciones entre ellos” (Scott y Carrington, 2011).

Un grafo conceptual (Sowa, 1992) es un sistema de notación simbólica y de representación del conocimiento. Presentado por John F. Sowa, se basa en los gráficos existenciales (Peirce, 1909) de Charles Sanders Peirce, en las estructuras de redes semánticas y en datos de la lingüística, la filosofía y la psicología.

2.2 Patrones en Grafos

El reconocimiento o detección de patrones dentro de grafos busca detectar un subgrafo (patrón) en un grafo (objetivo). Debemos considerar que esta búsqueda de coincidencias se puede descomponer en dos partes:

1. Una concordancia estructural, en donde los nodos y relaciones del patrón conforman una estructura existente en el grafo objetivo
2. Una concordancia a nivel de elementos, en donde los nodos y relaciones, a nivel de sus atributos

particulares, tiene los mismos valores que en la estructura encontrada en el grafo objetivo.

Muchas veces la búsqueda de estas dos concordancias se ejecuta de forma separada para optimizar los algoritmos o reducir el espacio de búsqueda (Fan, 2012).

En el dominio bajo estudio, la detección de un subgrafo (patrón), se realizará en los grafos (objetivos) que representan los contenidos de la materia y las respuestas de los alumnos en las instancias de exámenes.

2.3. Métricas en grafos

Una herramienta ampliamente utilizada para describir grafos y que muchas veces se utiliza para iniciar el análisis de patrones existentes en los mismos, es el cálculo de métricas (Van Steen, 2010), locales o globales, que permiten caracterizar el grafo objetivo o el grafo patrón. Las métricas se pueden dividir en dos grandes grupos:

- Métricas estáticas: Cuando se calculan sobre un grafo estático en un punto en el tiempo determinado. Se enfocan principalmente en las características estructurales del mismo.
- Métricas dinámicas: Tienen en cuenta la dimensión temporal de los cambios que se producen sobre el grafo. Están más enfocadas en las variaciones entre dos instantes de tiempo, antes que en las características propias del grafo en cada uno de esos instantes.

Otro enfoque para el análisis de las métricas radica en analizar sobre qué componentes del grafo se realizan las mediciones. Desde este punto de vista se tienen diversas perspectivas, siendo las más comunes:

- Métricas de redes (o globales): Son las métricas que toman como referencia el grafo completo, con todos los nodos y arcos que lo conforman. Por ejemplo: Centralidad, Conexionado, Cantidad de Componentes, Tamaño del

componente gigante, Ruta más corta/larga, entre otras.

- Métricas nodos (o locales): Son aquellas que toman como referencia un nodo o subconjunto de nodos para realizar los cálculos. Por ejemplo: Conectividad, Centralidad de grado, Centralidad de cercanía, Centralidad de intermediación, entre otras.

3. Objetivos, Avances y Resultados

3.1 Objetivos

El objetivo del proyecto es analizar, detectar y evaluar patrones topológicos frecuentes en un grafo conceptual construido para determinar la exactitud de las respuestas, escritas en forma textual sobre un tema específico, utilizando una base de conocimientos diseñada como un grafo dirigido. Para ello se han identificado los siguientes objetivos particulares:

1. Explorar patrones topológicos de grafos que contengan información relevante para la identificación de estructuras dentro de la base de conocimientos de la materia Paradigmas de Programación.
2. Analizar la existencia de patrones recurrentes o subyacentes en los grafos generados a partir de las respuestas base de los docentes y los obtenidos de las respuestas dadas por los alumnos.
3. Proponer algoritmos que permitan detectar patrones conocidos en la teoría de grafos como son las "comunidades", "pares", "rutas principales" y otros patrones comunes, por medios del análisis de las métricas sobre la base de conocimiento.
4. Identificar características como exactitud, coherencia y consistencia, entre otras, de las respuestas escritas en forma textual, en la base de conocimiento diseñada como un grafo dirigido.

3.2 Avances

Al momento de la confección de este trabajo se ha podido cumplir con los dos primeros objetivos particulares del proyecto.

Para comenzar se realizó la exploración de diferentes patrones topológicos de grafos que pudieran ser relevantes para la identificación de estructuras dentro de la base de conocimientos de la materia Paradigmas de Programación.

Luego se analizaron diferentes formas de diseñar un sistema de reconocimiento de patrones, llegando a establecer dicho diseño a través de tres fases (Alonso Romero y Calonge Cano, 2001):

1. Adquisición y preproceso de datos.
2. Extracción de características.
3. Toma de decisiones o agrupamiento.

1 Adquisición y preproceso de datos.

En la introducción del presente trabajo, se mencionó que se cuenta con la base de conocimiento de la materia, donde están almacenados en un grafo, los datos adquiridos y procesados relacionados a los contenidos, las preguntas y las respuestas, tanto las provistas por los docentes como las respondidas por los alumnos.

Para el manejo de la base de datos de grafos se utilizó el producto OrientDB Community Edition, es una aplicación de código abierto, con licencia Apache 2 y gratuita para todo tipo de uso. Esta base de datos de grafos implementa de forma nativa dos características que son centrales en el planteo del método de corrección. La implementación se realizó en el lenguaje Java, utilizando librerías de corrección ortográfica y de visualización de la base de grafos.

Para poder generar la segunda base de grafos, que contiene las respuestas de los alumnos, se realizó el análisis y diseño de la registración automática, cuando el sistema califica las respuestas de los alumnos en las distintas instancias de los exámenes.

2 Extracción de características y agrupamiento

Para las dos siguientes fases, extracción de características y toma de decisiones o agrupamiento, se consideraron los objetivos y contenidos de la materia, las necesidades de los docentes de la cátedra, respecto a la información que les podría brindar el prototipo de calificación y en general, las características que son de interés para evaluar o estudiar los mecanismos de evaluación de la materia y el grado de aprendizaje de los alumnos.

También se consideraron algunas métricas mencionadas en el apartado anterior, relacionadas a los patrones inicialmente identificados.

Se identificaron dos grandes clases de patrones, aquellos relacionados a la evaluación y aquellos relacionados al aprendizaje de los alumnos, a partir de allí se definieron los conjuntos de patrones asociados a cada uno.

3.2 Resultados

El reconocimiento de patrones es un ámbito de gran auge en la actualidad, si bien la mayoría de las aplicaciones se orientan al reconocimiento de patrones sobre imágenes u otro tipo de información bidimensional, el reconocimiento sobre grafos está cobrando cada vez mayor importancia debido en gran medida al auge de las redes sociales, naturalmente modeladas como grafos, principalmente en la búsqueda de información valiosa que ya existe en la estructura de datos y que se pueda aprovechar.

En el caso particular de este proyecto de investigación se busca detectar patrones subyacentes tanto en la información del programa de la materia como de los exámenes elaborados por los docentes o en los conocimientos aprehendidos por los estudiantes, de forma tal que se pueda mejorar el dictado de la materia y mejorar el aprendizaje, objetivo este central en cualquier proceso educativo.

Se estima que los patrones previamente detectados brindarán a la cátedra de Paradigmas de Programación un sustento fáctico sobre el cual basar las posibles modificaciones a los materiales de estudio a

los métodos de dictado y a los instrumentos de evaluación, así como una herramienta tecnológica eficaz para evaluar, en un tiempo relativamente breve, el impacto de esas modificaciones sobre el cursado y la evaluación.

Se proseguirá la investigación de forma tal que los docentes cuenten con herramientas informáticas para realizar consultas sobre la base de datos de grafos, que ayuden a detectar patrones diferentes a los actualmente planteados. El objetivo final es que los propios docentes puedan indicarle al sistema, a modo de aprendizaje supervisado, qué se debe buscar de acuerdo a los objetivos de la cátedra y tener una realimentación inmediata que posibilite el análisis de diversos escenarios a la hora de plantear herramientas de evaluación y planificar el dictado.

4. Formación de Recursos Humanos

La estructura del equipo de investigación está conformada por un director, un codirector, dos investigadores tesistas, un investigador de apoyo, un becario BINID, un becario alumno SAE y un becario alumno UTN-SCYT.

Relacionado con el tema del proyecto, hasta el momento se ha presentado y aprobado el trabajo integrador en la Especialidad en Sistemas de Información de uno de los investigadores tesistas. Se estima que en el transcurso del mes de agosto también presentará el plan de tesis para la Maestría en Sistemas de Información.

En el transcurso de los meses siguientes el otro investigador tesista presentará el trabajo integrador en la Especialidad en Sistemas de Información y el plan de tesis para la Maestría en Sistemas de Información.

Los becarios se incorporan con la finalidad de que inicien su formación en investigación científica y tecnológica, quienes también colaborarán en la recolección, manipulación y desarrollo de este marco metodológico.

En el marco del proyecto los estudiantes tendrán la posibilidad de hacer la Práctica Supervisada de quinto año. Los avances,

propuestas y herramientas construidas, estarán disponibles para su transferencia y aplicación en el Centro de Investigaciones, Desarrollo y Transferencia de Sistemas de Información - CIDS. Del mismo modo la detección de patrones sobre el dominio de conocimiento de la materia Paradigmas de Programación continuará beneficiando a los integrantes de la cátedra y a los estudiantes.

5. Publicaciones relacionadas con el PID

Desde el comienzo del proyecto se realizó la difusión de las actividades a través de la participación en congresos. Para la difusión del nuevo proyecto se presentó un trabajo en formato de paper, en WICC 2018 (WICC, 2018), en el cual se aprobó y se realizó la presentación resumen en formato de Poster. Para la difusión del primer resultado de los avances, se confeccionó un trabajo en formato de paper, presentado en CADI y CAEDI (4° CADI y 10° CAEDI, 2018), con aprobación sin objeciones.

Referencias

4° CADI y 10° CAEDI, 2018. *4to. Congreso Argentino de Ingeniería - CADI 2018 - y 10mo. Congreso Argentino de la Enseñanza de la Ingeniería - CAEDI -*, 19, 20 y 21 de septiembre de 2018, en Hotel Holiday Inn de la Ciudad de Córdoba – ARGENTINA.

Alonso Romero L. y Calonge Cano, T. (2001). *Redes neuronales y reconocimiento de patrones*.

Fan, W. (2012) *Graph pattern matching revised for social network analysis*. Proceedings of the 15th International Conference on Database Theory. ACM, p. 8-21.

Peirce, C (1909) *Existential graphs*, Collected Papers of Charles Sanders Peirce 4, pp. 1-7.

Scott, J. y Carrington, P. (2011). *The SAGE handbook of social network analysis*. SAGE publications.

Sowa, J. (1992) *Conceptual graph summary*, Conceptual Structures: Current Research and Practice. Ellis Horwood, New York London Toronto, pp. 3-66.

Van Steen, M. (2010) *An Introduction to Graph Theory and Complex Networks*. Copyrighted material.

Watanabe S. (1985) *Pattern Recognition: Human and Mechanical*. Wiley, New York.

WICC, 2018, 20° edición del workshop, Corrientes, Argentina, 26 y 27 de abril de 2018, organizado por Red de Universidades Nacionales con Carreras de Informática (RedUNCI) y la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste.