

EVALUACIÓN DE COMPETENCIAS SUBJETIVAS. EXPERIENCIA EN LA EVALUACIÓN DEL RENDIMIENTO DEL TRABAJO EN GRUPO DE LOS ESTUDIANTES

MONTERO MORALES, José Ant. ⁽¹⁾; GÓMEZ URGELLÈS, Joan ⁽²⁾

ALÍAS PUJOL, Francesc ⁽³⁾; GARRIGA BERGA, Carles ⁽⁴⁾

VICENT SAFONT, Lluís ⁽¹⁾; BADÍA FOLGUERA, David ⁽¹⁾

montero@salle.url.edu

⁽¹⁾Universidad Ramon Llull, España, ETSEEI La Salle, Dpto. Comunicaciones

⁽²⁾Universidad Politécnica de Cataluña, España, EPSEVG, Dpto. Matemáticas

⁽³⁾Universidad Ramon Llull, España, ETSEEI La Salle, Dpto. Tecnologías Audiovisuales y Multimedia

⁽⁴⁾Universidad Ramon Llull, España, ETSEEI La Salle, Dpto. Electrónica

RESUMEN

La evaluación de competencias subjetivas como la capacidad de trabajo en grupo suele provocar en el profesorado una sensación de ‘poco control’ sobre el proceso evaluador. En asignaturas con un gran número de alumnos y en las que conviven varios profesores, resulta muy difícil fijar los mismos criterios subjetivos y aplicarlos del mismo modo por parte de todos los profesores. Incluso resulta difícil para un único profesor asegurar que el modo de aplicar los criterios considerados para evaluar esa competencia es el mismo para todos los alumnos.

En este artículo se presenta una herramienta de soporte al profesorado en el proceso evaluador de competencias subjetivas. El objetivo es disponer de una herramienta automática que permite aplicar los criterios subjetivos definidos por el profesorado de antemano para la evaluación de una competencia, asegurando así que la evaluación se realiza de forma equitativa para todos los alumnos. Para ello se ha escogido, como técnica de modelado, la lógica difusa frente a otras posibles técnicas del mundo del aprendizaje artificial, dada su interpretabilidad y facilidad en la modelización del razonamiento seguido por el profesor cuando evalúa a sus alumnos. La aplicación de las técnicas de modelado difuso adecuadas permite obtener un sistema evaluador que supone una posible solución al problema planteado, con el valor añadido de conocer las reglas aplicadas por los profesores durante el proceso evaluador, dada la fácil interpretabilidad del modelo final obtenido.

Palabras clave: evaluación de competencias, trabajo en grupo, lógica difusa.

1. Introducción

En revistas y congresos del ámbito de la educación pueden encontrarse numerosas experiencias que presentan formas diferentes de evaluar el rendimiento del trabajo en grupo llevado a cabo por los estudiantes. Por ejemplo, en [1] el trabajo en grupo se evalúa a partir de los conocimientos en uno de los alumnos de cada grupo, tomado al azar. Es decir, se supone que los alumnos trabajan en grupo y que todos deberían saber lo mismo al final del proceso, por lo que se evalúan los conocimientos (a partir de una prueba) de uno de los miembros del grupo y su nota es, automáticamente, la de todo el grupo. En [2] se evalúa la calidad del trabajo en grupo en base a los documentos presentados por el grupo, relativos a actas de reuniones, planificaciones, repartos de tareas, etc. Sin embargo, el razonamiento seguido para dar una nota final a partir de esos documentos queda a criterio (subjetivo) del profesor. Es decir, ¿qué pasa si se planifica bien, pero las actas no son aceptables? ¿Se evaluaría del mismo modo que si las actas son buenas pero la planificación y el reparto de tareas es malo? Bajo este enfoque, el grado de subjetividad en manos del profesor es muy elevado, por lo que difícilmente podrá mantener el criterio evaluador al poner las diferentes notas a los diferentes grupos de alumnos. Otro posible modo de evaluar el trabajo en grupo es el que se basa en la opinión que cada miembro del grupo tiene de sus compañeros. Un ejemplo de esta filosofía de evaluación lo encontramos en la *University of Technology (Sydney)* [3], donde la idea básica es la de permitir que los alumnos de un grupo se evalúen ellos mismos y que evalúen también a sus compañeros de grupo. Así, en base a sus respuestas en los formularios, el profesorado define cómo se combinarán los datos recopilados con el fin de generar una nota de trabajo en grupo, individual para cada alumno.

Todas estas filosofías evaluadoras citadas anteriormente utilizan datos objetivos (notas en pruebas realizadas por los alumnos y/o puntuaciones en encuestas) para obtener una nota final asociada al trabajo en grupo desarrollado. La mayoría de estas filosofías definen previamente una ponderación para cada dato considerado y calculan la nota final asociada al trabajo en grupo realizando la combinación lineal correspondiente. Otras, sin embargo, dejan en manos del profesor la decisión final apelando a su experiencia, conocimientos y sentido común. En este segundo caso, se intentan tener en cuenta aspectos subjetivos que pueden desprenderse de los datos objetivos disponibles (por ejemplo la progresión observada en los estudiantes, la relación entre la nota de grupo y las notas individuales de sus miembros, la sensación de compromiso adquirido entre los miembros de un grupo, etc), aunque en este caso resulta mucho más difícil mantener la objetividad y la justicia (en el sentido de que se aplican los criterios evaluadores del mismo modo a todos los alumnos) durante el proceso evaluador. En esta última situación, se hacen necesarias herramientas que evalúen de forma automática esta competencia garantizando que se aplica el criterio subjetivo seguido por el experto (profesor) de forma objetiva y justa.

En este artículo se presenta una herramienta de soporte al profesorado en el proceso evaluador del rendimiento del trabajo en grupo realizado por los estudiantes. Una herramienta que no se limite a ponderar la nota de cada uno de los miembros del grupo, sino que aplique los criterios subjetivos que el profesorado considere necesarios y adecuados para tal evaluación, pero asegurando que se hace de forma equitativa para todos los alumnos. Una herramienta que permita realizar de forma automática dicha evaluación y que permita obtener un mayor conocimiento del modo en el que el profesorado razona para emitir la nota final correspondiente. Para ello se ha escogido, como técnica de modelado, la lógica difusa frente a otras posibles técnicas del mundo del aprendizaje artificial (redes neuronales o algoritmos genéticos, por ejemplo), dada su interpretabilidad y facilidad en la modelización del razonamiento seguido por el profesor cuando evalúa a sus alumnos. Con otras técnicas se podría abordar el problema considerado, aunque el sistema final alcanzado no permitiría obtener información asociada al razonamiento aplicado por el profesor, sino que debería ser tratado como una ‘caja negra’ entrenada para resolver, de forma desconocida para nosotros, el problema planteado. Sin embargo, la aplicación de las técnicas de modelado difuso adecuadas permite obtener un sistema evaluador que supone una posible solución al problema planteado, con el valor añadido de conocer las reglas aplicadas por los profesores durante el proceso evaluador, dada la fácil interpretabilidad del modelo final obtenido.

2. La lógica difusa como herramienta válida a la hora de evaluar competencias subjetivas

La lógica difusa fue formalizada por el profesor Lotfi A. Zadeh en 1965 y, desde entonces, ha sido y es aplicada en diferentes disciplinas. Aunque la aplicación de la teoría difusa destaca sobre todo en el campo del control de sistemas, algunos autores la han aplicado también en un contexto docente. La lógica difusa en el ámbito docente se ha utilizado para orientar al estudiante durante el proceso de aprendizaje [4], [5], o al profesor en el proceso tutorial [6]. En otros casos se ha utilizado para evaluar de forma diferente a la tradicional los conocimientos de los alumnos en una determinada materia [7], o para evaluar competencias desarrolladas en los estudiantes a partir de la opinión de otros estudiantes [8]. En este artículo, la lógica difusa se aplica como técnica base para la evaluación automática de una competencia subjetiva como es el rendimiento del trabajo en grupo de los alumnos, a partir de datos objetivos.

Aplicar la lógica difusa a un problema, supone trabajar con variables lingüísticas a las que se asocian conjuntos difusos, que son funciones definidas para todos los posibles valores que puede tomar la variable lingüística (universo del discurso). Es decir, un conjunto difuso es una función que asigna a cada uno de los valores posibles del universo del discurso un número real comprendido entre 0 y 1. Si dicho número es un 0 indica que no se pertenece en absoluto a dicho conjunto difuso; si es un 1 sí se pertenece totalmente al conjunto difuso en cuestión; y si es un valor intermedio éste indicará la mayor o menor pertenencia al conjunto difuso. A esta función se la denomina función de pertenencia y a cada posible valor asignado se le denomina grado de pertenencia al conjunto difuso correspondiente. Representa pues una extensión multivaluada del álgebra de Boole (bivaluada).

Conviene diferenciar claramente la teoría de los conjuntos difusos de la teoría de probabilidad. A continuación se presenta un ejemplo sencillo con el que dicha diferencia puede comprenderse con claridad: Supongamos que disponemos de dos botellas de agua y nos dicen que la primera de ellas contiene agua potable con un 90% de probabilidad mientras que la segunda contiene agua cuyo grado de pertenencia al conjunto difuso de bebidas potables es de 0,9. ¿Cuál de ellas escogería el lector? Sin duda la segunda, ya que la calidad del agua es muy alta (casi totalmente potable), mientras que si elige la primera podría darse la circunstancia de que ésta fuera no potable (hay un 10% de probabilidad de que esto sucediera).

Para emular la manera de razonar de los seres humanos no es suficiente obtener una representación de las palabras mediante los conjuntos difusos, sino que es necesario poder operar con ellos. Así, aparece el concepto de relación difusa. Las relaciones difusas se utilizan para relacionar los antecedentes con los consecuentes de sentencias lingüísticas. Un controlador difuso se implementa a partir de una serie de reglas que describen lingüísticamente el modo en el que deseamos que se comporte el mismo. Estas reglas de funcionamiento generalmente presentan la forma ‘si ... entonces...’ y recogen el conocimiento subjetivo que el diseñador (o el experto cuyo comportamiento pretende emularse) posee. Un ejemplo sencillo sería el siguiente:

SI asistencia a clase de un alumno es BAJA Y nota en examen es BAJA, ENTONCES interés por la asignatura es BAJO.

Como puede observarse, las reglas presentan un formato muy similar al razonamiento realizado por el experto cuando lleva a cabo algún proceso. Por tanto, permiten, de una forma sencilla, trasladar al controlador el conocimiento subjetivo que dicho experto posee. Asimismo, si las reglas proceden de un proceso previo de diseño, éstas permiten una interpretación relativamente sencilla del sistema final obtenido.

El proceso de evaluación de cada regla (es decir, el cálculo de la salida correspondiente a cada una de las reglas según los valores obtenidos en la fusificación) se denomina *implicación*, y gracias a él se obtiene el grado de activación de cada una de las reglas. Finalmente se dispone de un conjunto difuso de salida para cada regla previamente definida, que se modifica según el grado de activación de la regla correspondiente. El proceso de cálculo de un único valor de salida en el controlador difuso a partir de los conjuntos difusos de salida se denomina *defusificación*.

Para profundizar en aspectos relacionados con la lógica difusa, consultar [9].

3.- Evaluación del rendimiento del trabajo en grupo realizado por los estudiantes

3.1. Presentación

Desde hace varios años, en la asignatura de álgebra lineal impartida en la ETSEEI La Salle perteneciente a la Universidad Ramon Llull se han ido incorporando numerosos cambios en la metodología docente aplicada en el aula. Entre los más importantes pueden destacarse la incorporación de la modelización matemática (con numerosos ejercicios contextualizados en un ámbito técnico que son resueltos aplicando conceptos algebraicos) o la aplicación del aprendizaje cooperativo como estrategia de trabajo en grupo, haciendo énfasis en el cumplimiento de los principios que rigen este modo de trabajar con los alumnos (especialmente los principios de heterogeneidad y de interdependencia positiva). Durante el curso los alumnos realizan, entre otras actividades, numerosos controles en la clase, primero de forma individual, y posteriormente en grupo, entregando al profesor, además de la resolución individual del control, otra resolución del mismo control consensuada por todos los miembros del grupo. Las notas obtenidas en estos controles representan los datos a partir de los cuales se evaluará el rendimiento del trabajo realizado en grupo por los alumnos. El procedimiento seguido se presenta en las secciones siguientes.

Para profundizar en los detalles correspondientes a la metodología seguida en la asignatura de álgebra lineal en la ETSEEI La Salle, puede consultarse [10].

3.2. Parámetros para evaluar el trabajo en grupo

De todos los datos recogidos de los alumnos a lo largo del curso en la asignatura de álgebra lineal, se han considerado las notas individuales y las de grupo obtenidas en cada uno de los controles de evaluación continua realizados durante un parcial, con el fin de evaluar el rendimiento del trabajo en grupo realizado por los alumnos.

Conviene resaltar que, a medida que el número de controles propuestos a los alumnos aumenta, la cantidad de notas (individuales y de grupo) disponible hace mucho más difícil realizar la evaluación. Por ello, resultará necesario parametrizar el conjunto de notas de algún modo. En esta sección se presentan 4 parámetros definidos para facilitar el proceso posterior de automatización de la evaluación, así como para centrar la atención del profesor que evalúa, aumentar la coherencia del criterio aplicado, y reducir el tiempo final dedicado al proceso evaluador. Dichos parámetros han sido definidos por el profesorado de la asignatura de álgebra lineal y validados posteriormente por otros profesores (ver sección 3.3).

Homogeneidad (H)

Es un parámetro comprendido entre 0 y 1, que da idea de la homogeneidad, entre los miembros de un mismo grupo, de las notas obtenidas en todos los controles realizados. Es importante resaltar que en cada grupo conviven 3 alumnos con perfil diferente (un alumno con buenas notas previas de la asignatura, otro con notas previas intermedias, y otro con notas previas muy bajas), por lo que el valor de este parámetro, inicialmente, es muy similar en todos los grupos.

Las notas obtenidas en los últimos controles realizados en un parcial tienen una ponderación mayor en el cálculo de este parámetro que la obtenida en los primeros, ya que se considera que a medida que pasa el tiempo los alumnos de un mismo grupo deberían obtener notas similares si el trabajo en grupo a lo largo del parcial se ha llevado a cabo con éxito. La ecuación (1) muestra cómo se calcula el parámetro σ^2_{actual} necesario para calcular el valor final de H .

$$\sigma^2_{actual} = \lambda_1\sigma_1^2 + \lambda_2\sigma_2^2 + \dots + \lambda_n\sigma_n^2 \quad (1)$$

donde n representa el número de pruebas que se consideran, y σ_i^2 representa la varianza de las notas individuales de los miembros del grupo en el control i . El cálculo de los coeficientes de la combinación lineal se realiza según indica la expresión (2):

$$\lambda_i = \frac{i}{\sum_{j=1}^n j} \Rightarrow \sum_{i=1}^n \lambda_i = 1 \quad (2)$$

El parámetro de homogeneidad final H se calcula aplicando la ecuación (3) con el objetivo de adaptar la sensación subjetiva del profesor con el valor final de homogeneidad calculado.

$$H = \frac{1.5}{1.5 + \sigma} \quad (3)$$

Mejora (M)

Para el cálculo de este parámetro se toma la nota obtenida por el grupo cuando los miembros del mismo resuelven cada control de forma conjunta, y se compara con la nota más alta de los 3 miembros del grupo cuando ese mismo control fue realizado individualmente. Es decir, la mejora debida a un control concreto 'i' y el valor de mejora global tras la realización de N controles se calculan según se muestra en la ecuación (4).

$$\begin{aligned} \text{mejora}_i &= \text{notaEnGrupo}_i - \text{notaMaximaAlumnosGrupo}_i \\ M &= \frac{1}{N} \sum_{i=1}^N \text{mejora}_i \end{aligned} \quad (4)$$

Por tanto, este parámetro representa la mejora media de la nota obtenida en cada control realizado en grupo, con respecto a la mejor nota individual obtenida en cada control.

Asistencia (A)

Dado que los controles se realizan en el aula sin aviso previo, en ocasiones no se encuentran en clase los 3 miembros de un mismo grupo. Con este parámetro, cuyo valor está comprendido entre 0 y 1, se valora el grado de presencia de los miembros del grupo en todos los controles realizados. Es decir, un valor de '1' indicaría que todos los miembros del grupo estuvieron presentes en todos los controles realizados, mientras que un valor de '0,66' indicaría que, en promedio, sólo hubo dos miembros del grupo en cada control. Se calcula aplicando la ecuación (5).

$$A = \frac{1}{3N} \sum_{i=1}^N A_i \quad (5)$$

donde A_i representa el número de miembros del grupo que realizaron el control i , y N el número de controles considerado.

Promedio de las notas de grupo (P)

Este indicador se limita a calcular la nota media de los controles realizados en grupo (no se consideran los resultados obtenidos en los controles individuales).

3.3. Validación de los parámetros definidos

Introducción

No debe olvidarse cuál es el objetivo perseguido: evaluar el rendimiento del trabajo en grupo realizado por los estudiantes, a partir de los parámetros presentados en la sección anterior. Por tanto, antes de continuar es fundamental validar los parámetros definidos para poder construir con garantías el sistema evaluador automático a partir de ellos. Así, con el fin de conocer los criterios aplicados por los profesores cuando evalúan el rendimiento del trabajo en grupo, y confirmar que éstos quedan reflejados en los parámetros definidos con anterioridad, se propusieron 2 tests a 25 profesores universitarios pertenecientes a diferentes disciplinas dentro del mundo de la ingeniería. En el primero de ellos se pidió a los profesores que evaluaran el trabajo en grupo realizado por 11 grupos de 3 alumnos constituidos en el curso 2005-06 en la asignatura de álgebra lineal. Para ello se les facilitaron todas las notas disponibles, tanto individuales como de grupo, procedentes de todos los controles realizados. Al final del test cada evaluador indicó los criterios que había seguido en su evaluación y el tiempo invertido. En el segundo test se les pidió que evaluaran lo mismo, pero contando únicamente con el valor de los 4 parámetros definidos en la sección 3.2 y previamente calculados a partir de las notas obtenidas por los estudiantes en los controles. Es importante resaltar que los profesores no tuvieron acceso al segundo test antes de contestar el primero, para no condicionar sus respuestas.

Validación de los evaluadores

Antes de validar la viabilidad de los parámetros definidos es esencial evaluar la coherencia de criterio de los profesores cuando contestaron los dos tests, con el fin de utilizar datos fiables. Así, además de contener los datos ya mencionados correspondientes a los 11 grupos, ambos tests contenían 3 puntos de control (los datos correspondientes a tres grupos fueron presentados a los profesores dos veces en cada test) con la finalidad de evaluar la consistencia de criterio en los evaluadores. El error de coherencia EC en cada test para cada profesor se calculó según la ecuación (6):

$$EC = \frac{\sum_{i=1}^3 |nota1_i - nota2_i|}{\text{intervalo de notas utilizado}} \quad (6)$$

donde $nota1_i$ y $nota2_i$ representan las notas asignadas por el profesor dos veces para cada grupo duplicado, e *intervalo de notas utilizado* representa el rango de valores numéricos utilizado por el profesor cuando asigna las notas en el test.

Así, los profesores que obtuvieron en alguno de los dos tests un error de coherencia mayor a un umbral predefinido fueron descartados. Como resultado, 18 profesores superaron la condición impuesta, por lo que sus respuestas en ambos tests fueron consideradas como fiables.

Validación objetiva de los parámetros

Con el fin de validar la utilidad de los 4 parámetros definidos se comparó el resultado de la evaluación de los profesores considerados coherentes entre los dos tests realizados. Para ello, se calculó el coeficiente de correlación entre las notas asignadas en el test 1 y las asignadas en el test 2, obteniendo como resultado un valor de correlación $\rho=0.657$. El valor de correlación obtenido es considerado alto ($|\rho|>0.5$) cuando se aplica en el campo de la investigación psicológica [11] por lo que los parámetros se consideraron validados.

Evidentemente, si se desea ser más exigente en el proceso de validación de los parámetros definidos, sólo hay que aumentar el valor umbral de correlación. En caso de que el valor de correlación obtenido sea inferior al valor umbral utilizado, los parámetros deberían redefinirse nuevamente con la ayuda de los comentarios introducidos por los evaluadores en el test 1. Una vez redefinidos, el proceso de validación debería iniciarse nuevamente.

3.4. Obtención del sistema evaluador

Una vez que los parámetros definidos en la sección 3.2 fueron validados, se procedió a definir un sistema que evalúe automáticamente el rendimiento del trabajo en grupo de los alumnos a partir de estos 4 parámetros. El sistema debe aplicar en el proceso evaluador las reglas extraídas a partir de la evaluación realizada por todos los profesores considerados coherentes en el test 2, y debe permitir su interpretabilidad posterior con el fin de conocer con mayor profundidad cuál es el razonamiento aplicado por el sistema cuando evalúa. Es importante recordar que no deseamos un sistema que evalúe únicamente, sino que queremos obtener de él información que nos ayude a comprender las reglas que el profesorado aplica durante el proceso evaluador. Así, aunque existen diferentes herramientas para afrontar este problema (redes neuronales o algoritmos genéticos, entre otras) la imposición de esta interpretabilidad posterior del modelo obtenido hace que la herramienta más apropiada en esta situación sea la lógica difusa.

El método de modelado difuso aplicado ha sido el método FIIF (Fast, Incremental, Intelligible and Fuzzy). Éste fue propuesto por el Dr. Carles Garriga en el año 2005 [12], y es un método heurístico simple, inteligible e incremental concebido para tratar de forma sencilla los problemas inherentes en el diseño de un modelo difuso inteligible desarrollado a partir de un conjunto de datos que asocian diferentes entradas con una salida. FIIF llega a un modelo final a través de un método con un coste computacional bajo que, además, puede ser ajustado según la naturaleza del problema o las preferencias del usuario. En el ámbito del modelado difuso, algunos autores han desarrollado métodos que alcanzan modelos con una alta precisión, aunque a costa de dejar olvidadas las capacidades lingüísticas (inteligibilidad) de los sistemas difusos. De hecho, resulta extremadamente difícil la obtención de modelos con un alto nivel de inteligibilidad y precisión por lo que, con frecuencia, el modelado difuso se ha dividido en *modelado difuso preciso* y *modelado difuso lingüístico*. El método FIIF representa un equilibrio entre precisión e inteligibilidad, por lo que resulta especialmente interesante para ser utilizado en el contexto de este trabajo. Para más detalles sobre el método puede consultarse [12], donde además se compara el método FIIF con otros métodos de modelado difusos existentes y se detallan las ventajas que este método presenta cuando los criterios que se potencian son la inteligibilidad, la facilidad de uso del método y el coste computacional del proceso, sin sacrificar demasiado la precisión en el modelo obtenido.

Las evaluaciones realizadas por los profesores coherentes en el test 2 a los 11 grupos de alumnos fueron utilizadas para obtener un sistema evaluador automático aplicando el método FIIF. En el proceso de entrenamiento se utilizó el método de validación *leave-one-out-cross-validation*, de manera que se obtuvieron 11 sistemas intermedios previos al sistema final. Cada uno de esos sistemas intermedios fue entrenado con los datos correspondientes a 10 grupos, y después sometido a la fase de test con los datos del grupo restante. Para cada uno de los 11 sistemas intermedios se calcularon dos parámetros de error: el error de evaluación cometido por el sistema cuando se le introducían los datos utilizados en la fase de entrenamiento (TrE_i para valores de i entre 1 y 11), y el error cometido cuando se introducían los datos correspondientes al grupo de test (TE_i). Ambos errores fueron calculados utilizando el *NRMSE* (*Normalized Root Mean Square Error*) con respecto a la mediana de las notas asignadas por los profesores a cada uno de los 11 grupos.

El sistema difuso final fue obtenido utilizando en la fase de entrenamiento los datos de los 11 grupos disponibles. Los errores asociados al sistema final se estimaron a partir de la media de los errores obtenidos en los 11 sistemas intermedios. Estos valores fueron $TrE=0.7849$ and $TE=1.2108$ (sobre diez). La evaluación que realizó el sistema final para cada grupo aparece marcada con círculos en la figura 1, donde aparecen también las distribuciones de las notas asignadas a cada grupo por los diferentes profesores (representadas como *boxplots*) y la evaluación realizada, en las fases de entrenamiento y test correspondientes, por los diferentes sistemas difusos intermedios obtenidos durante el proceso.

El sistema difuso final obtenido tras la aplicación del método FIIF queda determinado por los conjuntos difusos definidos para cada una de las 4 variables de entrada y para la variable de salida (ver

figura 2) y por las reglas difusas aplicadas finalmente (ver tabla 1). Si se desea profundizar en los detalles del sistema presentado, consultar [13].

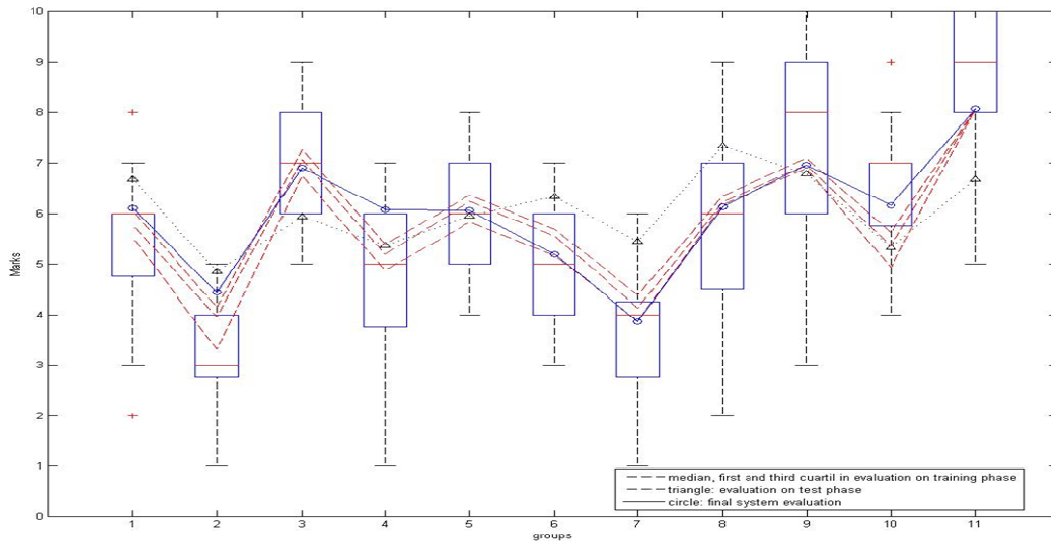


Figura 1: Evaluación realizada por el sistema difuso.

Tabla 1: Reglas difusas que definen el sistema evaluador final.

Identificador	Homogeneity	Improvement	Attendance	Average	Conjunto difuso de salida
1	m	h	h	h	A
2	l	m	h	l	B
3	l	m	h	h	B
4	l	h	h	l	B
5	l	h	h	m	B
6	l	h	h	h	B
7	m	h	h	l	B
8	m	h	h	m	B
9	l	m	l	m	C
10	l	m	l	h	C
11	l	h	l	m	C
12	l	h	l	h	C
13	m	m	l	h	C
14	m	m	h	m	C
15	m	m	h	h	C
16	m	h	l	m	C
17	m	h	l	h	C
18	h	m	h	l	C
19	h	m	h	m	C
20	h	h	h	l	C
21	h	h	h	m	C
22	l	m	h	m	D
23	m	m	l	m	D
24	m	m	h	l	D
25	m	h	l	l	D
26	h	m	l	l	D
27	h	m	l	m	D
28	h	h	l	l	D
29	h	h	l	m	D
30	l	l	l	h	E
31	l	l	h	h	E
32	m	l	l	h	E
33	l	l	l	l	F
34	l	l	l	m	F
35	l	l	h	m	F
36	l	m	l	l	F
37	m	l	l	l	F
38	m	l	l	m	F
39	m	l	h	m	F
40	m	l	h	h	F
41	m	m	l	l	F

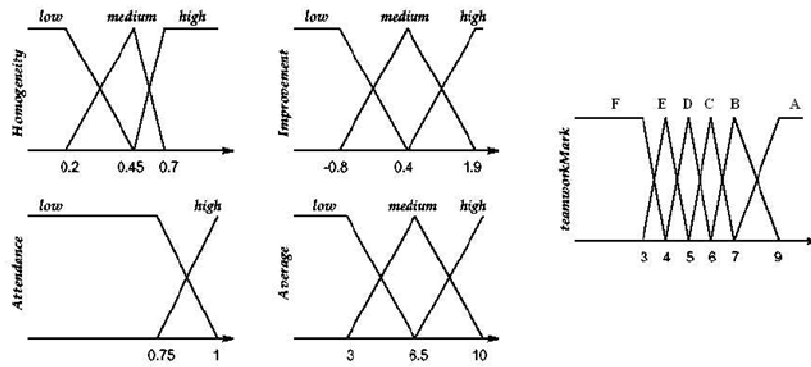


Figura 2: Conjuntos difusos definidos en el sistema difuso final

4. Conclusiones y líneas de futuro

Conviene recordar que la inminente entrada en vigor del EEES (Espacio Europeo de Educación Superior) trae consigo la necesidad de incorporar a las clases universitarias metodologías docentes activas que desarrollen y potencien nuevas competencias en los estudiantes. Esas competencias deben ser evaluadas de algún modo, aunque la naturaleza subjetiva de algunas de ellas, la falta de recursos y la falta de práctica de muchos profesores en la aplicación de metodologías de carácter constructivista, dificulta esa evaluación. Por ese motivo, resulta especialmente interesante la aplicación de un proceso sistemático de diseño como el propuesto en este artículo para obtener herramientas automáticas que permitan evaluar competencias de esta naturaleza de forma justa (del mismo modo para todos los estudiantes) adaptándose a los recursos disponibles.

En este artículo se ha presentado un sistema automático que evalúa el rendimiento del trabajo en grupo de los alumnos a partir de datos objetivos, como son las notas obtenidas por éstos en los controles (tanto individuales como en grupo) realizados a lo largo del curso. Para aplicar dicho sistema es importante seguir la política de creación de grupos adecuada, que garantiza homogeneidad entre los diferentes grupos y heterogeneidad dentro de cada uno de ellos.

El sistema de evaluación automática obtenido ha sido generado a partir del razonamiento aplicado por diferentes profesores universitarios en el momento de evaluar el rendimiento del trabajo en grupo realizado por los estudiantes. Se han definido 4 parámetros (validados experimentalmente) que concentran la mayor parte de la información que los profesores perciben tras observar todos los datos disponibles. A partir de ellos, y de la evaluación realizada por los profesores, se han aplicado técnicas de modelado difuso para obtener un sistema evaluador interpretable capaz de evaluar automáticamente la competencia considerada.

El interés del sistema diseñado es doble, ya que permite:

- garantizar que todos los grupos de alumnos son evaluados por el profesor aplicando los mismos criterios, de forma rápida y equitativa. Además, en situaciones en las que varios profesores comparten una misma asignatura y desean aplicar el mismo criterio evaluador, la utilización de un sistema evaluador automático como el generado en este trabajo permite que la evaluación de esta competencia se realice del mismo modo, independientemente del profesor del que dependa cada grupo.
- conocer mejor el razonamiento aplicado por el profesorado (como ente global) cuando se evalúa el rendimiento del trabajo en grupo. Ese conocimiento procede del análisis detallado de las reglas difusas que controlan el sistema difuso implementado, y nos permite descubrir la importancia que tiene cada uno de los criterios aplicados en el proceso evaluador.

Como líneas de trabajo futuro se proponen: (i) la generación de nuevos sistemas difusos que evalúen otras competencias de carácter subjetivo, mediante la aplicación del procedimiento de diseño seguido en este artículo; y (ii) el estudio de estrategias de validación de los sistemas difusos obtenidos con el fin de ajustar las reglas del sistema final y evitar así problemas derivados de un entrenamiento realizado a partir de un conjunto de datos reducido.

Actualmente nos encontramos en un momento en el que muchos profesores, ante la inminente entrada en vigor del EEES, se plantean cómo pueden evaluar de forma explícita competencias que hasta ahora no habían evaluado. El trabajo presentado en este artículo pretende ser una posible respuesta a esta necesidad.

5. Referencias

- [1] VIVARACHO, C., SIMÓN, A., PRIETO, O. *Ampliación de una experiencia de aprendizaje cooperativo a varios grupos de primero y con aulas masificadas*. En *Actas V Jornadas sobre Aprendizaje Cooperativo*, 2005.
- [2] MARTÍNEZ, F., HERRERO, L. C., GONZÁLEZ, J.M., DOMÍNGUEZ, J.A. *Project based learning experience in industrial electronics and industrial applications design*. En *Proceedings of The International seminar on innovative teaching and learning in engineering education*, Valladolid (España), 2006, pp. 299-312.
- [3] FREEMAN, MARK, MCKENZIE, JO. *Spark, a confidential web-based template for self and peer assessment of student teamwork: benefits of evaluating across different subjects*. *British Journal of Educational Technology*, 2002, vol. 33, pp. 551-569.
- [4] KINSHUK, A., PATEL, A. *Adaptative tutoring in business education using fuzzy backpropagation approach*. En *Proceedings of the 9th International Conference on Human-Computer Interaction*, New Orleans (USA), 2001.
- [5] KAVCIC, A., PEDRAZA, R. MOLINA, H. VALVERDE, F. *Student modelling base don fuzzy inference mechanisms*. En *Proceedings of The International Conference on Computer as a tool*, 2003.
- [6] NYKANEN, O. *Inducing fuzzy models for student classification*. *Educational Technology and society*, 2006, vol. 9(2), pp. 223-234.
- [7] CHENG, P., KILIS, D., KNIGHT, G. *Knowledge assessment using fuzzy conceptual representation*. En *Proceedings of the ACM symposium on Applied Computing*, 1997, pp. 3-9.
- [8] MA, J., ZHOU, D. *Fuzzy set approach to the assessment of student-centered learning*. *IEEE Transactions on Education*, 2000, vol. 43(2), pp. 237-241.
- [9] WANG, LI-XIN. *A course in fuzzy systems and control*. Prentice-Hall, 1997.
- [10] MONTERO, J.A., MORÁN, J.A., GÓMEZ, J., ALÍAS, F., VICENT, L., BADIA, D. *Moving from classical teaching to active learning in algebra for engineers*. En *Proceedings of the International Seminar on innovative teaching and learning in engineering education*, Valladolid (España), 2006, pp. 313-332.
- [11] COHEN, J. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, second edition, 1988.
- [12] GARRIGA, C. *A new approach to the synthesis of fuzzy systems from input-output data*. PhD thesis, Universidad Ramon Llull, octubre, 2005.
- [13] MONTERO, JOSÉ A. *Hacia una metodología docente basada en el aprendizaje activo del estudiante presencial de ingeniería, compatible con las exigencias del EEES*. PhD thesis, Universidad Ramon Llull, febrero, 2008.