

Construcción de Clasificadores Automáticos de Habilidades de E-Tutores de Aprendizaje Colaborativo

Pablo Santana Mansilla ^(1,2), Rosanna Costaguta ⁽²⁾, Daniela Missio ⁽²⁾

(1) CONICET, Comisión Nacional de Investigaciones Científicas y Técnicas

(2) Instituto de Investigación en Informática y Sistemas de Información (IISI), Departamento de Informática, Facultad de Ciencias Exactas y Tecnologías (FCEyT), Universidad Nacional de Santiago del Estero (UNSE), Avda. Belgrano (S) 1912, Santiago del Estero, 4200, Argentina
{psantana, rosanna, dmissio}@unse.edu.ar

Resumen. El Aprendizaje Colaborativo Soportado por Computadora (ACSC) permite a los estudiantes aprender por medio de la interacción con personas localizadas en marcos temporales y espaciales diferentes. Disponer de software especializado que apoye el aprendizaje en grupo, no garantiza que los estudiantes colaboren, porque no es la tecnología utilizada lo que soporta la colaboración sino la forma en que la utilizan los e-tutores (docentes) para coordinar la interacción grupal. Teniendo en cuenta que el e-tutor es clave para el éxito del ACSC, y que poco se sabe sobre como los docentes intervienen en las actividades de aprendizaje colaborativo de los estudiantes, se creó en primera instancia un esquema de clasificación de habilidades de e-tutores de ACSC. Considerando que el análisis manual de las interacciones registradas en entornos de ACSC requiere considerable tiempo y esfuerzo, en segunda instancia se desarrollaron y evaluaron métodos automáticos de análisis de interacciones y reconocimiento de habilidades.

Palabras Clave: Aprendizaje Colaborativo Soportado por Computadora, e-tutor, clasificación de habilidades de e-tutor, reconocimiento automático de habilidades.

1 Introducción

Con el rápido desarrollo de la sociedad basada en conocimiento ha crecido la importancia que se le da a la creación colaborativa de conocimientos. En este contexto, el ACSC se ha vuelto una nueva forma de educación donde los estudiantes aprenden por medio de la interacción con personas que pueden estar localizadas en marcos temporales y espaciales diferentes [1]. Por otro parte, las investigaciones realizadas demuestran que los sistemas de ACSC brindan un entorno apropiado para el desarrollo de habilidades tales como solución de problemas, pensamiento crítico, establecimiento de metas, interpretación, y análisis [1, 2]. Sin embargo, disponer de herramientas de software que soporten el aprendizaje en grupo no garantiza que los estudiantes colaboren, porque no son las características de la tecnología utilizada sino la forma en que se la utiliza lo que soporta la colaboración [3, 4, 5].

Según [6] y [7] tanto la calidad como la profundidad del aprendizaje dependen de la forma en que la colaboración es coordinada pero, la mayoría de los e-tutores no están familiarizados con las técnicas que se requieren en ACSC, ni se conocen con precisión las habilidades que deberían poseer para desempeñarse adecuadamente [5, 7, 8]. Si además se tiene en cuenta que las iniciativas de ACSC tienen pocas chances de éxito sin docentes con habilidades para sacarle provecho a las herramientas tecnológicas disponibles [5, 7], es evidente la necesidad de plantear mecanismos que les permitan a los e-tutores adquirir las habilidades necesarias para desempeñarse adecuadamente en entornos de ACSC.

De acuerdo con [9] una manera de propiciar la adquisición de habilidades consiste en desarrollar sistemas de software que entrenen a los e-tutores. No obstante, tanto la falta de conocimiento sobre las habilidades requeridas en un e-tutor colaborativo, como los efectos negativos sobre la comunicación que presentan las técnicas de análisis de interacciones basadas en interfaces estructuradas o semi estructuradas [10, 11], constituyen los principales obstáculos a vencer para desarrollar aplicaciones que permitan a los e-tutores adquirir las habilidades que manifiestan con deficiencia.

En el marco del proyecto de investigación “Sistemas de información web basados en agentes para promover el Aprendizaje Colaborativo Soportado por Computadoras” (CICyT-UNSE Código 23/C097), se dieron varios pasos tendientes a propiciar el futuro desarrollo de sistemas informatizados que entrenen a los e-tutores de ACSC. Primero, se definió un esquema de clasificación de las habilidades que debería poseer un e-tutor de ACSC. Luego, de forma complementaria a la taxonomía de habilidades creada, se aplicaron técnicas de minería de textos para construir clasificadores con la capacidad de identificar automáticamente las habilidades manifestadas por e-tutores al interactuar con sus alumnos. El trabajo realizado se documenta en el presente artículo que se estructura como sigue. La próxima sección describe brevemente el esquema de clasificación de las habilidades que deberían poseer los e-tutores de ACSC. La sección 3 presenta las dos maneras en que puede realizarse el análisis de las interacciones en entornos de ACSC (manual y automático). La sección 4 describe la aplicación de la técnica de análisis de contenido. La sección 5 comenta las adaptaciones realizadas a GATE¹ (la herramienta de minería de textos empleada) a fin de procesar textos en español. La sección 6 especifica los pasos metodológicos ejecutados para construir los clasificadores, la experimentación realizada, y los resultados obtenidos con diferentes algoritmos de clasificación. Por último, en la sección 7 se enuncian algunas conclusiones.

2 Clasificación de Habilidades de E-Tutores

A pesar de que las habilidades propias de un e-tutor de ACSC no han sido descritas detalladamente, entre los investigadores que han estudiado esta temática se pueden identificar dos enfoques. Por un lado, se proporcionan guías y recomendaciones para e-tutores de ACSC pero sin brindar mayores precisiones. Por el otro lado, se analiza el rol de los e-tutores de ACSC tomando como base el esquema de organización por roles, heredado de la literatura sobre habilidades y competencias para docentes de

¹ <http://gate.ac.uk/>

e-learning que no recurren al aprendizaje colaborativo como estrategia pedagógica sino que se basan en el modelo de interacción docente-estudiante. Como parte de la creación del esquema de clasificación de habilidades de docentes de ACSC en [12] se integraron los dos enfoques citados, teniendo en cuenta además los planteos pertenecientes al área del *e-learning* no colaborativo. Con el propósito de facilitar la comprensión de la labor de los docentes de ACSC se organizó la taxonomía en tres niveles (Figura 1a). El primer nivel responde a las habilidades vinculadas con los roles que desempeñan los e-tutores al coordinar el aprendizaje: Administrativo, Pedagógico, Social, Técnico, Comunicación y Evaluación (Figura 1b). El segundo nivel considera las sub habilidades relacionados con cada habilidad, y finalmente, el tercer nivel especifica atributos para cada sub habilidad.

Si bien en la Figura 1b se muestran seis categorías de habilidad, se trabajó sólo con las habilidades sociales para evaluar la viabilidad del uso de la minería de textos en el reconocimiento automático de habilidades.

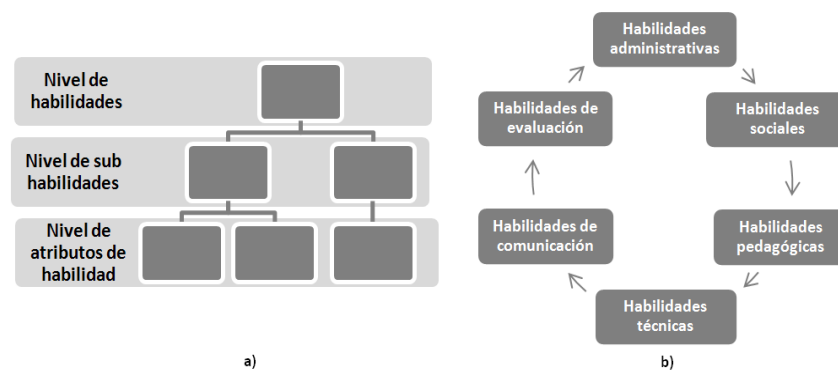


Fig. 1. Niveles y categorías de la clasificación de habilidades.

3 Análisis de Interacciones en ACSC

Tomando como base la clasificación de habilidades descrita en [12] se podrían analizar las contribuciones docentes en un ambiente ACSC a fin de identificar las habilidades manifestadas. Pero, tal proceso de análisis no es sencillo ya sea que se realice de forma manual o automática. Por una parte, cuando se tiene una cantidad considerable de interacciones el análisis manual es prácticamente inviable debido al tiempo y el esfuerzo que demanda [13]. Por otra parte, en el análisis automático de la información registrada por un sistema de ACSC se ha recurrido tanto a interfaces estructuradas o semi estructuradas como a texto libre para modelar las interacciones. Las interfaces estructuradas y semi estructuradas (oraciones de apertura, contribuciones dirigidas por menú, y contribuciones basadas en diagramas) simplifican el proceso de análisis al no tener que usar técnicas de procesamiento de lenguaje natural [10, 11, 14]. Sin embargo, este tipo de interfaces socava el proceso de colaboración porque restringe las posibilidades de interacción, provoca que la

comunicación sea lenta, y crea stress relacional [3, 10, 11]. En relación al texto libre, al permitir que la comunicación se realice sin restricciones ya no es posible usar la interfaz para hacer inferencias sobre el proceso de interacción, y por consiguiente, es necesario recurrir a técnicas de procesamiento de lenguaje natural. Hasta el momento, este tipo de tecnología se usó de manera limitada en tareas como la clasificación de los tópicos de una conversación y la caracterización de patrones de discusión [14].

Debido a las deficiencias asociadas con el análisis manual y con el análisis automático de interacciones estructuradas o semi estructuradas, surgió la necesidad de contar con una técnica capaz de reconocer las habilidades que poseen los e-tutores sin afectar negativamente la dinámica de trabajo de los estudiantes o docentes, ni representar un alto costo o sobrecarga. En este contexto la minería de textos (en particular sus técnicas de clasificación) se vislumbró como una alternativa para cumplir con las condiciones mencionadas, dada su capacidad para manejar la incertidumbre, borrosidad, diversidad de estructuras, y gran cantidad de palabras que caracterizan al lenguaje natural [15].

4 Aplicación de Análisis de Contenido

Para obtener el conjunto de datos (entrenamiento y prueba) que se usó para aplicar minería de textos, se realizaron varias experiencias de ACSC en el entorno de aprendizaje Moodle² con docentes y estudiantes de la FCEyT de la UNSE. En pos de incrementar la cantidad de datos usados en la construcción de los clasificadores se solicitó la colaboración de investigadores que participaron del proyecto Tactics [16], quienes brindaron acceso a las sesiones de chat y comunicaciones vía foro de las experiencias de ACSC en las que estuvieron involucrados. Sin embargo, las interacciones por sí mismas no son suficientes para construir los clasificadores porque también se necesitan conocer los atributos de habilidad manifestados por los e-tutores. Por consiguiente, sobre el conjunto de interacciones se aplicó análisis de contenido (siguiendo el planteamiento metodológico de Krippendorff [17]). Así, expertos humanos (un psicopedagogo y un especialista en ACSC) hicieron corresponder un atributo de habilidad social de la clasificación comentada en la sección 2 a cada oración que formaba parte de los mensajes o contribuciones de los docentes. Cabe aclarar que los expertos no tuvieron que catalogar todos los mensajes publicados por los e-tutores sino solamente aquellos pre seleccionados mediante la técnica de muestreo de relevancia [17] por considerar que contenían al menos una oración donde se manifestaba un atributo de habilidad social. El nivel de acuerdo entre los expertos humanos (para las 891 oraciones resultantes) fue de $\alpha = 0.938$. Si bien este valor es superior al 0.80 recomendado por Krippendorff para aceptar los resultados de un estudio de análisis contenido como fiables, no representaba un nivel de acuerdo perfecto ($\alpha = 1$) sino que indicaba que existían oraciones a las que los expertos asignaron atributos de habilidad diferentes. Para resolver estas discrepancias se realizó un segundo proceso de análisis donde los expertos unificaron criterios y

² <https://moodle.org/>

acordaron el atributo de habilidad que le correspondía a cada oración sobre la que existía desacuerdo.

5 Adaptaciones Realizadas a GATE

Teniendo en cuenta que en cada mensaje o contribución de los e-tutores se pueden manifestar múltiples habilidades, como paso previo a la construcción de los clasificadores, fue necesario dividir en oraciones a los mensajes. La descomposición de los mensajes de los e-tutores en oraciones tiene influencia directa en los resultados que se obtienen con los clasificadores de habilidades ya que una identificación errónea de oraciones llevaría a que las habilidades no se reconozcan apropiadamente. Si bien GATE cuenta con recursos de procesamiento para identificar oraciones en un documento, los mismos fueron creados para ser aplicados a textos en inglés y por consiguiente fue preciso adaptarlos para que trabajen correctamente al procesar textos en español.

GATE ofrece dos recursos de procesamiento para la segmentación de textos en oraciones: *Regex Sentence Splitter* y *ANNIE Sentence Splitter* [18]. El primero de estos tuvo que ser descartado ya que no se pueden reconocer correctamente oraciones cuando se tiene abreviaturas seguidas de signos de puntuación. En [19] se brindan detalles sobre las reglas que se crearon de modo tal que mediante *ANNIE Sentence Splitter* se puedan identificar oraciones en textos en español. Las *stop words* son palabras de uso frecuente en un lenguaje (artículos, preposiciones, conjunciones, pronombres, etc.) que pueden ser descartadas porque tienen muy poco contenido de información para distinguir entre categorías y una capacidad predictiva escasa [15, 20]. GATE no cuenta con un recurso dedicado exclusivamente a la supresión de *stop words* por lo cual en [19] también se describe como implementar esta operación de pre procesamiento mediante *JAPE Transducer* y *ANNIE Gazetteer* [18].

6 Construcción de los Clasificadores

La minería de textos o *text mining* se refiere al proceso de extracción de patrones interesantes y no triviales o conocimiento desde documentos de texto [15, 21]. La minería de textos intenta revelar la información oculta por medio de métodos que son capaces de tratar con la vaguedad, incerteza, borrosidad, y gran cantidad de palabras y estructuras que caracterizan al lenguaje natural [15]. Si bien la minería de textos utiliza técnicas de minería de datos, en el proceso de descubrimiento de conocimiento la minería de textos parte de datos textuales no estructurados mientras que la minería de datos se aplica sobre bases de datos estructuradas [22]. Las técnicas de clasificación de la minería de textos, que consisten en asignar objetos a categorías predefinidas, se adecuan naturalmente al problema de identificar las habilidades manifestadas por los profesores de ACSC, porque la intención es vincular cada contribución de los e-tutores con una o más habilidades de la clasificación propuesta en [12].

Realizadas las adaptaciones a GATE que se comentaron en la sección 5, el paso siguiente en la investigación consistió en construir los clasificadores siguiendo la

metodología CRISP-DM [23]. Si bien CRISP-DM fue propuesta para guiar el desarrollo de proyectos de minería de datos, puede usarse para abordar problemas de minería de textos en la medida que los datos textuales se transformen en un formato estructurado o semi estructurado [24].

La primera tarea de preparación de los registros de las sesiones de ACSC para la minería de textos consistió en convertirlos a texto plano. Las sesiones de ACSC estaban almacenadas en 4 formatos de archivo (htm, rtf, doc, y pdf) y para evitar problemas de compatibilidad en la representación de caracteres los mensajes muestreados se guardaron en archivos de texto plano con un archivo por mensaje. A continuación se aplicaron las siguientes tareas de limpieza de datos: corrección de errores de ortografía, reemplazo de emoticones por caracteres Unicode equivalentes, supresión de archivos duplicados, y eliminación de archivos que contenían solo nombres de lugares o de personas por cuanto no brindan información relevante para la clasificación.

Los sistemas de minería de textos no aplican sus algoritmos de descubrimiento de conocimiento a colecciones de documentos no estructurados por lo tanto, es necesario recurrir a operaciones de pre procesamiento o de preparación de datos de manera de transformar documentos de texto no estructurados en un formato intermedio estructurado más explícitamente [20, 22]. Las operaciones de pre procesamiento se centran en la identificación, extracción, refinamiento y adición de características representativas de los documentos en lenguaje natural de modo tal que, las características más representativas sean usadas para el *text mining* mientras que las restantes descartadas [22]. En el presente trabajo de investigación se recurrió a las operaciones de pre procesamiento: identificación de token, identificación de oraciones, normalización (*inflectional stemming* y lematización), *Part-of-Speech Tagging* (POST), y supresión de *stop words* [15, 20, 22]. Estas operaciones no se aplicaron de manera aislada sino que se combinaron para determinar si tenían influencia en la efectividad de los algoritmos de clasificación. La Tabla 1 muestra las combinaciones posibles de las cinco operaciones de pre procesamiento utilizadas.

Tabla 1. Combinaciones de operaciones de pre procesamiento

| Combinaciones | Identificación de Token | Identificación de oraciones | Eliminación de <i>stop words</i> | Lematización | <i>Stemming to a root</i> | POST |
|-----------------------|-------------------------|-----------------------------|----------------------------------|--------------|---------------------------|------|
| Pre procesamiento 1 | x | x | | | | |
| Pre procesamiento 2 | x | x | x | | | |
| Pre procesamiento 3 A | x | x | | x | | |
| Pre procesamiento 3B | x | x | | | x | |
| Pre procesamiento 4 | x | x | | | | x |
| Pre procesamiento 5 | x | x | x | | | x |
| Pre procesamiento 6 A | x | x | x | x | | |
| Pre procesamiento 6 B | x | x | x | | x | |
| Pre procesamiento 7 A | x | x | | x | | x |
| Pre procesamiento 7 B | x | x | | | x | x |
| Pre procesamiento 8 A | x | x | x | x | | x |
| Pre procesamiento 8 B | x | x | x | | x | x |

Cuando un clasificador se construye mediante minería de textos se necesita que un conjunto de documentos manualmente clasificados por expertos del dominio, sea dividido en dos subconjuntos, uno de entrenamiento y otro de prueba. El clasificador se construye mediante un proceso inductivo, donde observando las características del subconjunto de entrenamiento se infieren las condiciones que documentos previamente no examinados deberían cumplir para ser clasificados bajo una u otra categoría [25]. Luego, la comparación de las decisiones de categorización realizadas por el clasificador con las efectuadas por expertos humanos (sobre el subconjunto de prueba) permite evaluar la efectividad de la clasificación automática. En el marco de la presente investigación, para evaluar la efectividad de los clasificadores en el reconocimiento de habilidades, se calcularon las métricas de precisión y recall con validación cruzada 10-fold [25]. La precisión indica el porcentaje de oraciones clasificadas correctamente entre todas las oraciones a las que se les asignó un atributo de habilidad, independientemente de si el clasificador no asignó atributos de habilidad a oraciones que debieron ser clasificadas [18, 22]. Por su parte, el recall señala el porcentaje de oraciones clasificadas correctamente entre todas las oraciones a las que se les debía asignar un atributo de habilidad, independientemente de cuantas clasificaciones erróneas se realizaron [18, 22].

Para la construcción de los clasificadores se utilizaron los algoritmos KNN, SVM, PAUM, Naïve Bayes y C4.5 [20, 22, 25, 26]. Un análisis detallado del desempeño de cada uno de los cinco algoritmos mencionados se brinda en [27], pero cabe resaltar que los clasificadores con los niveles más altos de precisión se obtuvieron con los algoritmos SVM y PAUM. En el caso del algoritmo SVM se decidió recurrir a una variante conocida como SVM con márgenes desiguales [28] puesto que para todos los atributos de habilidad social la cantidad de ejemplos de entrenamiento positivos era pequeña en relación a la cantidad de ejemplos negativos. Se probaron solo 4 valores del parámetro τ (0.3, 0.4, 0.5 y 0.6) puesto que fuera de los mismos la precisión o el recall asumen valores muy bajos. Los experimentos realizados combinando los 4 valores de τ con las diversas operaciones de pre procesamiento (cuyos resultados se resumen en Figura 2) indican que la preparación de datos tiene una influencia variable sobre la efectividad de la identificación de habilidades. Mientras que la eliminación *stop words* y POST tienen un efecto casi imperceptible sobre la precisión (Figura 2a) y el recall (Figura 2b), la normalización puede tanto ocasionar una disminución como un incremento de la precisión dependiendo de si aplica *inflectional stemming* o lematización. Por su parte, el algoritmo PAUM utiliza un margen positivo (τ_{+1}) y un margen negativo (τ_{-1}) para hacer frente a problemas de clasificación donde los ejemplos positivos son escasos en relación a los ejemplos negativos. La Figura 3 muestra las medidas de efectividad de solo cuatro combinaciones de valores de los márgenes puesto que son los que permitieron: obtener el valor más alto de precisión ($\tau_{+1}=-1$ y $\tau_{-1}=-1.5$), llegar al valor más alto de recall ($\tau_{+1}=10$ y $\tau_{-1}=0.5$), y tener la menor diferencia entre recall y precisión ($\tau_{+1}=5$ y $\tau_{-1}=0$ al igual que $\tau_{+1}=10$ y $\tau_{-1}=-0.5$). Se pudo comprobar que las operaciones de pre procesamiento no incrementan la precisión sino que tienden a disminuirla (Figura 3a), mientras que el recall permanece relativamente estable (Figura 3b). En la Figura 3 también se aprecia que para valores positivos de τ_{+1} no existen diferencias significativas entre las cifras de precisión y recall, pero para valores negativos de τ_{+1} la precisión supera ampliamente al recall.

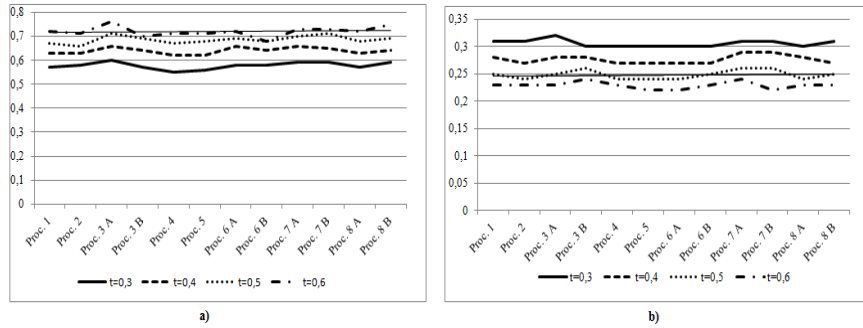


Fig. 2. Precisión y recall micro promediados para los clasificadores construidos con SVM

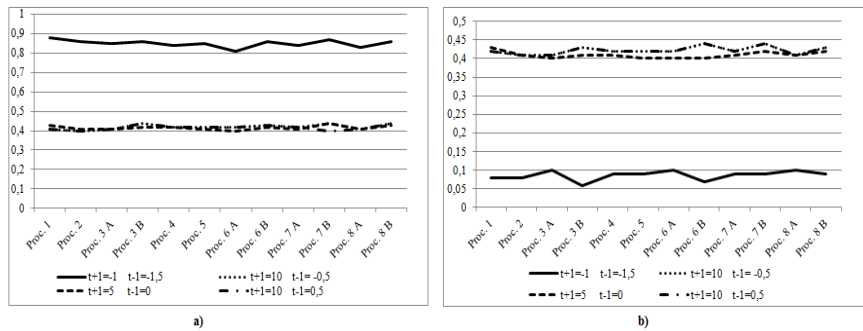


Fig. 3. Precisión y recall micro promediados para los clasificadores construidos con PAUM

7 Conclusiones

Durante la creación del esquema de clasificación de habilidades de e-tutores de ACSC quedó en evidencia que las distintas facetas del trabajo de un e-tutor (social, administrativa, pedagógica, técnica, comunicacional, y de evaluación) están vinculadas y es difícil separarlas. La taxonomía de habilidades se planteó con la intención de mejorar la comprensión de la tarea de los e-tutores pero, de ninguna manera puede considerarse definitiva sino que, por ser el fruto de un trabajo interdisciplinario con la pedagogía, se piensa que la clasificación podrá ser tomada como punto de partida por los profesionales del área interesados en realizar futuras investigaciones que permitan su validación científica. En este sentido, los resultados del análisis de contenido son auspiciosos porque en las experiencias de ACSC analizadas los expertos humanos reconocieron el 75% de los atributos de habilidad pertenecientes a la categoría social. Esto significa que en la práctica realmente se manifiestan los comportamientos y las conductas contempladas en la taxonomía de habilidades elaborada.

Teniendo en cuenta que los clasificadores automáticos construidos con los algoritmos PAUM y SVM permiten identificar habilidades en los mensajes de los

e-tutores con niveles de precisión superiores al 0.7, se puede afirmar que la minería de textos es una alternativa viable para conocer el desempeño de los e-tutores de ACSC sin demandar el tiempo y el esfuerzo requeridos para un análisis manual de las interacciones. A pesar de que se necesita seguir trabajando para mejorar el recall, los valores de precisión logrados son más que destacables considerando que, la cantidad de ejemplos utilizada para construir los clasificadores es muy inferior a la cantidad empleada en otros trabajos de investigación donde se utiliza minería de textos para tareas de clasificación. Así por ejemplo, en [26] con la colección Mod-Apte (una muestra de Reuters-21578 con 12902 documentos) se reportaron valores de precisión de 0.75, mientras que en la presente investigación se lograron niveles de precisión mayores al 0.7 con apenas 891 oraciones. Por otro lado, la experimentación realizada permite afirmar que la influencia de las operaciones de pre procesamiento sobre la efectividad del reconocimiento de habilidades depende del algoritmo de clasificación. Para obtener resultados verdaderamente concluyentes, se cree conveniente incrementar el número de interacciones, y también ampliar las categorías a reconocer ya que en esta investigación el reconocimiento se restringió a los atributos de la habilidad social.

Por último, los resultados obtenidos son alentadores sobre el uso de la minería de textos como parte de un futuro sistema software de entrenamiento personalizado de habilidades para e-tutores de ACSC. Una aplicación de este tipo tendría que analizar las interacciones grupales de modo de identificar conflictos que requieran la intervención de los e-tutores para resolverse. En estos casos, el sistema se encargaría de sugerir a los e-tutores las acciones a llevar a cabo para mejorar el aprendizaje grupal y simultáneamente practicar las habilidades que no hayan manifestado adecuadamente.

Referencias

1. Suh, H., Lee, S.: Collaborative Learning Agent for Promoting Group Interaction. *ETRI Journal*, 28(4), 461--474 (2006)
2. Day, T.W.: Online Collaborative Learning and Leadership Development. En: Rogers, P., Berg, G., Boettcher, J., Howard, C., Justice, L., Schenk, K. (eds) *Encyclopedia of Distance Learning*, Second Edition, pp. 1488--1492. Information Science Reference (2009)
3. Olivares, O. J.: Collaborative vs. Cooperative Learning: The Instructor's Role in Computer Supported Collaborative Learning. En: Orvis, K. L., Lassiter, A. L.R (eds) *Computer-Supported Collaborative Learning: Best Practices and Principles for Instructors*, pp. 20--39. Information Science Publishing, USA (2007)
4. Onrubia, J., Engel, A.: The role of teacher assistance on the effects of a macro-script in collaborative writing tasks. *International Journal of Computer-Supported Collaborative Learning*, 7(1), 161--186 (2012)
5. Orvis, K.L., Lassiter, A.L.R.: Computer-Supported Collaborative Learning: The Role of the Instructor. En: Ferris, S., Godar, S.H. (eds) *Teaching and learning with virtual teams*, pp. 158--179. Information Science Publishing (2006)
6. Borges, M.A.F., Baranauskas, M.C.C.: CollabSS: a Tool to Help the Facilitator in Promoting Collaboration among Learners. *Educational Technology & Society*, 6 (1), 64--69 (2003)
7. Kukulska-Hulme, A.: Do Online Collaborative Groups Need Leaders? En: Roberts, T.S. (ed) *Online Collaborative Learning: Theory and Practice*, pp. 262--280. Information Science Publishing (2004)

8. Greiffenhagen, C.: Making rounds: The routine work of the teacher during collaborative learning with computers. *International Journal of Computer-Supported Collaborative Learning*, 7(1), 11–42 (2012)
9. Barker P.: Skill set for online teaching. *Word Conference on educational, Multimedia, Hipermedia, & Telecomunicatios* (2002)
10. Chen, W.: Supporting teachers' intervention in collaborative Knowledge building. *Journal of Network and Computer Applications*, 29, 200–215 (2006)
11. Soller, A., Martínez, M. A., Jermann, P., Muehlenbrock, M.: From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning. *International Journal of Artificial Intelligence in Education*, 15 (4), 261–290 (2005)
12. Santana Mansilla, P., Costaguta, R., Missio, D.: Habilidades de E-tutores en Grupos Colaborativos. En: Peñaranda, N., Zazarini, S., Bejarano, I. F. (eds) *Experiencias Innovadoras en Investigación Aplicada*, pp. 687–704. Ediciones DASS-UCSE, Jujuy (2012)
13. Rosé, C., Wang, Y., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F.: Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3 (3), 237–271 (2008)
14. Tchounikine, P., Rummel, N., McLaren, B.M.: Computer Supported Collaborative Learning and Intelligent Tutoring Systems. En: Nkambou, R., Bourdeau, J., Mizoguchi, R. (eds) *Advances in Intelligent Tutoring Systems. SCI*, vol. 308, pp. 447–463. Springer, Heidelberg (2010)
15. Hotho, A., Nürnberger, A., Paaß, G.: Brief survey of text mining. *LDV Forum GLDV Journal for Computational Linguistics and Language Technology*, 20 (1), 19–62 (2005)
16. Juárez Pacheco, M.: Recomendaciones para el uso académico de herramientas web gratuitas. *Revista Mexicana de Investigación Educativa*, 10 (25), pp. 577–584 (2005)
17. Krippendorff, K.: *Content analysis: an introduction to its methodology*, Second Edition. SAGE Publications, USA (2004)
18. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., Gorrell, G., et al.: *Developing Language Processing Components with GATE Version 7 (a User Guide)*. Departamento de Ciencias de la Computación, Universidad de Sheffield, Reino Unido (2013)
19. Santana Mansilla P., Costaguta R., y Missio D.: Uso de la Arquitectura GATE para la Identificación de Sentencias en Textos en Español. *Primer Congreso Argentino de la Interacción-Persona Computador@, Telecomunicaciones, Informática e Información Científica* (2012)
20. Weiss, S. M., Indurkha, N., Zhang, T., Damerau, F. J.: *Text Mining Predictive Methods for Analyzing Unstructured Information*. Springer, USA (2005)
21. García Adeva, J. J., Calvo, R. A.: Mining Text with Pimiento. *IEEE Internet Computing*, 10 (4), 27–35 (2006)
22. Feldman, R., Sanger, J.: *The text mining handbook. Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press (2007)
23. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: *CRISP-DM 1.0 Step-by-step data mining guide*. SPSS Inc, USA (2001)
24. Magalhães, S. E.: *Descoberta de Conhecimento com o uso de Text Mining: Cruzando o Abismo de Moore*. No publicado, Tesis de Master, Universidad Católica de Brasil (2002)
25. Sebastiani, F.: Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1–47, (2002)
26. Li, Y., Zaragoza, H., Herbrich, R., Shawe-Taylor, J., Kandola, J.: The Perceptron Algorithm with Uneven Margins. En: *Nineteenth International Conference on Machine Learning*, pp. 379–386. Morgan Kaufmann Publishers Inc., San Francisco (2002)
27. Santana Mansilla P., Costaguta R., y Missio D.: Reconocimiento de Habilidades de Docentes en Aprendizaje Colaborativo Soportado por Computadora Mediante Minería de Textos. *42 Jornadas Argentinas de Informática e Investigación Operativa (JAIIO)* (2013)
28. Li, Y., Shawe-Taylor, J.: The SVM with uneven margins and Chinese document categorization. En: *17th Pacific Asia Conference on Language Information and Computation (PACLIC17)*, pp. 216–227. Colips Publications, Singapur (2003)