

# Estudo Exploratório sobre a Utilização de Algoritmos de Mineração de Dados na Predição e Caracterização da Qualidade de Objetos de Aprendizagem no Repositório MERLOT

Cristian Cechinel<sup>1</sup>, Salvador Sánchez-Alonso<sup>2</sup>, Miguel-Ángel Sicilia<sup>2</sup>, Merisandra Côrtes de Mattos<sup>3</sup>

<sup>1</sup>Curso de Engenharia de Computação– Universidade Federal do Pampa (UNIPAMPA)  
Caixa Postal 07 – 96400-970 – Bagé – RS – Brasil

<sup>2</sup>Information Engineering Research Group – Universidad de Alcalá  
Alcalá de Henares, Espanha

<sup>3</sup>Curso de Ciência da Computação  
Universidade do Extremo Sul Catarinense – Criciúma, SC – Brasil

contato@cristiancechinel.pro.br, {salvador.sanchez, msicilia}@uah.es,  
mem@unesc.net

**Abstract.** *In the last years, learning object repositories have become a reliable source of information about the current state of art regarding learning resource technologies. The large amount of learning objects stored in (or referenced by) the repositories allied with the availability of evaluative metadata provided by their communities has opened the possibility to search for patterns and rules that could be used in order to classify learning objects according to their quality. This paper presents an exploratory analysis about the use of data mining algorithms in the process of predicting and characterizing the quality of resources inside the MERLOT repository.*

**Resumo.** *Nos últimos anos, os repositórios de objetos de aprendizagem se tornaram uma confiável fonte de informações com relação ao estado da arte das tecnologias de ensino-aprendizagem. A grande quantidade de recursos armazenados nos (ou referenciados pelos) repositórios aliada à disponibilidade de metadados avaliativos fornecidos pela comunidade dos mesmos abriu a possibilidade de buscar por padrões e regras que possam ser utilizadas no processo de classificar os materiais com relação a sua qualidade. O presente trabalho apresenta um estudo exploratório sobre o uso de algoritmos de mineração de dados no processo de predição e caracterização da qualidade de recursos dentro do repositório MERLOT.*

## 1. Introdução

Objetos de Aprendizagem (OA) são normalmente definidos como qualquer recurso digital que pode ser reutilizado no apoio ao processo de ensino-aprendizagem e são considerados por muitos como a base para a larga disseminação de iniciativas de *e-learning* na internet. Diversas iniciativas e propostas para a avaliação da qualidade de um OA vêm sendo discutidas ao longo dos últimos anos (Diaz et al., 2002; Williams,

2000; Nesbit et al., 2003; por exemplo), entretanto, ainda não existe consenso sobre o que constitui um objeto de aprendizagem de qualidade, nem sobre qual a melhor maneira de conduzir um processo de avaliação. Em parte, isso pode ser atribuído à natureza heterogênea desses recursos, pois uma vez que eles podem diferir em diversos aspectos (tamanho, granularidade, tecnologia usada, padrão de metadados, duração, entre outros) (Churchill, 2007), é razoável assumir que os critérios de qualidade e as maneiras de medição dos mesmos também serão diferentes de acordo com esses muitos aspectos. Em todo caso, assim como o número de OAs cresceu consideravelmente nos últimos anos, também cresceu o número de repositórios de objetos de aprendizagem (ROAs) disponíveis para ajudar na organização e facilitar a busca dos recursos existentes. Esses ROAs, por sua vez, são potenciais agregadores de comunidades de prática (Han et al., 2008), ou seja, pessoas que compartilham interesses e preocupações sobre algo que fazem e que aprendem a partir de suas interações (Wenger, 2006). Subsequentemente, alguns desses repositórios aproveitam as características de tais ambientes sociais e adotam estratégias para o estabelecimento da qualidade que se baseiam nas impressões de uso e nas avaliações dadas por usuários e especialistas que são membros da comunidade do repositório. Essas são formas de metadados avaliativos (Vuorikari et al., 2008) que servem para o adequado ranqueamento e recomendação de recursos para os usuários. A atual abundância de recursos armazenados em repositórios (Ochoa e Duval, 2009) e a disponibilidade dessas avaliações contextuais em alguns deles abriu a possibilidade de buscar relações entre as características dos recursos existentes e a qualidade observada pelos usuários sobre esses recursos.

O presente trabalho apresenta um estudo inicial sobre a aplicação de algoritmos de mineração sobre um conjunto de dados coletados junto ao repositório MERLOT ([www.merlot.org](http://www.merlot.org)). Esse trabalho tem como objetivo explorar a possibilidade de utilização de algoritmos de mineração na predição da qualidade de objetos dentro do repositório, além de observar possíveis relações existentes entre as características dos OAs e sua qualidade. O restante desse artigo está estruturado da seguinte maneira: a seção 2 descreve o repositório MERLOT e algumas formas de medição da qualidade implementadas no mesmo. A seção 3 reporta a análise exploratória desenvolvida e discute alguns dos resultados observados, e a seção 4 apresenta uma breve caracterização da qualidade dos OAs no repositório. Finalmente, a seção 5 apresenta as conclusões iniciais e perspectivas de trabalhos futuros.

## **2. O Repositório MERLOT**

O MERLOT (*Multimedia Educational Resource for Learning and Online Teaching*) é uma iniciativa internacionalmente conhecida que permite que usuários cataloguem recursos educacionais com o objetivo de facilitar o uso e o compartilhamento de tecnologias de aprendizagem online (Cafolla, 2002). O repositório contém metadados de materiais de aprendizagem classificados em diferentes categorias de acordo com sua disciplina (área de estudo) e seu tipo (objetivos pedagógicos, granularidade e formato técnico), e possui um robusto sistema de avaliação da qualidade que é baseado em diferentes tipos de metadados avaliativos (comentários e notas de usuários, revisão de pares, prêmios, coleções pessoais).

Os recursos no MERLOT estão organizados em sete disciplinas diferentes: Artes, Negócios, Educação, Humanidades, Matemática e Estatística, Ciência e Tecnologia, e Ciências Sociais, que por sua vez também estão subdivididas em diversas

subcategorias. Quando um usuário cadastra um novo recurso no repositório, além da categoria da disciplina, ele precisa também informar qual o tipo daquele material, que é dividido nas seguintes classes: Animação, Ferramenta de Avaliação, Tarefa, Estudo de Caso, Coleção, Ferramenta de Desenvolvimento, Exercício e Prática, Repositório de Objetos de Aprendizagem, Curso Online, Jornal ou Artigo Aberto, Livro Texto Aberto, Aula/Apresentação, Material de Referência, Simulação, Ferramenta de Redes Sociais, Tutorial, Exame/Teste, Material para Treinamento e Workshop.

Com o objetivo de garantir qualidade dentro do repositório, o MERLOT adota um modelo de revisão por pares (Cafolla, 2002), onde os materiais já catalogados são revisados por especialistas que são membros da comissão editorial de alguma comunidade de disciplina (Artes, Matemática, Negócios, entre outras). As comissões editoriais do MERLOT decidem sobre o processo de selecionar os materiais que consideram dignos de serem avaliados, e esses materiais selecionados são então revisados e pontuados pelos pares. Após esse processo, o editor chefe da comissão compõe um único relatório (e uma média das notas) e publica o mesmo no repositório. Além da avaliação por pares, o MERLOT também permite que usuários registrados no repositório forneçam comentários e notas sobre os materiais, complementando a sua estratégia de avaliação com um mecanismo mais informal. As notas de ambos (usuários e revisores) variam entre 1 e 5 (com 5 como a melhor pontuação). Ainda, o MERLOT permite que seus usuários adicionem os recursos em coleções pessoais, fornecendo uma maneira dos mesmos organizarem seus materiais favoritos de acordo com seus interesses individuais. Todos esses metadados avaliativos juntos são então utilizados na ordenação dos materiais quando um usuário realiza uma busca dentro do repositório.

### **3. Metodologia**

A metodologia utilizada nesse estudo foi semelhante à utilizada por Romero et. al (2008) e está dividida em quatro etapas: 1) recolhimento dos dados, 2) pré-processamento, 3) aplicação dos algoritmos de mineração e 4) interpretação dos resultados. As etapas 2, 3 e 4 foram executadas em ciclos iterativos, uma vez que em alguns momentos a interpretação dos resultados trouxe a necessidade de uma nova organização dos dados para serem mais uma vez minerados.

#### **3.1. Recolhimento dos Dados**

Um total de 20267 de registros de OAs foram coletados (Dezembro de 2009) por meio de um *webcrawler* desenvolvido especificamente com esse propósito. As informações coletadas sobre cada um dos OAs foram: 1) categoria de disciplina (CD), 2) tipo de material (TM), 3) pontuação média obtida por meio da revisão dos especialistas (NRE), 4) pontuação média obtida por meio da revisão dos usuários (NRU), e 5) número de coleções pessoais em que o recurso está inserido (CP).

#### **3.2. Pré-Processamento**

Dentre os 20267 registros do conjunto original de dados, 33 não continham a informação relacionada ao tipo de material e foram desconsiderados. A maioria dos recursos não tinha nenhuma pontuação obtida por meio da revisão dos especialistas (NRE) ou por meio da revisão dos usuários (NRU), e do total de dados coletados, somente 3.42% apresentaram ao menos uma nota dada por especialistas e uma nota dada por usuários ao mesmo tempo. Cada uma dessas situações foi utilizada para gerar

um conjunto de dados diferente que foi posteriormente utilizado na etapa de aplicação dos algoritmos de mineração (Tabela 1).

**Tabela 1. Conjuntos de dados C1, C2 e C3 gerados na etapa de pré-processamento**

Total	C1 = NRE > 0		C2 = NRU > 0		C3 = NRE > 0 □ NRU > 0	
	Tamanho	%	Tamanho	%	Tamanho	%
20234	2586	12.78	2499	12.35	694	3.42

Com relação aos conjuntos, é importante ressaltar que C1 é utilizado para a extração de conhecimento sobre a qualidade dos objetos desde o ponto de vista dos especialistas, C2 para a extração do conhecimento sobre a qualidade desde o ponto de vista dos usuários, e C3 para a extração do conhecimento desde os dois pontos de vista. Isso se justifica no fato de estudos anteriores mostrarem que as duas comunidades de avaliadores no MERLOT (especialistas e usuários) possuem impressões diferentes com relação à qualidade dos objetos de aprendizagem (Cechinel et. al, 2010).

Em seguida, os valores das notas dadas por usuários e especialistas (variando de 1 a 5) foram transformados em classes nominais de qualidade (*ruim*, *médio* e *bom*) a partir da utilização dos tercís dessas variáveis como os limites dessas classes em cada um dos conjuntos de dados existentes. Assim, as pontuações abaixo do primeiro limite foram classificadas como *ruins*, as pontuações entre o primeiro limite e o segundo foram classificadas como *médios*, e as pontuações acima do segundo limite como *bons* (ver Tabela 2).

**Tabela 2. Conjuntos de dados gerados na etapa de pré-processamento**

Conjunto de Dados	Variável	Tercis para cada Classe		
		Ruim	Médio	Bom
C1	NRE	0 a 3.9	4. a 4.74	4.75 a 5
C2	NRU	0 a 3.9	4 a 4.4	4.5 a 5
C3	NRE	0 a 3.9	4 a 4.9	5
C3	NRU	0 a 3.9	4. a 4.74	4.75 a 5

A última parte da etapa de pré-processamento consistiu na transformação do formato dos dados para o formato requerido pela ferramenta de mineração de dados que seria utilizada. No nosso caso, os dados estavam armazenados em um banco de dados MySQL e foram exportados para um arquivo texto com formato ARRF (*Attribute-Relation File Format*).

### 3.3. Mineração de Dados

A etapa de mineração de dados foi realizada utilizando a ferramenta WEKA (Hall et. al, 2009). A tarefa de mineração inicialmente aplicada nos conjuntos foi a de classificação, que tem como objetivo a construção de modelos capazes de associar cada registro do conjunto de dados a um rótulo categórico, ou classe (Goldschmidt e Passos, 2005). Foram selecionados 10 algoritmos disponíveis na ferramenta WEKA, cinco deles cuja saída gera uma árvore de decisão (J48, SimpleCart, REPTree, RandomTree e Breath First Tree – BFTree ) e cinco cuja saída gera um conjunto de regras SE ENTÃO (Ripple-Down Rule Learner – Ridor, PART, Non-Nested generalized – Nnge, JRip, e Decision Table). Outros algoritmos de classificação que geram redes neurais, funções matemáticas e redes bayesianas não foram utilizados nesse primeiro estudo uma vez que também desejávamos estudar possíveis relações entre os valores das variáveis a partir do conteúdo das regras geradas.

Os algoritmos selecionados foram aplicados nos conjuntos C1, C2 e C3 utilizando como atributo de saída as variáveis NRE (para os conjuntos C1 e C3) e NRU (para os conjuntos C2 e C3). Como atributos de entrada foram informados: o tipo do material (TM), a categoria de disciplina a qual o material pertence (CD), e o número de coleções pessoais às quais o material está associado (CP). Esta última variável foi previamente identificada como um bom prognosticador de qualidade no repositório MERLOT (García-Barriocanal e Sicilia, 2009).

#### 3.3.1 Resultados para os conjuntos C1 e C2

A tabela 3 mostra os resultados obtidos na execução dos algoritmos de classificação para os conjuntos C1 e C2.

**Tabela 3. Resultados obtidos pelos algoritmos de mineração na classificação de OAs entre Bons, Médios e Ruins (Conjuntos C1 e C2)**

Algoritmo	Conjunto C1				Conjunto C2			
	Atributo de Saída = NRE				Atributo de Saída = NRU			
	CC	IC	K	E	CC	IC	K	E
J48	55.80	44.20	0.27	37.22	51.42	48.58	0.25	39.08
Simple Cart	53.91	46.09	0.24	38.26	48.90	51.10	0.22	40.51
REPTree	55.34	44.66	0.27	36.39	48.06	51.94	0.19	40.34
RandomTree	64.54	35.46	0.44	28.21	60.90	39.10	0.40	30.68
BFTree	50.54	49.46	0.16	39.54	55.14	44.86	0.32	35.73
Ridor	51.31	48.69	0.19	32.46	45.5	54.4	0.19	36.28

					8	2		
PART	58.93	41.07	0.33	34.38	54.18	45.82	0.30	37.17
NNge	58.20	41.80	0.35	27.87	54.14	45.85	0.30	30.57
JRip	52.59	47.41	0.20	39.37	44.82	55.18	0.11	41.93
Decision Table	50.62	49.38	0.16	39.46	46.78	53.22	0.17	41.10

Na tabela 3, CC representa o percentual de registros corretamente classificados pelo algoritmo, IC representa o percentual de registros incorretamente classificados e K representa o valor do coeficiente Kappa (uma medida que varia entre 0 e 1 e indica a concordância geral entre os dados observados e os esperados, sendo 0 para nenhuma concordância e 1 para total concordância), e E representa o erro médio geral. Como pode ser observado na tabela, os resultados obtidos pelos algoritmos testados não são muito animadores, apresentando um alto percentual de registros classificados de forma incorreta, coeficientes Kappa baixos, e um alto percentual de erros.

Considerando que os algoritmos não apresentaram bons resultados na classificação de OAs entre *bom*, *médio* e *ruim*, voltamos à etapa de pré-processamento e reclassificamos as variáveis NRE (conjunto C1) e NRU (conjunto C2) para assumir apenas os valores *bom* e *não bom* (sendo esse último uma união entre *médio* e *ruim*). Uma nova rodada de execuções dos algoritmos foi realizada alcançando melhores resultados (tabela 4).

**Tabela 4. Resultados obtidos pelos algoritmos de mineração na classificação de OAs entre Bons e Não Bons (Conjuntos C1 e C2)**

Algoritmo	Conjunto C1				Conjunto C2			
	Atributo de Saída = NRE				Atributo de Saída = NRU			
	CC	IC	K	E	CC	IC	K	E
J48	72.00	28.00	0.37	39.31	70.19	29.81	0.23	40.08
Simple Cart	67.75	32.25	0.28	43.48	67.15	32.85	0.16	42.30
REPTree	70.46	29.54	0.32	39.48	69.11	30.89	0.26	40.24
RandomTree	77.53	22.46	0.49	28.91	77.11	22.89	0.47	29.54
BFTree	67.75	32.25	0.28	42.78	67.47	32.53	0.13	42.01

Ridor	67.63	32.37	0.22	32.37	67.59	32.41	0.18	32.41
PART	72.20	27.80	0.39	37.64	70.91	29.09	0.28	38.59
NNge	73.16	26.84	0.42	26.84	68.91	31.09	0.35	31.09
JRip	68.91	31.09	0.31	42.13	67.47	32.53	0.20	43.37
Decision Table	67.75	32.25	0.28	42.71	67.15	32.85	0.16	42.3

Como é possível observar na tabela 4, para o conjunto C1 os algoritmos que apresentaram os melhores desempenhos foram: RandomTree (CC = 77.53% , K= 0.49 e E = 28.91), NNge (CC = 73.16%, K = 0.42 e E=26.84), PART (CC = 72.20, K = 0.39, e E = 37.64) e J48 (CC = 72.00, K = 0.37 e E = 39.31). Para o conjunto C2, os melhores resultados foram obtidos pelos algoritmos RandomTree (CC = 77.11, K = 0.47 e E = 29.54 ), PART (CC = 70.91, K = 0.28 e E = 38.59), NNge (CC=68.91, K =0.35 e E = 31.09) e J48 (CC = 70.19, K = 0.23 e E = 40.08). É importante ressaltar que todos os resultados obtidos para o conjunto C1 (variável de saída NRE) são melhores dos que o obtido para o conjunto C2 (variável de saída NRU), com apenas três exceções com relação aos valores de erro médio (E).

### 3.3.2 Resultados para o conjunto C3

Uma última rodada de execução dos algoritmos de classificação foi realizada para o conjunto de dados C3. Considerando que as execuções anteriores dos algoritmos de mineração nos conjuntos C1 e C2 obtiveram melhores resultados na classificação de OAs entre *bons* e *não bons*, as variáveis NRE e NRU desse conjunto também foram reclassificadas para trabalhar apenas com esses valores. A tabela 5 apresenta os resultados obtidos para o conjunto C3.

**Tabela 5. Resultados obtidos pelos algoritmos de mineração na classificação de OAs entre Bons e Não Bons (Conjunto C3)**

Algoritmo	Conjunto C3 Atributo de Saída = NRE				Conjunto C3 Atributo de Saída = NRU			
	CC	IC	K	E	CC	IC	K	E
J48	79.11	20.89	0.51	30.88	64.99	35.01	0.14	44.86
Simple Cart	76.66	23.34	0.44	33.55	63.83	36.17	0.05	46.03
REPTree	78.67	21.33	0.50	31.41	66.57	33.43	0.17	43.46
RandomTree	90.49	9.51	0.79	11.23	87.18	12.82	0.73	15.21
BFTree	77.65	22.34	0.45	31.65	63.83	36.17	0.05	46.03

Ridor	74.21	25.79	0.35	25.79	64.12	35.88	0.07	35.88
PART	80.26	19.74	0.55	26.99	70.75	29.25	0.30	39.37
NNge	89.34	10.66	0.76	10.66	85.01	14.99	0.68	14.99
JRip	72.91	27.09	0.42	37.52	65.13	34.87	0.13	45.35
Decision Table	72.91	27.09	0.42	35.86	63.69	36.31	0.08	45.82

Com relação ao atributo de saída NRE, os melhores resultados foram obtidos pelos algoritmos NNge (CC = 89.34, K = 0.76, E = 10.66), RandomTree (CC = 90.49, K = 0.79, E = 11.23), PART (CC = 80.26, K = 0.55, e E = 26.99) e J48 (CC = 79.11, K = 0.51, E = 0.51). Esses mesmos algoritmos, com exceção do J48, também obtiveram os melhores resultados para o atributo de saída NRU. Na tabela 5 também podemos observar que todos os algoritmos apresentaram melhores desempenhos (em todos os indicadores) para a variável de saída NRE.

Ao compararmos o desempenho dos algoritmos para a variável de saída NRE entre os conjuntos C1 e C3, avaliamos que os indicadores de qualidade são melhores para o conjunto C3 em absolutamente todas as situações observadas. Isso indica que, ao prever a qualidade de um OA desde o ponto de vista de um especialista, as relações entre as variáveis utilizadas para a mineração dos dados são mais fortes quando esse OA também já foi avaliado por algum usuário.

O mesmo não acontece quando comparamos os resultados para a variável de saída NRU entre os conjuntos C2 e C3. Aqui, apenas os algoritmos RandomTree e NNge (que já apresentavam o melhor desempenho para essa variável no conjunto C2) conseguiram melhorar seus índices de desempenho no conjunto C3.

#### 4. Considerações sobre a caracterização da qualidade dos OAs

De acordo com Goldschmidt e Passos (2005), o predicado ou ponto de separação em tarefas de classificação baseadas em árvores de decisão é escolhido como sendo a condição que melhor separa ou discrimina uma classe de saída. Com base nisso, os predicados dos resultados gerados pelos algoritmos de classificação por árvores de decisão para as variáveis NRE e NRU do conjunto C3 foram avaliados, tendo sido possível observar que:

- a) Para a variável NRE, todas as árvores de decisão foram geradas separando as classes primeiramente a partir do número de coleções pessoais (CP), seguido pela separação por categoria de disciplina (CD) e por último, pelo tipo de material (TM).
- b) Para a variável NRU, todas as árvores de decisão foram geradas separando as classes primeiramente pelo tipo de material (TM), sendo que as próximas separações se alternaram entre o uso do número de coleções pessoais (CP) e da categoria de disciplina (CD).

Com o objetivo de observar um pouco mais as relações entre as variáveis estudadas e a qualidade dos OAs, aplicamos também a tarefa de associação (com o algoritmo Predictive a Priori) no conjunto C3. O algoritmo foi executado duas vezes,



uma para o conjunto de variáveis NRE, CP, MT e CP, e outra para o conjunto de variáveis NRU, CP, MT e CP. Na mineração de dados, a tarefa de associação é utilizada para buscar relacionamentos significativos entre os itens que ocorrem simultaneamente em um conjunto de dados permitindo a observação de tendências e padrões.

Como o algoritmo Predictive a Priori permite apenas a utilização de dados categóricos, a variável CP (número de coleções pessoais) foi transformada utilizando metodologia semelhante a que foi utilizada para a categorização de NRE e NRU (ver seção 3.2), porém dividindo a mesma em cinco categorias, sendo elas:  *muito pouco*,  *pouco*,  *médio*,  *alto*,  *muito alto*.

Diferentemente da tarefa de classificação, na associação não se define qual a variável de saída desejada para as regras, ou seja, o algoritmo busca livremente no conjunto fornecido as relações existentes. Ao executar o algoritmo para o conjunto C3 e fornecendo como entrada as variáveis NRE, CD, TM, e CP foram geradas 29 regras com um grau de confiança superior a 90% (o grau de confiança indica a probabilidade de ocorrência da consequência considerando o antecedente da regra). Dessas 29 regras, 23 traziam como consequência a qualidade do OA (NRE). Essa informação demonstra a força da relação existente entre as variáveis CD, TM e CP e a qualidade do OA quando consideramos o ponto de vista do especialista. Ainda, as regras geradas apontam que OAs que possuem um número de coleções pessoais  *muito pouco*,  *pouco* ou  *médio* sempre estão associados a uma qualidade  *não bom*, confirmando a influência anteriormente observada dessa variável na qualidade do OA (García-Barriocanal e Sicilia, 2009)

Ao aplicar o algoritmo de associação para as variáveis NRU, CD, TM e CP foram geradas apenas 11 regras com grau de confiança superior a 90%, sendo que dentre estas, somente 7 traziam como consequência a qualidade do OA (NRU). Aqui, o número de coleções  *muito pouco*,  *pouco* e  *médio* normalmente também aponta para um OA de qualidade  *não bom*, entretanto, aparecem regras com exceções, em que OAs com um número de coleções pessoais  *muito alto* é classificado como  *não bom*.

## 5. Conclusões

O presente estudo apresenta algumas contribuições importantes sobre a predição de qualidade de OAs por meio da mineração de dados para o contexto específico do repositório estudado (MERLOT). Alguns dos algoritmos de classificação aplicados demonstraram um bom desempenho na predição da qualidade dos OAs, sendo eles: NNge, RandomTree e PART e indicam uma possibilidade real de utilização da mineração de dados para a predição da qualidade dentro do repositório. Outro aspecto relevante observado foi que os algoritmos de mineração alcançam resultados significativamente melhores quando classificam OAs entre  *bons* e  *não bons*, do que quando classificam OAs entre  *ruins*,  *médios* e  *bons*. Ainda, foi possível verificar que as relações entre as variáveis utilizadas (categoria de disciplina, tipo de material e número de coleções pessoais) e a qualidade dos OAs são mais fortes desde o ponto de vista da comunidade de especialistas, do que do ponto de vista da comunidade de usuários, reforçando a conclusão de trabalhos anteriores de que essas duas comunidades comunicam impressões diferentes com relação à qualidade dos OAs (Cechinel et. al, 2010). Sobre esse tópico, o estudo das árvores de decisão geradas permitiu demonstrar que para a predição de qualidade no contexto da percepção dos especialistas, a variável

mais importante é o número de coleções pessoais (CP), seguida pela categoria de disciplina (CD) e o tipo de material (TM), e que para o contexto da percepção dos usuários, a variável mais importante é o tipo de material (TM). Ainda sobre esse assunto, foi possível observar que, ao predizer a qualidade de um OA desde o ponto de vista de um especialista, as relações entre as variáveis utilizadas para a mineração dos dados são mais fortes quando esse OA também já foi avaliado por algum usuário.

Nos próximos trabalhos serão explorados outros algoritmos de mineração (com métodos que utilizam redes bayesianas, redes neurais, lógica nebulosa, entre outros) para avaliar a possibilidade de melhorar o desempenho e diminuir o erro médio nesse processo de predição, assim como a inserção de outras variáveis relacionadas aos OAs e possíveis de serem extraídas do repositório, como, por exemplo, o formato do mesmo e algumas medidas intrínsecas dos recursos (números de links, número de imagens, entre outros).

## Referências

- Cafolla, R. (2002). Project Merlot: Bringing Peer Review to Web-based Educational Resources. In: Proceedings of the USA Society for Information Technology and Teacher Education International Conference, pp. 614– 618.
- Cechinel, C., Sanchez-Alonso, S., Sicilia, M-A (2010). Exploratory analysis of the correlations between peer-reviewers and users ratings on MERLOT repository. In: Proceedings of the II Conferencia conjunta Iberoamericano sobre Tecnologías del Aprendizaje. Cadiz. Spain.
- Churchill, D. (2007). Towards a useful classification of learning objects. *Educational Technology Research and Development* 55 (5), 479-497.
- Diaz, P., Sicilia, M.-A. (2002). Evaluation of hypermedia educational systems: criteria and imperfect measures. In: Proceedings of the International Conference on Computers in Education (ICCE'02). Auckland, New Zealand, pp. 621-626 vol.1.
- García-Barriocanal, E., Sicilia, M-A. (2009). Preliminary explorations on the statistical profiles of highly-rated learning objects. In: *Metadata and Semantic Research*. Vol. 46 of Communications in Computer and Information Science. Springer Berlin Heidelberg, pp. 108-117.
- Goldschmidt, R., Passos, E. (2005). *Data Mining: um guia prático*. Rio de Janeiro: Elsevier, 2005.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., (2009); *The WEKA Data Mining Software: An Update*; *SIGKDD Explorations*, 11(1).
- Han, P., Kortemeyer, G., Kramer, B.J., and Prummer, C.: Exposure and Support of Latent Social Networks Among Learning Object Repository Users. In: *Journal of Universal Computer Science (J.UCS)*, Vol. 14, Issue 10, 1717-1738 (2008)
- Nesbit, J. C., Belfer, K. & Leacock, T. (2003). *Learning Object Review Instrument (LORI)*. E-Learning Research and Assessment Network.
- Ochoa, X., Duval, E. (2009). Quantitative analysis of learning object repositories. In: *IEEE Transactions on Learning Technologies*, 2 (3), 226-238.

- Romero, C., Ventura, S., García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. In: *Computers & Education*, 51 (1), pp. 368-384.
- Vuorikari, R., Manouselis, N., Duval, E. (2008). Using Metadata for Storing, Sharing and Reusing Evaluations for Social Recommendations: the Case of Learning Resources. In: *Social Information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively*. Idea Group Inc., New York, NY, pp. 87–107.
- Wenger, E.: *Communities of practice: a brief introduction*. (2006). Available at [http://www.ewenger.com/theory/communities\\_of\\_practice\\_intro.htm](http://www.ewenger.com/theory/communities_of_practice_intro.htm). [Last access 25th February 2010]
- Williams, D. D. (2000). Evaluation of learning objects and instruction using learning objects. D. A. Wiley (Ed.), *The Instructional Use of Learning Objects: Online Version*. URL <http://reusability.org/read/chapters/williams.doc>