

7. Conclusão

A motivação principal deste trabalho é a busca de um processo que permita que um usuário pesquise em documentos, LOs que ele possa reutilizar na composição de seu material didático. Partindo do tipo de LO desejado, que especificaria a sua proposta instrucional, e de palavras-chave indicando o tema de interesse, o sistema proposto encontrará partes elementares de documentos que contenham o assunto procurado no formato especificado (por exemplo, uma definição). Deste modo facilitamos a criação de ALOs que poderão ser recuperados em consultas integradas durante a composição de LOs complexos (aulas, cursos etc...).

Propusemos a aplicação de um modelo onde o objeto LO é segmentado em pedaços menores, permitindo facilmente a reutilização desta pequena porção de conhecimento, agora apreendida em um ALO, em diversas situações. Podemos, por exemplo, utilizar a definição de um termo ou palavra, em qualquer conteúdo instrucional que esteja tratando de algum conceito relacionado a este termo.

Como apoio na definição, com maior nível semântico, do que pode ser considerado um ALO propusemos o uso de uma Ontologia de Objetos de Aprendizagem. Utilizamos técnicas de aprendizado de máquina para apreender dos exemplos, informações sobre os textos que queremos encontrar, gerando um Corpus e um Modelo Bayesiano de Misturas Multinomiais.

Utilizando a solução proposta fizemos alguns experimentos e alcançamos um nível de precisão nesta classificação dentro do esperado, com valores de *recall* altos para as classes positivas indicando que o algoritmo encontrou a grande maioria dos textos que deveriam ser encontrados. Por outro lado, diversos resultados dos experimentos principalmente relativos a métrica *precision* nos mostraram as dificuldades que podem ser encontradas na aplicação do aprendizado de máquina a este problema.

7.1. Contribuições

A principal contribuição deste trabalho foi o desenvolvimento de um processo para obtenção de LOs que foi formalizado com a criação do sistema SQLLOMining. Este sistema propiciou um ambiente para a execução de experimentos que poderá ser utilizado como um laboratório, dando prosseguimento a execução de novos testes que ficam aqui listados como trabalhos futuros.

No sistema SQLLOMining foi implementado o algoritmo Expectation-Maximization (EM) associado ao modelo Bayesiano de misturas multinomiais, proposto por (Nigam, Maccallum, Thrun e Mitchell, 2000) disponibilizando um classificador de textos que executa o aprendizado semi-supervisionado, recurso bastante poderoso no processo de mineração de texto. A aplicação desta técnica ao problema de obtenção de LOs trouxe contribuições tanto para a área de aprendizado de máquina quanto para a área de e-learning.

Adicionalmente, com a realização de um experimento bastante simples utilizando o sistema SQLLOMining, pudemos observar que o uso da ontologia apoiando a criação das classes favoreceu o processo de aprendizado. A ontologia aumentou o nível semântico do modelo de LOs e proporcionou um maior grau de comprometimento em manter coerentes os tipos de ALOs utilizados na classificação. Desta forma procuramos satisfazer minimamente as premissas utilizadas na especificação do modelo de misturas permitindo a utilidade do aprendizado semi-supervisionado. Este experimento mostrou que o processo de separação dos textos existentes entre as classes possíveis é uma parte do problema de extrema importância, que nos remete ao já bastante estudado “Problema de Clusterização” (Ochi, 2004).

7.2. Trabalhos Futuros

Como trabalhos futuros existem diversos experimentos que podem trazer resultados muito interessantes sobre como a ontologia pode auxiliar no processo de obtenção de LOs. Um exemplo é analisar o que ocorreria com as métricas de desempenho se passássemos a utilizar os níveis mais altos da ontologia, como classes para o algoritmo de classificação. Intuitivamente podemos imaginar que quando trabalhamos com uma classe de um nível mais elevado perdemos

características específicas que facilitariam a sua identificação. É razoável supor que quanto mais elevado este nível for mais difícil fica a classificação.

A justificativa teórica que podemos encontrar para esta intuição está nas duas premissas relativas à geração de documentos: (1) existe um modelo de misturas, (2) existe uma correspondência de um para um entre as componentes da mistura e as classes. Para nos mantermos coerentes com estas premissas, é necessário subdividir as classes, permitindo que a relação de um para um seja verdadeira. Se utilizarmos, por exemplo, apenas uma superclasse chamada conceito englobando as suas subclasses: Definição, Lei, Fato, Procedimento etc..., e criarmos uma segunda classe chamada de Outros que englobaria todo o resto, claramente teremos uma relação de um para n entre as componentes e as classes. Neste caso esperamos obter um desempenho pior.

Outros trabalhos futuros que podem se feitos utilizando o sistema SQLLOMining:

- Aumento do corpus:
 - Permitir a execução de um n-fold cross validation aumentando a qualidade dos experimentos
 - Aumentar a quantidade de classes
- Aprimoramentos no algoritmo EM para melhorar o desempenho do aprendizado semi-supervisionado:
 - Fator de peso para as amostras
 - Mistura para cada componente - similar a utilização de mais classes indicadas pela ontologia
- Aprimoramento no pré-processamento do texto:
 - Separação dos termos mais importantes
 - Separação de n-gramas considerando todos os grupos possíveis

Finalmente podemos concluir que este trabalho também contribui para a seqüência de pesquisas voltadas para e-Learning do Laboratório TecBD do Departamento de Informática visando o descobrimento, obtenção, integração, armazenamento e compartilhamento de Learning Objects.