

5 Conclusão

A Análise Probabilística de Semântica Latente é uma modelagem que permite o aprendizado de grupo de usuários através de sua verossimilhança com os dados de treinamento. Esta técnica é adaptável ao contexto de sistemas de recomendação, onde é necessário recomendar itens a usuário conforme suas preferências. Neste trabalho avaliamos o desempenho do PLSA em dois problemas de sistemas de recomendação: a recomendação de anúncios na web e a recomendação de filmes. Na recomendação de anúncios na web o PLSA apresentou resultados excelentes na Precisão do Posicionamento (PP), atingindo 18,1% de melhoria sobre a média das médias, e 10,7% sobre o SVD para as 10 primeiras posições de anúncios. Além desta melhoria foi atingido um nível de precisão de 73,2% na primeira posição dos anúncios. Já na recomendação de filmes foi atingida uma melhoria de 3,78% no RMSE sobre a Média das Médias, mostrando a eficiência do algoritmo mesmo com pouca informação sobre os dados. Juntamente com os resultados analíticos o algoritmo implementado foi desenvolvido de forma a ser flexível e reutilizável, sendo possível com isto integrá-lo ao framework de recomendação de anúncios na web LearnAds.

Apesar dos resultados obtidos o algoritmo apresenta algumas limitações. O PLSA tem complexidade computacional de $O(I \times Z)$, onde I é o número de exemplos e Z a quantidade de grupos latentes, tornando o treinamento de modelos grandes uma tarefa bastante demorada. Além da complexidade computacional elevada o PLSA apresenta uma dependência forte da inicialização das probabilidades do modelo, tornando o conhecimento a priori um fator importante de convergência do algoritmo.

Devido a estas limitações algumas oportunidades de melhorias são sugeridas como trabalhos futuros: a primeira delas é a distribuição do algoritmo, onde apesar de termos mostrado que a separação total dos dados dificilmente é obtida, o desenvolvimento de uma estratégia inteligente de separação parcial dos mesmos

tornaria viável a distribuição do algoritmo. Também é possível a implementação da estratégia apresentada no *sPLSA* (Lin et al., 2007) que elimina a computação de probabilidades muito baixas diminuindo assim o tempo de treinamento. Um trabalho futuro sugerido para a questão da inicialização do PLSA é a utilização de usuários típicos do conjunto de treino como âncoras dos grupos, dando a eles probabilidade 100% de estar em um grupo desejado. Esta estratégia pode diminuir o número de mínimos locais do algoritmo, aumentando assim as chances de convergência. Finalmente o módulo desenvolvido apresenta a possibilidade de integração do PLSA com algoritmos de Boosting, porém não foram realizados ainda experimentos utilizando esta técnica juntamente com o PLSA (Freund & Schapire, 1996).

Com isto mostramos que o PLSA é um modelo que se adapta bem ao contexto de sistemas de recomendação, atingindo bons níveis de precisão em sua recomendação. Além disto, o este permite uma grande flexibilidade para comportar diferentes semânticas do modelo, bastando adaptar as probabilidades em sua inicialização. Finalmente ele oferece campo de pesquisa para sistemas distribuídos e otimização, necessitando de um empenho para melhorar seu tempo de treinamento e conseqüentemente a sua usabilidade em sistemas comerciais.