

5. Conclusões e trabalhos futuros

5.1. Resumo do trabalho

Sistemas de recomendação já estão consolidados como ferramentas importantes que auxiliam os usuários da Web em diversos contextos. Seja em lojas virtuais ou em indicações de filmes em cartaz, os sistemas de recomendação tem facilitado enormemente a tomada de decisão dos usuários na internet.

Neste trabalho, abordamos o problema da recomendação de segundo nível, que é similar à recomendação baseada em filtragem de conteúdo, porém, com um fator maior de subjetividade, pois trata da recomendação sem a utilização do perfil de consumo do usuário. Propusemos um sistema de recomendação, o **GRNews**, que auxilia os jornalistas na produção de conteúdo relacionado às matérias publicadas para os usuários do site na Web. Como esta recomendação é feita para o jornalista, não utilizamos os dados de perfil de usuário, que são típicos em sistemas de recomendação baseados em filtragem de conteúdo. Por este motivo, denominamos de *segundo nível* nosso modelo para sistema de recomendação.

O **GRNews** foi construído com módulos acopláveis como um *framework*, em que podemos substituir partes do sistema por componentes que forneçam os melhores resultados. Como exemplo, as *features* apresentadas no Capítulo 3 podem ser adicionadas ou removidas do sistema sem afetar o funcionamento do sistema como um todo.

A implementação do **GRNews** é independente da plataforma de publicação de conteúdo, podendo ser acoplada a qualquer base de dados de matérias. Nosso primeiro piloto foi com a base de dados de matérias do portal **G1**. Porém, temos a previsão de aplicar também o modelo sobre a base de dados de matérias do portal **Globoesporte.com**.

5.2. Principais contribuições

A principal contribuição desta dissertação refere-se ao modelo de recomendação de segundo nível, pois se trata da utilização de técnicas de recomendação para suporte à tomada de decisão da equipe de jornalistas.

Como não encontramos na literatura um trabalho similar, entendemos que esta abordagem pode ser utilizada para auxiliar outras equipes de jornalistas na publicação de conteúdos em outros portais na Web.

Outra contribuição consiste no extrator de entidades nomeadas. Embora não tenha sido o foco principal do nosso trabalho, o nosso extrator de entidade apresentou bons resultados e entendemos que ele pode ser melhorado e utilizado em outros trabalhos de processamento de texto.

5.3. Limitações do sistema

Algumas limitações puderam ser observadas durante o desenvolvimento da dissertação. Poderíamos aumentar o número de *features* utilizadas levando em consideração, por exemplo, os *links* existentes nas matérias. De fato, observamos alguns trabalhos na literatura que utilizaram os documentos referenciados por *links* no corpo do texto como possíveis fontes de informações.

Também temos limitações na informação extraída. Alguns trabalhos utilizam uma abordagem de expansão de conceitos, ou seja, dado um conceito extraído, o sistema utilizava um serviço de suporte que permite identificar informações relacionadas a este conceito. De fato, poderíamos utilizar a Wikipédia para extrair mais informações relacionadas às *features* extraídas.

Percebemos como outra limitação do sistema, o uso de ngrams como base para a identificação de entidades nomeadas. já é conhecido na literatura [30,31] que o uso de chunkers apresenta melhores resultados no processo de indentificação de entidades. Uma abordagem futura para este trabalho poderia ser a substituição do módulo de identificação de entidades nomeadas por outro que fizesse uso de chunkers.

5.4. Trabalhos futuros

Como uma primeira iniciativa de trabalho futuro, pretendemos criar subeditorias das editorias para as quais o sistema apresentou uma baixa acurácia. O objetivo da criação destas subeditorias seria identificar se, com grupos mais homogêneos, o sistema **GRNews** conseguiria elevar o percentual de acurácia. Esta hipótese foi levantada já que as editorias *Rio de Janeiro* e *São Paulo* obtiveram os melhores índices de assertividade.

Outra atividade futura é um estudo para aperfeiçoamento do componente de extração de entidades nomeadas, visando melhorar o percentual da sua acurácia.

Pode ser observado também a possibilidade de adicionar novos critérios de relacionamento de notícias como é o caso do relacionamento por analogia semântica. Em nosso trabalho apenas utilizamos as sugestões por similaridade textual e por popularidade.