

Inferência de Geolocalização de Usuários de Redes Sociais

Silas Fernandes Moreira, Luciano Antonio Digiampietri

Escola de Artes Ciências e Humanidades - Universidade de São Paulo

silas.moreira@usp.br, digiampietri@usp.br

Objetivos

Conhecer a geolocalização de usuários é uma informação importante para diversas tarefas, como a recomendação de locais, rotas e produtos. Essa informação pode ser fundamental para o correto funcionamento de determinadas funções, porém, pode não ser estritamente necessária para a interação do usuário com a aplicação – como acontece em diversas redes sociais online populares.

Quando a informação de geolocalização não é requerida pela aplicação, observa-se que a frequência do uso de marcadores precisos de localização é bastante baixa. Assim, é interessante que sejam desenvolvidos sistemas computacionais que possam inferir a geolocalização desses usuários com base em suas postagens nas redes sociais.

O objetivo deste trabalho é desenvolver um sistema de inferência de geolocalização de usuários em redes sociais online.

Métodos e Procedimentos

Para desenvolvimento do projeto, foi escolhida a rede social *Facebook* para aplicação do sistema de inferência. Mais especificamente, foi utilizada uma série de postagens públicas de uma página brasileira sobre viagens.

O código fonte do sistema de inferência é construído na linguagem *JavaScript* utilizando as APIs do *Facebook* e do *Google Places*.

A API do *Facebook* permite acesso direto às postagens da página selecionada e a API do *Google Places* facilita a busca por *matches* de lugares.

O sistema percorre as postagens da página – da mais nova à mais antiga e, para cada palavra ou grupo de palavras, procura alguma

correspondência no banco de dados de lugares da API do *Google Places*.

Resultados

Todas as entradas exatas foram bem sucedidas na busca pelo respectivo local de referência e até entradas com algumas letras erradas (erros de digitação), a mais ou faltantes obtiveram sucesso de forma bastante rápida.

As principais dificuldades encontradas foram: a) ambiguidade de lugares (como o estado e a cidade de São Paulo); b) palavras homógrafas, considerando ou não acentos gráficos (preposição “para” e estado Pará); c) palavras que referenciam lugares mas não condizem com o nome real do lugar (palavra “Sampa” referenciando São Paulo).

Conclusões

O desenvolvimento de um sistema de inferência de geolocalização não é uma tarefa fácil, principalmente porque é necessário o tratamento de casos não convencionais e imprevisíveis. Os próximos passos do projeto incluem o tratamento mais adequado a esses casos. Entretanto, em boas condições de entrada, o sistema já desenvolvido se mostra bastante eficaz e eficiente na busca pelos lugares referentes às entradas.

Referências Bibliográficas

RIBEIRO, Silvio; PAPPA, Gisele L. Strategies for combining Twitter users geo-location methods. *Geoinformatica*, p. 1-25.

HAN, B.; COOK, P.; BALDWIN, T. Geolocation prediction in social media data by finding location indicative words. In *Proceedings of COLING*, 2012. p. 1045–1062.