

MC886 - Proposta de projeto

Guilherme P. Gonçalves (RA 091429)

Agosto de 2013

O problema

Este trabalho explora a aplicação de conceitos de Machine Learning para modelar os padrões de acesso a rede de um navegador web, de forma a tomar ações que antecipem as necessidades do usuário e tornem a navegação mais rápida. Especificamente, pretendo trabalhar na camada de rede do navegador Firefox para implementar ações como iniciar resoluções DNS e conexões TCP antes que o usuário as requisite explicitamente, baseado nos padrões de uso anterior do usuário.

Dados disponíveis

O algoritmo de predição em que pretendo trabalhar deve, dado o atual estado do navegador e as informações já conhecidas sobre os padrões de navegação do usuário, determinar quais URLs serão acessadas no futuro próximo. Essas URLs podem apontar tanto para domínios que o usuário visita normalmente quanto para recursos (arquivos de CSS e Javascript) carregados por páginas desse domínio.

Dentre os dados disponíveis para a predição, o algoritmo terá acesso a informações referentes ao estado do navegador – página atual e seu conteúdo, conteúdo do cache de navegação, requisições de rede feitas anteriormente – e informações referentes ao usuário – seu histórico, tempo de navegação e horário de acesso a determinadas páginas.

Proposta de trabalho

Já existe uma [implementação preliminar](#) de um algoritmo desse tipo esperando revisão para ser incorporado ao Firefox. Esse algoritmo é bastante simples e não existem métricas para validar sua eficácia. Pretendo começar meu trabalho definindo uma métrica para que se possa avaliar a qualidade das predições de forma simples – a princípio, pretendo usar a taxa de uso das conexões preditas e taxa de acertos ao cache DNS do navegador. Depois disso, pretendo estender (ou reescrever) esse algoritmo existente de forma a tomar proveito de ideias relacionadas a padrões comuns de navegação.

A primeira ideia seria explorar a relação entre as páginas acessadas pelo usuário e o horário em que esse acesso acontece. Por exemplo, um padrão de navegação comum consiste em abrir o navegador e, nos primeiros minutos, acessar um conjunto de páginas mais ou menos estável dado um mesmo horário. Essa informação pode ser usada para que o navegador tome ações de predição ao ser aberto em determinados horários, carregando preventivamente páginas que o usuário costuma acessar.

Também pretendo usar o fato de que o conteúdo acessado por um usuário em uma determinada página pode apresentar localidade espacial – por exemplo, um usuário que gosta de notícias sobre esportes costuma clicar em links referentes a esse assunto ao abrir a página de seu portal de notícias preferido, e o navegador poderia abrir conexões a esses links antes do clique do usuário. Essa localidade pode ser explorada tanto usando a informação visual (identificar regiões “quentes” da página, com as quais o usuário costuma interagir, e agir em seus links) quanto estrutural da página (agir sobre elementos “quentes” da árvore DOM em vez de regiões).

Finalmente, pretendo avaliar o impacto de usar parâmetros adicionais para o algoritmo de predição, como presença ou não da página em cache ou no histórico e “profundidade” das páginas acessadas dentro de um domínio na hierarquia de páginas, de forma a refinar as predições com amparo dos dados experimentais. Outras aplicações de predições também podem ser implementadas ao longo do desenvolvimento do projeto conforme novos dados e ideias surgirem, em caráter opcional.