

6 Considerações Finais

Os resultados experimentais apresentados em nossa dissertação mostram uma grande variação de desempenho entre os diferentes algoritmos em um determinado cenário e de um mesmo algoritmo em cenários variados. Por outro lado, o pequeno desvio padrão na maior parte das simulações, mostra que não há muita variação nos resultados de vários dias de processamento mantendo-se o mesmo cenário, quando temos uma distribuição de consultas razoavelmente bem comportada.

Estes resultados nos levam a concluir que, uma análise de vários dias com dados reais, seria uma boa estratégia para apontar a melhor solução para um caso específico, mas isto nem sempre é possível, por falta de tempo, recursos ou outros motivos.

Neste capítulo, consolidamos as análises experimentais e apontamos direções para a escolha de um algoritmo a ser utilizado na solução de casos reais, discutimos o uso de predição nos algoritmos e sugerimos algumas possibilidades de pesquisa visando ampliar os resultados aqui obtidos.

6.1 Sobre o Desempenho dos Algoritmos

Considerando os cenários descritos no capítulo anterior, apresentamos na Tabela 6.1 o percentual de simulações em que cada algoritmo obteve o melhor desempenho. Como ocorreram casos de empate entre os algoritmos, o total de cada conjunto de cenários não é de 100%. Consideramos empatados os resultados com uma diferença de até 0,1%. Não apresentamos o desempenho do algoritmo Prima-Dual, por ter resultados muito próximos aos do algoritmo AdWords. Mantivemos apenas este último como representante deste tipo de estratégia.

O desempenho dos algoritmos que utilizam predição sobre a distribuição diária das consultas, como podemos ver pelos resultados, mostrou-se bem superior aos outros. O algoritmo Aleatório foi o melhor quando $m \leq Q_{max}$ e o algoritmo Preditivo nos casos em que $m > Q_{max}$.

Logo, se considerarmos a implementação de um algoritmo híbrido que,

quando a previsão de consultas for menor ou mesmo pouco superior ao limite total de exibições, funcione como o algoritmo Aleatório e, caso contrário, como o algoritmo Preditivo, este novo algoritmo será o melhor em 100% dos casos.

Algoritmo	Anunciantes Similares	Anunciantes Diferenciados
Aleatório	65%	65%
Preditivo	35%	35%
Guloso	26%	14%
AdWords	1%	2%

Tabela 6.1: Estatística sobre desempenho dos algoritmos

Considerando a relação entre o número de consultas diárias e o limite total de exibições (m/Q_{max}) que utilizamos para dividir os cenários avaliados, na Tabela 6.2, apresentamos o algoritmo de melhor desempenho em cada situação.

Consultas / Limites	Anunciantes Similares	Anunciantes Diferenciados
1/4	Aleatório e Guloso	Aleatório e Guloso
1/2	Aleatório	Aleatório e Guloso
1	Aleatório	Aleatório
2	Preditivo	Preditivo
4	Preditivo	Preditivo

Tabela 6.2: Melhores algoritmos

Apesar do excelente desempenho dos algoritmos que utilizam predição, é importante destacarmos dois fatores que podem impedir o seu uso: o primeiro é não existir um padrão de distribuição que possa ser capturado e o segundo é não termos recursos suficientes para o cálculo diário da solução ótima sobre uma distribuição média.

Em vista disso, na Tabela 6.3, apresentamos os resultados considerando apenas os algoritmos que não utilizam predição e, na Tabela 6.4, qual seria a melhor escolha em cada cenário.

Algoritmo	Anunciantes Similares	Anunciantes Diferenciados
Guloso	66%	40%
AdWords	35%	60%

Tabela 6.3: Estatística sobre desempenho dos algoritmos sem predição

Consultas / Limites	Pouca Competição	Muita Competição
1/4	Guloso	Guloso
1/2	Guloso	Guloso
1	Guloso	AdWords
2	Guloso	AdWords
4	Guloso	AdWords

Tabela 6.4: Melhores algoritmos sem predição

Apresentamos os resultados de forma distinta nas Tabelas 6.2 e 6.4 porque, ao compararmos apenas os algoritmos Guloso e AdWords, algo que realmente impacta no melhor comportamento de um ou de outro, além da relação entre a quantidade de consultas e o limite de exibições (m/Q_{max}), é o nível de competição entre os anunciantes.

6.2 Sobre o Uso de Predição nos Algoritmos

Durante nossa pesquisa, fizemos uma tentativa de mesclar o uso da predição com o algoritmo AdWords. Dentro do algoritmo Preditivo que desenvolvemos, quando este não selecionava um anunciante para uma determinada consulta, utilizávamos a função do algoritmo AdWords para selecionar o anunciante.

Para os cenários onde o número de consultas é maior que o limite de exibições ($m > Q_{max}$), o fato deste algoritmo sempre atribuir uma consulta a um anunciante, fazia o algoritmo esgotar os limites dos anunciantes mais cedo e não permitia ao algoritmo ampliar a receita com as consultas que chegavam depois.

Para os cenários com quantidade de consultas menor ou igual ao limite total de exibições ($m \leq Q_{max}$), este mecanismo até melhorava um pouco o desempenho do Preditivo, mas não ficava tão bom quanto o algoritmo Guloso, por exemplo. Neste caso, o fato de sempre selecionarmos um anunciante para cada consulta, gerava um aumento da receita em relação ao Preditivo sem Adwords, mas também causava uma perda de oportunidade de uma receita melhor mais a frente.

No fim, o desempenho deste algoritmo ficou sempre próximo do resultado obtido pelo AdWords e resolvemos deixá-lo de lado.

Como o algoritmo Preditivo sem o AdWords apresentou ótimos desempenho para os casos em que $m > Q_{max}$, a alternativa encontrada foi desenvolver o algoritmo Aleatório, que também utiliza predição. O resultado obtido

por este algoritmo foi um ótimo desempenho exatamente quando o algoritmo Preditivo não tinha um desempenho tão bom, ou seja, quando $m \leq Q_{max}$.

6.3

Trabalhos Futuros

Acreditamos que esta dissertação contribuiu para esclarecer sobre o desempenho prático dos algoritmos de seleção de links e que os algoritmos propostos podem ser uma boa alternativa para problemas reais. Sabemos também que ficaram muitas questões em aberto e que merecem uma maior investigação.

Um trabalho importante seria avaliar o desempenho dos algoritmos propostos e adaptados com dados reais. Não fizemos este trabalho, pois não conseguimos estes dados.

Após o trabalho de pesquisa e avaliação dos algoritmos, surgiram algumas idéias sobre variações nos conjuntos de cenários apresentados que poderiam enriquecer o entedimento sobre o comportamento destes algoritmos. Descrevemos algumas a seguir.

Sobre o conjunto de cenários com anunciantes similares, a variação que nos parece mais interessante é utilizar uma distribuição normal para os valores das ofertas para cada tipo de consulta. Após definida a faixa de ofertas $[b_{min}, b_{max}]$, para cada tipo de consulta sorteamos um valor médio inicial e as ofertas dos anunciantes numa distribuição normal em torno deste valor. O objetivo é simular uma situação onde algumas consultas tenham um valor médio de leilão mais alto e outras mais baixo, mas com a maioria das ofertas próximas para um mesmo tipo de consulta. Para estes tipo de cenário utilizaríamos valores de b_{min} e b_{max} que não fossem muito próximos.

Uma outra variação parecida com esta, mas para o conjunto de cenários com anunciantes diferentes, é utilizar faixas de oferta diferenciadas por tipo de consulta, mas dentro de cada faixa continuaríamos a utilizar a distribuição de Zipf, como nos cenários que avaliamos.

Outra possibilidade é variar os tipos de distribuição sobre as consultas, já que em nossos cenários utilizamos sempre distribuições de Zipf. O objetivo seria verificar se a relação de desempenho entre os algoritmos se mantem.

Um outro caminho a ser explorado seria utilizar outras ferramentas e equipamentos para o cálculo da solução ótima e assim trabalhar com valores de entrada maiores.

No aspecto teórico, seria interessante uma discussão mais aprofundada sobre a competitividade dos algoritmos considerando um intervalo entre as ofertas mínima e máxima $[b_{min}, b_{max}]$.