

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
CURSO DE CIÊNCIA DA COMPUTAÇÃO

PEDRO EDUARDO BAHÍ DE SOUZA

**Um estudo empírico sobre técnicas de melhoria do desempenho de  
Comparadores**

Monografia apresentada como requisito parcial para a  
obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Leandro Krug Wives

Porto Alegre  
2015

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Graduação: Prof. Sérgio Roberto Kieling Franco

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do Curso de Ciência da Computação: Prof. Raul Fernando Weber

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

## RESUMO

Comparadores são serviços encontrados na Web que auxiliam consumidores no processo de decisão de compra, expondo produtos semelhantes lado a lado, mostrando os preços de diferentes lojas ao mesmo tempo. Essa facilidade de uso também é percebida pelo varejista, que obtém bons resultados de receita com investimentos abaixo da média em comparação com outras mídias. Os varejistas pagam por cada clique em seus produtos expostos na vitrine do comparador, por isso é preciso continuamente revisar os produtos, removendo produtos que estão gerando custo sem retorno de receita e recolocando produtos que apresentem melhor probabilidade de venda. Este controle exige análise contínua e trabalho repetitivo realizado pelos responsáveis pela exposição dos produtos. Neste trabalho são estudados, descritos e desenvolvidos processos para auxiliar varejistas a anunciar em comparadores com investimento otimizado, reduzindo os custos com anúncios sem afetar as vendas, através da seleção de produtos com melhor probabilidade de venda. Como resultado, foi demonstrado que os produtos escolhidos com maior probabilidade apresentaram efetivamente melhores resultados.

**Palavras-chave:** comparadores, aprendizado de máquina, adaptativo, reação, otimização, cpa, leilão, publicidade, comércio eletrônico.

## **An empirical study of machine learning techniques for improving the performance of Comparison Shopping Sites**

### **ABSTRACT**

Comparison Shopping Sites help consumers in the purchase decision process, exposing similar products side by side, showing the prices of different stores simultaneously. This ease of use is also perceived by the retailer, achieving good revenue at low cost when compared to other medias. Retailers pay for each click on their products exposed in the comparison shopping, so products must be continuously review, by removing products that are generating cost-counter return and by placing products with better probability of sale. This control requires analysis continuous and repetitive work done by those responsible for product display. This work describe the process model designed to help retailers advertise in comparators with optimized investment by reducing costs without affecting the sales, by selecting products with better probability of sale. As a result, it was demonstrated that the products chosen exhibited better results.

**Keywords:** comparison shopping, machine learning, adaptative, reaction, optimization, CPA, bidding, marketing, e-commerce.

## LISTA DE FIGURAS

- Figura 2.1 - Processo de comparação e escolha em um comparador
- Figura 3.1 – Comparativo do KUBE com variações de *Epson-first* com recompensas homogêneas
- Figura 3.2 – Comparativo do KUBE com variações de *Epson-first* com variação nas recompensas
- Figura 3.3 – Comparativo do KUBE com variações de *Epson-first* com recompensas muito diversas
- Figura 4.1 - Arquitetura do projeto
- Figura 4.2 - Comparação entre CPA e Cliques entre Aquisições
- Figura 4.3 - Relação entre cliques e conversão
- Figura 4.4 - Processo de clique em um produto
- Figura 4.5 - Processo de conversão
- Figura 4.6 - Processo de redução de preço de produtos
- Figura 4.7 - Processo de conversão em origens diferentes
- Figura 4.8 - Processo de decaimento de produtos pausados
- Figura 4.9 - Processo de criação de novos produtos
- Figura 5.1 – Eficiência de produtos em função da posição no rank

**LISTA DE TABELAS**

Tabela 5.1 – Eficiência de produtos por grupos de ranking

## LISTA DE ABREVIATURAS E SIGLAS

UFRGS - Universidade Federal do Rio Grande do Sul

CPA - Custo por Aquisição

CPC - Custo por Clique

ROI - Retorno sobre Investimento

KUBE - *knapsack-based upper confidence bound exploration and exploitation*

MAB - *Multi-armed Bandit*

API - *Application Programming Interface*

EPGP - *Effort Points / Gear Points*

CRM - *Customer Relationship Management*

XML - *Extensible Markup Language*

## SUMÁRIO

1. INTRODUÇÃO.....	10
1. FUNDAMENTAÇÃO TEÓRICA .....	11
1.1. Sites Comparadores .....	11
1.2. <i>Multi-armed Bandit</i> (MAB).....	12
1.3. Leilão baseado no problema da mochila .....	12
1.4. <i>Starvation</i> .....	13
1.5. Indicadores de <i>E-Commerce</i> .....	13
1.5.1. Custo por Aquisição (CPA).....	13
1.5.2. Custo por Clique (CPC).....	15
1.5.3. Retorno sobre Investimento (ROI) .....	15
1.6. Série Temporal - ARIMA.....	16
1.7. Resumo do capítulo .....	17
2. TRABALHOS RELACIONADOS .....	18
2.1. Ranking para recompensas (EPGP).....	18
2.2. <i>Knapsack Based Optimal Policies for Budget-Limited Multi-Armed Bandits</i> .....	18
2.3. Zoom.....	21
2.4. Resumo do capítulo .....	22
3. MÉTODO PROPOSTO.....	23
3.1. Captura de dados .....	24
3.2. Cliques entre Aquisições .....	24
3.3. Ranking de produtos.....	25
3.4. Adaptação do <i>Multi-armed Bandit</i> .....	30
3.5. <i>Starvation-freedom</i> baseado em decaimento.....	31
3.6. <i>Keywords</i> em comparadores.....	31
3.7. Resumo do capítulo .....	31
4. VALIAÇÃO .....	32
4.1. Séries temporais.....	32
4.2. Reação por correlação .....	32
4.3. Elasticidade de preços .....	32
4.4. Ranking.....	33
4.5. Resumo do capítulo .....	34
5. CONCLUSÕES .....	35
REFERÊNCIAS .....	36



APÊNDICE A - COLETA DE DADOS ..... 38

## 1. INTRODUÇÃO

O *E-Commerce* trouxe benefícios tanto para o consumidor quanto para o varejo, consumidores ampliaram as possibilidades de compra com um mercado sem limitações geográficas. Para lidar com uma variedade maior de opções foram criadas ferramentas de busca e comparação de produtos. Nesse contexto, sites de comparação, aqui chamados de comparadores, são relevantes. Os sites de comparadores, de maneira geral, possuem um bom retorno sobre o investimento (ROI), eles facilitam o trabalho do consumidor, agregando muitas lojas e as ajudando a escolher entre preços e características de produtos que melhor atendem suas necessidades.

Os varejistas pagam por cada clique em seus produtos expostos na vitrine do comparador, por isso é preciso continuamente revisar os produtos, removendo produtos que estão gerando custo sem retorno de receita e recolocando produtos que apresentem melhor probabilidade de venda. Este controle exige análise contínua e trabalho repetitivo realizado pelos responsáveis pela exposição dos produtos. Este trabalho tem como objetivo melhorar a utilização de comparadores do ponto de vista do varejista, propondo políticas e algoritmos, que reduzem o custo com mídia, mantendo a receita e dispensando a necessidade de revisar todo inventário de produtos várias vezes por dia e tomar decisões de investimento sobre cada um.

O texto está organizado da seguinte forma. O capítulo seguinte descreve os conceitos e fundamentos teóricos relacionados com a área de comparadores. Em seguida, o capítulo três apresenta trabalhos relacionados. No capítulo quatro é descrito o método proposto e seus processos. No capítulo cinco são apresentados os resultados dos algoritmos avaliados. O capítulo seis contém as conclusões deste trabalho.

## 1. FUNDAMENTAÇÃO TEÓRICA

Este capítulo descreve os conceitos relacionados com o trabalho desenvolvido. Inicia-se com o conceito de Site Comparador. Alguns indicadores de marketing são necessários nesta seção, pois estão diretamente relacionados com as otimizações propostas e na elaboração de algoritmos para a área.

### 1.1. Sites Comparadores

Os *sites* comparadores, de maneira geral, têm uma API muito limitada se comparada a outros veículos, que contam com sofisticados sistemas de leilão e análises detalhadas de transações. Os comparadores cobram os varejistas por clique e o varejista decide o quanto vai pagar pelos cliques, propostas com maior valor pelo clique recebem mais exposição, o ideal seria pagar mais por produtos mais lucrativos, porém a venda de cliques não é feita por produto, mas por categoria fazendo que produtos com baixo rendimento fiquem em igualdade de exposição com produtos mais lucrativos, prejudicando a estratégia de *bidding*. A atualização de produtos é realizada por meio de *Feed XML* com apenas uma atualização diária, limitando a flutuação de preços. Caso um anunciante apresente em seu *site* preços diferentes dos anunciados no *Feed*, pode ser banido, limitando a liberdade de flutuação de preços.

Figura 2.1 - Processo de comparação e escolha em um comparador



Fonte: [www.zoom.com.br](http://www.zoom.com.br)

## 1.2. *Multi-armed Bandit* (MAB)

O *Multi-armed Bandit* é um modelo de problema muito utilizado, atualmente, em algoritmos de inteligência artificial e é aplicado para fundamentar o algoritmo de ranking proposto neste trabalho. O dilema de descobrir (*exploration*) versus utilizar (*exploitation*) pode ser descrito como a busca da melhor relação entre gastar recursos para descobrir um ambiente ou para tirar proveito dele por meio de ações possivelmente mais lucrativas (AUER; CESA-BIANCHI; FISCHER, 2002).

A instância mais simples desse dilema, talvez, seja o *Multi-armed Bandit*, um problema estudado em estatística (BERRY; FRISTEDT, 1985 apud AUER; CESA-BIANCHI; FISCHER, 2002) que acabou se tornando fundamental em diferentes áreas de inteligência artificial como o aprendizado por reforço (SUTTON; BARTO, 1998 apud AUER; CESA-BIANCHI; FISCHER, 2002) e a programação evolutiva (HOLLAND, 1992 apud AUER; CESA-BIANCHI; FISCHER, 2002).

O problema descreve uma experiência hipotética, na qual uma pessoa enfrenta diversas máquinas caça-níqueis (chamadas de *one-armed bandits*, em inglês) com resultados esperados potencialmente diferentes. Deve-se encontrar a máquina com a melhor taxa de resultado, mas também se quer maximizar os ganhos. A tensão fundamental é entre "explorar" os cenários com bom desempenho no passado e "estudar" cenários novos ou aparentemente inferiores, caso eles possam ter um desempenho ainda melhor (SCOTT, 2010).

Tomando como exemplo o contexto de comparadores e produtos expostos, cada produto é considerado uma máquina caça níquel e os cliques são as moedas. Produtos com venda frequente são considerados com maior probabilidade de ganho, e os que precisam de muitos cliques para ceder uma recompensa são considerados pouco lucrativos. Os pouco lucrativos são deixados de fora e as apostas se concentram naqueles que são vistos com maior probabilidade. As máquinas ativas, então, estarão recebendo apostas continuamente e, em determinado momento, uma delas receberá muitas apostas sem ceder recompensa, até que ela será considerada pior que alguma desativada, então será retirada e a que estava em espera tomará seu lugar.

## 1.3. Leilão baseado no problema da mochila

Encontrar as melhores campanhas a serem exibidas é o principal fator de performance, que pode ser formulada como um problema de otimização, que maximiza receita e é limitado pela

receita e disponibilidade de estoque (CHEN; ANDERSON; BERKHIN; DEVANUR, 2011). A estratégia utilizada por Zhou, Lukose e Chakrabarty, em 2007, é tratar o leilão pelas melhores *keywords* como o problema da mochila e desenvolver algoritmos que, provavelmente, obtenham as melhores taxas, permitindo a automação do processo de *bidding*, enquanto otimizam os lances para atingir os diversos objetivos da campanha (ZHOU; LUKOSE; CHAKRABARTY, 2007).

#### **1.4. Starvation**

No caso de produtos se deseja que produtos com baixo desempenho, eventualmente, tenham espaço na vitrine, comparando produtos com processos, a vitrine pode ser considerada um recurso que o produto precisa para funcionar, em computação esse problema é denominado *Starvation* e é o problema, em que um processo fica perpetuamente esperando por recursos que precisa para funcionar. Em um sistema dinâmico, requisição simultânea de recursos ocorre constantemente, sendo necessária uma política, que decida qual ficará com o recurso, contudo essa política pode fazer com que um processo jamais obtenha o recurso necessário (Tanenbaum, 2009) devido a sua baixa prioridade em recursos muito concorridos. O conceito de Starvation-freedom consiste em utilizar uma política com garantia, que nenhum processo fique em espera por tempo indeterminado, de forma que em algum momento ele tenha acesso ao recurso, mesmo que competindo com processos de maior prioridade.

#### **1.5. Indicadores de E-Commerce**

A seguir são descritos uma série de indicadores e métricas comumente usados em e-Commerce.

##### **1.5.1. Custo por Aquisição (CPA)**

O denominado custo por aquisição (CPA) é uma das métricas mais importantes utilizadas para medir o custo envolvido para “Adquirir” uma venda por meio de veículos de mídia, por isso custos de produção e logística, por exemplo, não são incluídos nessa métrica (STOKES, 2008). Ela pode ser utilizada para medir veículos de mídia individualmente, como para medir globalmente, envolvendo custos com agências, equipes de criação, análise de dados e serviços de algoritmos de otimização.

Este trabalho não envolve encontrar valores ideais de CPA, apenas o utiliza como *input* para otimizar outros indicadores. Existem diversas métricas de *e-commerce*, que parecem não fazer sentido do ponto de vista econômico, mas fatores como sigilo de informações, comparações entre plataformas de mídia e simplificações para facilitar entendimento e tomada de decisões faz com que sejam adotadas como padrões de mercado métricas simples demais, que ocultam detalhes importantes. Os algoritmos utilizados para otimizar o CPA são alvo de muita discussão, de maneira geral, os veículos não conseguem encontrar um bom CPA, pois para essa decisão é necessário entender os custos envolvidos do lado do varejista que, por sua vez, se recusa a revelar tais informações, passando para o varejista a responsabilidade de como seu CPA deve variar, normalmente, essa decisão é estática, sendo definida baseada em cálculos estimados e teorias econômicas, não são utilizados algoritmos capazes de reagir dinamicamente a situações reais, o custo para o desenvolvimento ou contratação deste algoritmo e os riscos envolvidos não são proporcionais aos ganhos, fazendo da decisão empírica a mais comum, mesmo entre grandes grupos de varejo.

Existem diversas anomalias relacionadas com o CPA, este possui um ponto ideal, que não é conhecido e varia constantemente, valores muito altos, assim como muito baixos são igualmente prejudiciais, assim como encontrar quais são considerados como valores muito altos e muito baixos que se apresentam como outro problema.

Um CPA muito baixo traz questionamento sobre o volume de vendas, para atingir baixos níveis de CPA é necessário excesso de cautela, focando apenas em indivíduos e situações com grande probabilidade de compra, limitando o público-alvo e reduzindo as vendas (PERLICH; DALESSANDRO; HOOK; STITELMAN; RAEDER, 2013). Um CPA muito alto pode inviabilizar uma campanha, fazendo com que orçamentos disponíveis sejam consumidos antes de atingir metas de vendas.

Esses comportamentos levam a um comportamento dual, em que é necessário definir dois inputs, o CPA e o orçamento diário ou mensal. Enquanto o CPA estiver abaixo do CPA ideal, o investimento pode continuar mesmo após ter ultrapassado o orçamento, pois mesmo que o custo em mídia seja exorbitante, o faturamento será proporcional e irá cobrir todos os custos. Se ao final do período, o CPA estiver abaixo da meta será um resultado ruim, mesmo que com vendas muito acima do esperado, pois indica que poderia investir mais e obter resultados melhores ainda. Quando o CPA estiver operando acima do ideal, então, ele será limitado pelo orçamento, fazendo que o investimento seja paralisado até o término do período.

Por fim, é necessário observar a saturação do CPA, aumentando os lances no leilão, pela mídia é possível aumentar a exposição por um custo maior e causar uma elevação no CPA médio, mas existe o risco de já ter atingido a exposição máxima e aumentar os lances não irá melhorar a exposição ou as vendas (HU; SHI; TANG, 2013). Nesta situação, é necessário manter o investimento estável e redirecionar os recursos excedentes, investindo em outras frentes, para isso seria necessário uma atuação multi veículos, que não será abordada neste trabalho.

### **1.5.2. Custo por Clique (CPC)**

O custo por clique (CPC) é a forma mais antiga de aquisição de mídia *online*, a simplicidade faz dela a métrica mais popular e disponível em todos os veículos, sua apuração é direta (STOKES, 2008). Ao contrário do CPA, que é decidido de maneira empírica, o CPC é extremamente flutuante e algoritmos para otimizar lances são usados em larga escala, composta por muitas camadas, passando por diversas redes de anunciantes, que tentando comprar espaço, da mesma forma que produtores de conteúdo, tentando vender espaço com leilões multilaterais em tempo real. O valor do CPC não oferece um bom suporte para tomar decisões, mesmo que ele seja elevado poderá ser vantajoso se levar a uma alta taxa de conversão, pelo mesmo motivo que um CPC muito baixo pode ser desfavorável, se não resultar em conversões. Não podendo ser utilizado como base para comparações entre veículos, mas é possível utilizar o CPC combinado com as conversões para compor o CPA, ao permitir comparações e tomada de decisões (HU; SHI; TANG, 2013).

Esta conversão não é precisa por diversas razões, um consumidor pode ser impactado por diferentes veículos, que influenciaram na sua decisão de compra, mas a conversão será atribuída ao último veículo que recebeu o clique, então, mais de um veículo foi responsável pela venda, mas somente o CPA do último teve melhora, aos demais é atribuída apenas a assistência à conversão. Como este trabalho leva em consideração um único veículo, as assistências serão desconsideradas e somente o CPA absoluto será utilizado.

### **1.5.3. Retorno sobre Investimento (ROI)**

O retorno sobre investimento (ROI) é uma estimativa da relação entre recompensa e custo em um veículo de mídia, toma como base o total do preço de venda de todas as transações dividindo-se pelo total de investimento em mídia, essa relação de somas ao longo de um período é necessária, pois o preço de venda varia constantemente, assim como o custo do clique, oferecendo

uma perspectiva mais ampla de uma campanha (STOKES, 2008). O ROI junto com o CPA são as métricas favoritas dos executivos de marketing, mas assim como outras métricas, a análise isolada pode levar a diversas inconsistências, uma vez que é possível, por exemplo, obter excelentes métricas de ROI e CPA com resultados globais péssimos. O ROI não observa a lucratividade de um produto. Um produto com uma boa margem de lucro é comparado, da mesma maneira que outro com uma margem melhor, além deste problema é uma prática comum utilizar campanhas por categorias, em que muitos produtos diferentes compõem a mesma métrica, agravando o problema de mensurar a lucratividade (JEFFERY, 2013). Outro problema relacionado com ROI e CPA é que não consideram o volume de vendas, tecnicamente é possível obter maiores lucros sacrificando o desempenho destes dois indicadores.

Mesmo não oferecendo as medidas mais apuradas de lucratividade, e com o risco de interpretações incorretas estas são simples de apurar e oferecem uma boa comparação entre veículos ou mídias semelhantes. Fazendo deles indicadores amplamente utilizados, auxiliando a unificar a linguagem nas diversas organizações envolvidas no processo de publicidade.

## 1.6. Série Temporal - ARIMA

Durante a fase inicial deste trabalho, o algoritmo ARIMA foi avaliado com a intenção de prever os produtos ou *biddings*, que deveriam ser utilizados, durante o desenvolvimento outras abordagens foram adotadas, mas a descrição desta série é importante para apresentar as dificuldades relacionadas em aplicar o algoritmo ao caso estudado.

ARIMA é um algoritmo que se aplica a séries temporais. Uma série temporal é um conjunto de valores, que ocorrem em um período de tempo com um determinado padrão. Os padrões mais comuns são tendência de crescimento ou queda, ciclo, sazonalidade e flutuações irregulares (BOWERMAN; O'CONNELL; KOEHLER 2005 apud NGO, 2013).

Para modelar séries temporais, como uma função de valores anteriores, os analistas precisam assumir que o padrão ocorrerá no futuro (NGO, 2013).

O modelo ARIMA (*Autoregressive integrated moving average*) utiliza todos estes padrões para realizar previsões, este é amplamente utilizado na estatística e em modelos econômicos, possui diversas variações entre as mais importantes estão as que permitem fazer previsões com e sem sazonalidade (MAKRIDAKIS; WHEELWRIGHT; HYNDMAN, 1998).



## 1.7. **Resumo do capítulo**

Para formulação deste capítulo foi necessário entender o funcionamento de diversas mídias, a forma como o mercado as trata, entendendo suas vantagens e inconsistências, auxiliando a escolher os métodos mais apropriados para compor a solução. O entendimento constituído sobre as métricas de *E-commerce* interfere negativamente sobre a pesquisa, pois estas métricas procuram maximizar indicadores, que devem ser desconsiderados na construção de um modelo de decisões.

## 2. TRABALHOS RELACIONADOS

Este capítulo descreve os trabalhos relacionados com o trabalho desenvolvido. Inicia-se com o *addon* EPGP para o jogo *World of Warcraft*, o algoritmo KUBE para estimar orçamentos e, por fim, apresenta uma descrição sobre o comparador Zoom.

### 2.1. Ranking para recompensas (EPGP)

A ideia inicial de utilizar sistema de ranking neste trabalho veio de um *addon* para o jogo *World of Warcraft*, jogando cooperativamente ao derrotar inimigos, o grupo ganha recompensas (*loot*) que os jogadores deverão entrar em consenso de qual deles merece ou precisa mais acerca de determinados itens. Essa distribuição nem sempre é justa, sendo o motivo de muitos conflitos dentro do grupo. Cada grupo é livre para decidir quais regras irá utilizar na distribuição, uma opção é o EPGP - *Effort Points / Gear Points* um *addon*, que contabiliza quantas vezes um jogador já ajudou o grupo e quantas recompensas ele já recebeu, a relação dessas duas métricas gera um ranking, em que jogadores com melhor escore têm prioridade na disputa por recompensas.

Mesmo parecendo justa, tal condição gera outro problema, constantemente grupos organizados precisam procurar novos jogadores para substituir os que param de jogar, sem um grupo completo é praticamente impossível alcançar as recompensas. Quando novos jogadores descobrem, que o grupo utiliza EPGP eles desistem, imediatamente, de participar no grupo, pois sabem que estarão concorrendo com jogadores antigos com muitos pontos acumulados e que jamais terão oportunidade de participar, igualmente, na disputa pelas recompensas. Para amenizar esse problema, o *addon* implementa decaimento, fazendo que jogadores antigos e com muitos pontos acumulados não se tornem donos absolutos de todas futuras recompensas.

### 2.2. *Knapsack Based Optimal Policies for Budget-Limited Multi-Armed Bandits*

O problema do *multi-armed bandit* (MAB) foi inicialmente proposto por Robbins (1952), em que se apresenta um dos melhores exemplos de troca entre *exploration* e *exploitation* em aprendizado por reforço.

Entretanto, o modelo MAB fornece uma descrição incompleta para o problema de tomada de decisão quando confrontado a cenários do mundo real (TRAN-THANH; CHAPMAN; ROGERS; JENNINGS, 2012).

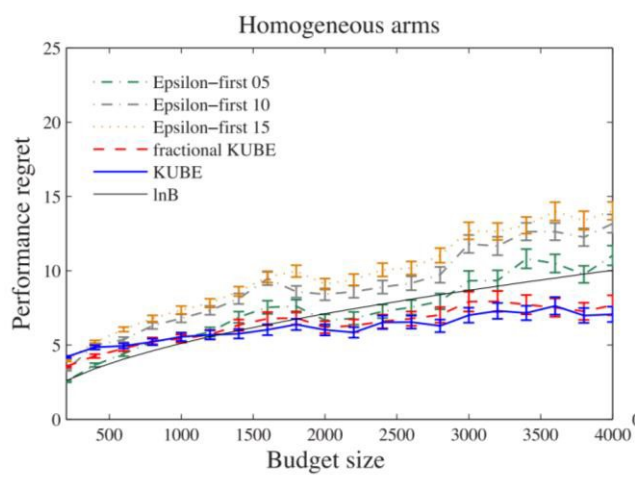
Uma variação é o modelo de MAB com restrição de orçamento, em que a aposta possui um custo e é limitada por um orçamento prefixado, ou seja, existe um limite para quantidade de apostas que um agente pode realizar para tentar descobrir a recompensa de uma máquina na fase inicial de exploração. Na fase seguinte de *exploit*, o agente deve seguir a política de apostar na opção com maior expectativa de recompensa.

O problema deste método é semelhante ao problema do CPA ideal, eles dependem de um orçamento, em que um agente externo precisa tomar a decisão de quanto será o orçamento, com efeitos diretos sobre a performance. Neste método, o orçamento é dividido em duas partes, uma que será usada na primeira fase e o restante na segunda, investir muito na primeira fase proporciona uma boa exploração, mas prejudica a segunda fase de *exploitation* e vice-versa. Dado o problema este trabalho relacionado propõe algoritmos para solucioná-lo, chamados KUBE (*knapsack-based upper confidence bound exploration and exploitation*) e KUBE fracional, o segundo é uma versão com menor custo computacional, baseado no algoritmo *fractional relaxation* de (KELLERER; PFERSCHY; E PISINGER 2004).

A abordagem adotada pelo KUBE não é determinar como deve ser a divisão entre as duas fases, mas trabalhar com as duas fases ao mesmo tempo, em que após cada tentativa, todas as opções são recalculadas para encontrar a melhor próxima tentativa.

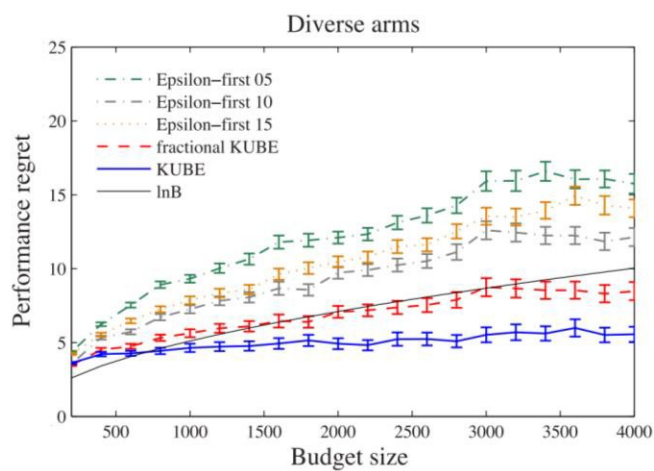
No comparativo apresentado na Figura 3.1 são demonstrados experimentos com resultados de performance, comparando resultados entre MAB com orçamento fixo e as duas variações de KUBE variando entre apostas com custo homogêneo e extremamente diverso. As divisões nas séries *Epsilon-first* representam divisões entre as duas fases de um orçamento máximo de 20, dos possíveis valores de 0 a 20 se escolheu [5,10,15] para demonstração.

Figura 3.1 – Comparativo do KUBE com variações de *Epsilon-first* com recompensas homogêneas



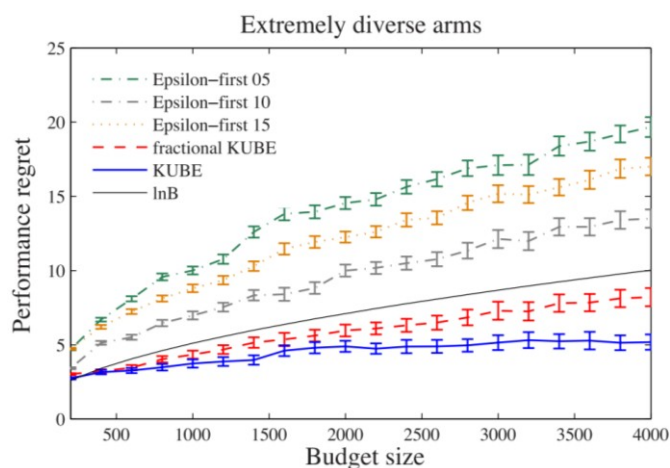
Fonte: Tran-thanh, Chapman, Rogers e Jennings (2012)

Figura 3.2 – Comparativo do KUBE com variações de *Epsilon-first* com variação nas recompensas



Fonte: Tran-thanh, Chapman, Rogers e Jennings (2012)

Figura 3.3 – Comparativo do KUBE com variações de *Epsilon-first* com recompensas muito diversas.



Fonte: Tran-thanh, Chapman, Rogers e Jennings (2012)

Nos casos com custo de aposta homogêneo as duas variações do KUBE obtiveram resultado similar, mas de maneira geral o KUBE superou a variação fracional, o que era esperado, pois o método guloso ordenado por densidade utilizado no KUBE oferece uma aproximação para o problema da mochila ilimitada melhor que o método *fractional relaxation*. No entanto, em todos os casos KUBE e sua variação obtiveram resultados de performance superiores ao método de orçamento prefixado, apresentando redução de perdas entre 70% e 50%. Além de estender o problema de MAB com orçamento fixo, também se consegue demonstrar que a performance de seus algoritmos está limitada, superiormente, por  $O(\ln B)$ , em que  $B$  é o orçamento disponível.

### 2.3. Zoom

O site Zoom<sup>1</sup> não é líder de mercado de comparadores, mas vem apresentando excelentes resultados e crescimento, seus principais diferenciais são a compra garantida, em que o Zoom atua como intermediário em caso de discordância entre o consumidor vendedor e garante tempo máximo para disputas ou o dinheiro de volta e gráficos com histórico de preços, que ajudam os consumidores a reconhecer descontos reais com a opção de avisar, quando o produto estiver pelo menor preço histórico. Mesmo oferecendo boas facilidades ao consumidor, o mesmo não ocorre

<sup>1</sup> Comparador Zoom - <http://www.zoom.com.br>

com a interface de varejista, os relatórios são limitados, não oferecem a visibilidade necessária para tomada de decisões e o sistema de leilão não é disponível via API.

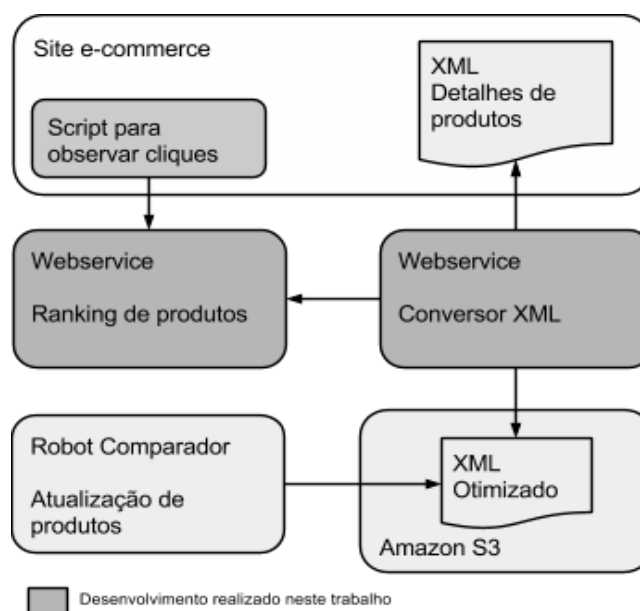
#### **2.4. Resumo do capítulo**

Neste capítulo foi apresentado um trabalho realizado por uma equipe de física, aparentemente sem relação com publicidade digital, apresentando alta similaridade com trabalhos utilizados pelo Google e Yahoo na otimização de seus resultados, semelhanças tanto na abordagem quanto nos obstáculos e objetivos. Durante o estudo foram analisados algoritmos utilizados na área de Marketing, que oferecem abordagens avançadas para relacionamento com o cliente, mas fogem completamente da natureza do problema abordado neste trabalho, mesmo pertencendo ao mesmo domínio.

### 3. MÉTODO PROPOSTO

O método proposto visa obter respostas imediatas às variações percebidas no comportamento de compra, tendo em vista que utilizando séries temporais a previsão de eventos com natureza caótica fica comprometida. Espera-se que a reação rápida baseada em eventos recentes, sem o uso de séries temporais, seja capaz de reduzir o investimento sem prejudicar o volume de vendas. A escolha do método proposto ocorreu após realização de testes com diferentes técnicas, entre elas séries temporais, elasticidade de preços e otimização de ROI e CPA, algumas foram descartadas e outras mantidas parcialmente, entre os motivos podem ser citados dificuldade de obtenção de dados ou por não se aplicar corretamente a solução proposta, os motivos de cada escolha são descritos no capítulo cinco com avaliação dos resultados de cada uma. Influenciou na decisão da abordagem do trabalho, que poderia tratar questões como segmentação de público baseado em preferências do consumidor, localização geográfica ou variação de lances por posição na vitrine a exemplo do que ocorre no Google Adwords, sendo constatado que em comparadores essas abordagens não são relevantes ou não são possíveis, e com isso se optou pela abordagem de remover ou colocar produtos do *Feed* baseada na probabilidade de venda. Além de testar técnicas existentes também foram criadas variações como o Clique entre Aquisições e ordenação por Ranking serão detalhadas neste capítulo.

Figura 4.1 - Arquitetura do projeto



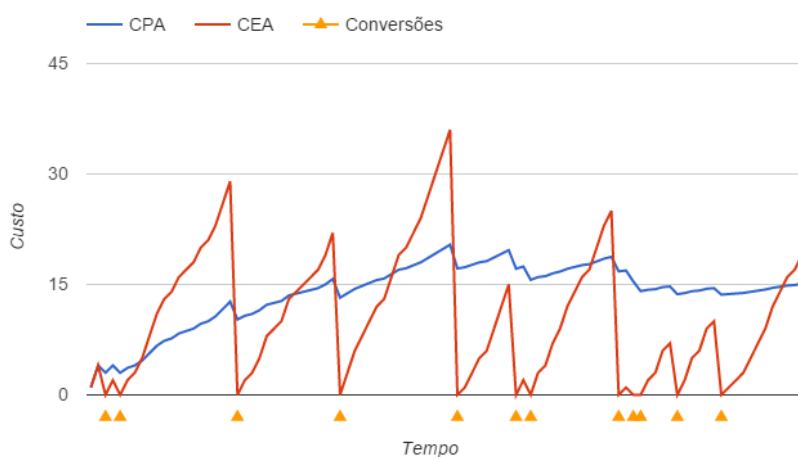
### 3.1. Captura de dados

Os dados são capturados com a inserção de código no *site* do varejista, que informa a origem do clique e eventos de conversão, esses eventos são enviados ao *webservice* implementado neste trabalho, que armazena os eventos, organiza estatísticas e compõe *Feed* com *ranking* de produtos.

### 3.2. Cliques entre Aquisições

Este indicador foi desenvolvido neste trabalho e não é utilizado pelo mercado. O principal objetivo é oferecer uma alternativa ao CPA, que depende de muitas estimativas e precisa de períodos fechados para ser construído, por se tratar de uma média não reage imediatamente aos eventos recentes, que podem ocorrer, de maneira pontual, por um curto intervalo e levar a perda de oportunidades prejudicando o desempenho. Pode ser usado para medir quantos cliques ocorreram desde a última conversão, sendo este indicador quanto menor melhor, logo após uma conversão passa para zero e é incrementado a cada novo clique, oferecendo resposta instantânea a eventos como mudanças de preços ou problemas de estoque de concorrentes.

Figura 4.2 - Comparação entre CPA e Cliques entre Aquisições



O gráfico desta simulação mostra a latência do CPA para convergir quando o comportamento muda repentinamente. O indicador de Cliques entre Aquisições possui uma convergência mais rápida, oferecendo melhor suporte para construção de um sistema autônomo, que precisa reagir a eventos externos, baseados no comportamento de consumidores.



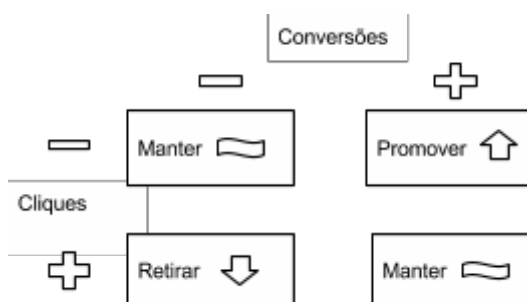
### 3.3. Ranking de produtos

Utilizando Cliques entre Aquisições será formado um ranking de produtos, os que apresentarem muitos cliques desde a última aquisição serão removidos do Feed, deixando mais recursos para os produtos com melhor taxa de conversão.

No caso de comparadores, o consumidor já está decidido pela compra e está avaliando as formas de adquirir entre diversos anunciantes, a qual oferece as melhores condições e segurança de compra, que de forma diferente de uma abordagem mais comum de *Customer Relationship Management* (CRM), em que se busca entender o comportamento do consumidor, bem como recomendar itens baseados em seu perfil, no cenário corrente o consumidor não pode ser considerado, o comparador não oferece ao vendedor formas de seleção de público, ou o que mostrar para cada pessoa, ficando totalmente a cargo do comparador e do consumidor o que será apresentado, algoritmos como *outliers* e *clusters*, baseados em comportamentos de consumidor não são aplicáveis. Outras dimensões também não podem ser escolhidas como a de horário, localidade ou características do público, das dimensões entre as poucas que podem ser trabalhadas estão: o preço, o tempo de entrega e a posição do produto na vitrine.

Pode-se avaliar o desempenho de cada produto pelas métricas de conversão e quantidade de cliques, quanto mais vendas e menos cliques melhor é o desempenho do produto, avaliando suas combinações apenas como baixa e alta se tem quatro possíveis casos, descritos a seguir, conforme definidos por PERLICH; DALESSANDRO; HOOK; STITELMAN; RAEDER (2013):

Figura 4.3 - Relação entre cliques e conversão



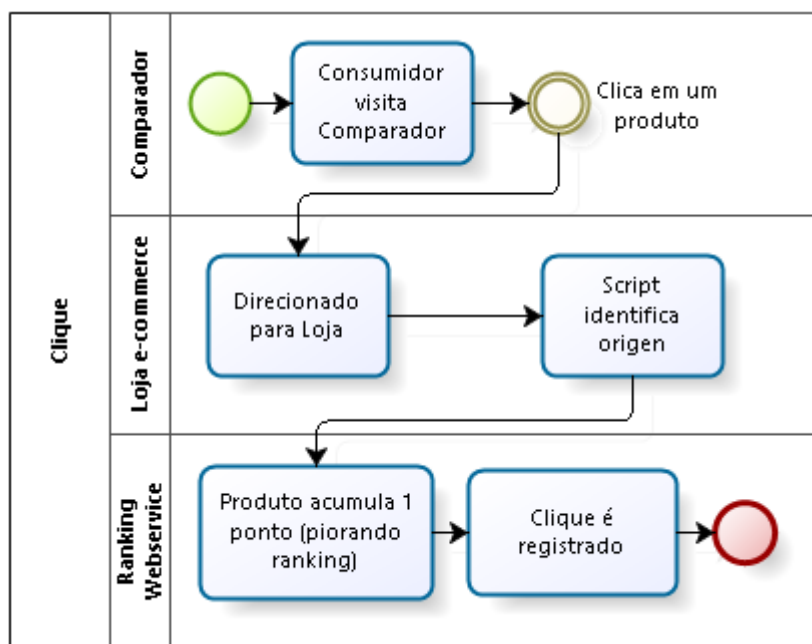
- Baixa conversão e poucos cliques - A quantidade de vezes que um produto é exibido não é cobrada fazendo com que produtos, que não vendem e não são procurados, acaba por se tornar uma boa maneira de exposição gratuita da marca e permitem que novos produtos, ainda desconhecidos pelo público, sejam apresentados com baixo custo inicial.

- Baixa conversão e muitos cliques - São produtos deficitários do ponto de vista de mídia, e principal foco deste trabalho, eles precisam ter a exposição reduzida, mas não devem ser removidos permanentemente, sendo necessário que, eventualmente, apareçam, pois se trata de uma condição momentânea, que é causada por fatores de muita variação de preços, condição de pagamento e promoções.
- Alta conversão e poucos cliques - É o cenário ideal e raro de acontecer e indica produtos com pouca concorrência, com público fiel, que sabe exatamente o que está procurando, normalmente, este público procura o produto diretamente no *site* ou loja do vendedor sem utilizar meios intermediários como comparadores.
- Alta conversão e muitos cliques - Mesmo apresentando custos acima do ideal são justificados pelo volume de vendas, representam os principais produtos oferecidos e removê-los do feed depende de decisões executivas, que podem envolver até remodelar o produto para mantê-lo competitivo.

Cabe lembrar que se deseja evitar apenas a alternativa que apresenta muitos cliques e poucas conversões. Para tanto, foram elaborados os seguintes processos:

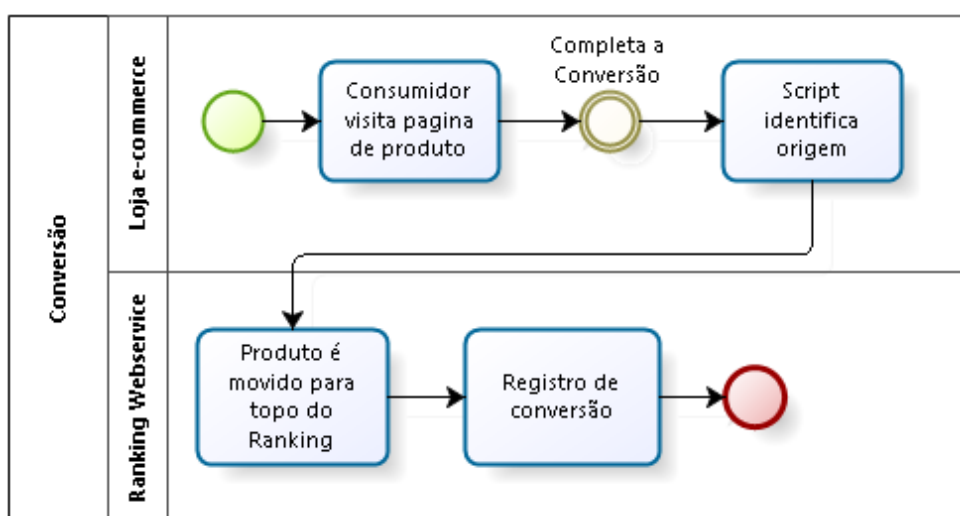
- Escore - Quanto mais baixo o escore melhor a posição no *ranking*.
- Clique - O produto ganha um ponto no escore, prejudicando seu *ranking*.

Figura 4.4 - Processo de clique em um produto



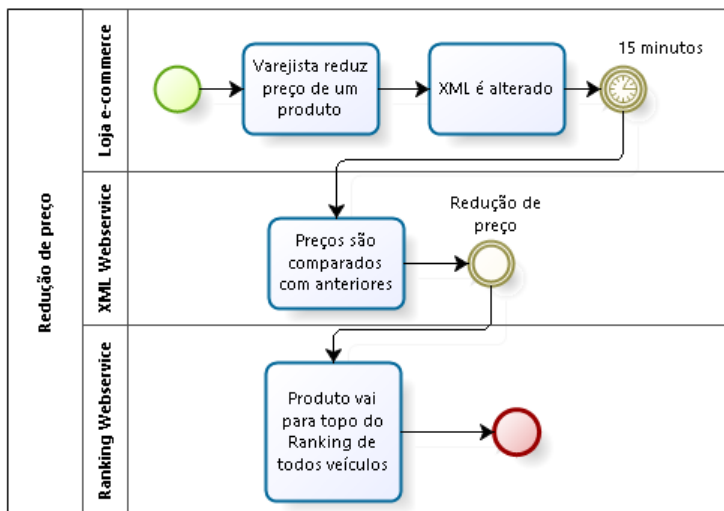
- Conversão - O escore é zerado e o produto volta ao topo do *ranking*.

Figura 4.5 - Processo de conversão



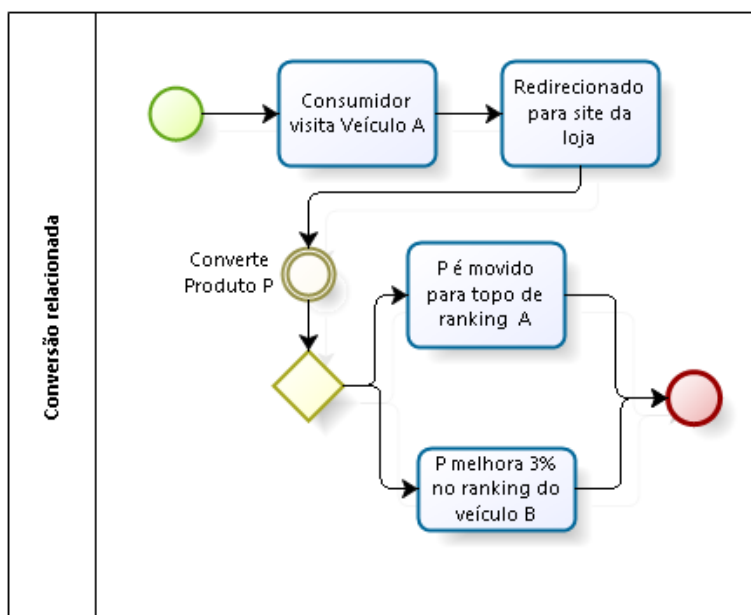
- Redução de preço - O escore é zerado e o produto volta ao topo do *ranking*.

Figura 4.6 - Processo de redução de preço de produtos



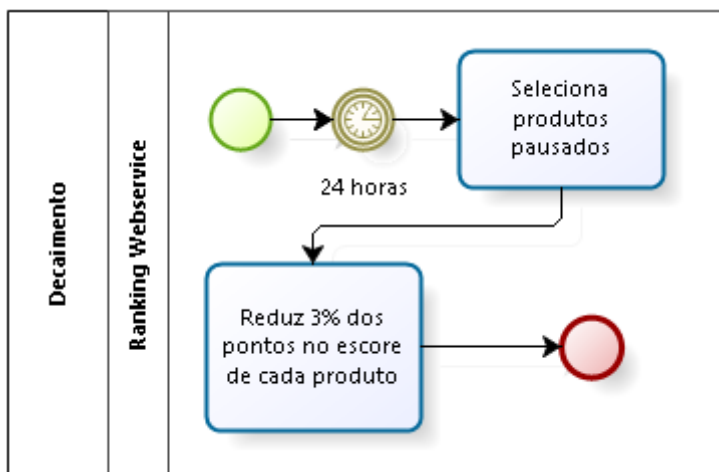
- Conversão externa - O objetivo é utilizar informações de outros veículos, em que o mesmo produto é anunciado, de forma que quando ocorrer uma conversão, em veículos semelhantes, o produto tenha seu escore melhorado obtendo maior exposição.

Figura 4.7 Processo de conversão em origens diferentes



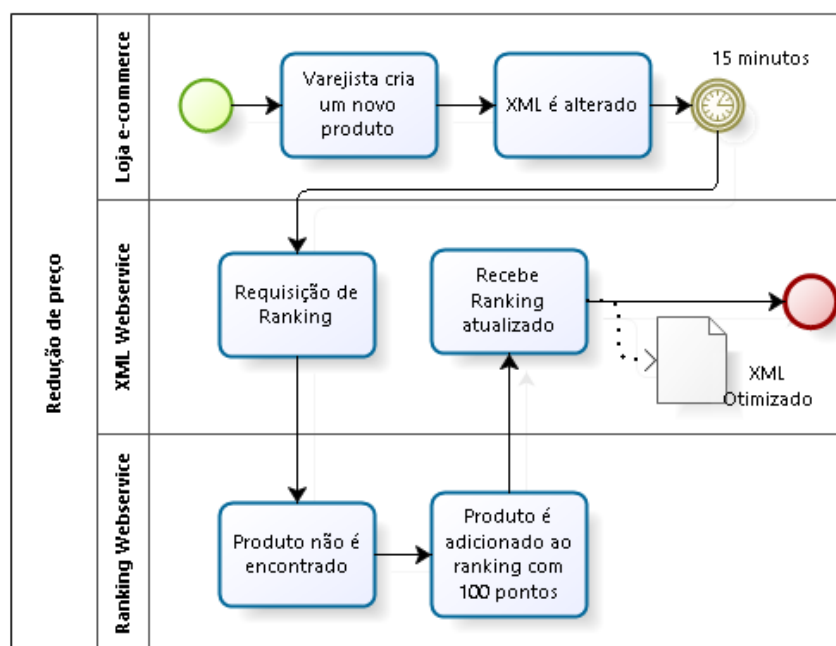
- Decaimento - Uma vez por dia os produtos pausados têm seu escore reduzido em 3%.

Figura 4.8 - Processo de decaimento de produtos pausados



- Novos produtos - Novos produtos iniciam com o escore de 100 pontos

Figura 4.9 - Processo de criação de novos produtos



### 3.4. Adaptação do *Multi-armed Bandit*

Existem muitas adaptações deste problema, basicamente é um problema de otimização com o objetivo de maximizar a receita, se for utilizada a variação gulosa ele permanecerá focado nos produtos com melhor desempenho, mas neste caso é importante que todos os produtos tenham visibilidade, mesmo os produtos considerados como não lucrativos, por questões de exposição de marca. Existem variações híbridas, que utilizam decaimento, em que escores ruins ao longo do tempo vão sendo melhorados, permitindo a definição em um grupo sub-ótimo de opções. Segundo Scott (2010) essa variação é um desperdício, mas atende ao objetivo de visibilidade, então, mesmo que o custo não seja convertido em receita, possibilita a exposição e exploração, pois constantemente produtos de baixo desempenho serão reavaliados, ganhando a oportunidade de melhorarem seus escores. Os melhores caça-níqueis serão os produtos com a maior probabilidade de conversão.

Para utilizar este algoritmo é preciso estabelecer informações como recompensa, custo e risco. A recompensa é igual para todos os produtos, possuindo diferentes ganhos envolvidos, uma vez que esses valores não são acessíveis, o desempenho é medido apenas por peças vendidas, então, todos os caça-níqueis fornecem exatamente a mesma recompensa. O custo é o CPC x Cliques que será adotada na simplificação de todos os produtos com o mesmo CPC, fazendo com que o custo atual seja simplesmente a quantidade de cliques desde a última recompensa. O Risco é baseado na probabilidade de conversão de cada produto, assim, produtos bem ranqueados possuem poucos cliques entre as conversões e produtos mal ranqueados recebem muitos cliques e raramente convertem, nesse ponto será utilizada uma estratégia gulosa, os produtos com menos cliques serão os escolhidos com melhor chance de recompensa. Para garantir a que todos apareçam no *feed*, eventualmente, se vai utilizar uma estratégia de decaimento, em que a cada ciclo é subtraído uma parte do excesso de cliques, que prejudica o escore do produto, isso gera uma região sub-ótima em que produtos com desempenho em declínio serão, constantemente, confrontados com produtos beneficiados pelo decaimento.

### 3.5. *Starvation-freedom* baseado em decaimento

Utilizando o conceito de *starvation-freedom* será reservado espaço para produtos que estão fora do *feed* por muito tempo, oferecendo a este a oportunidade de converter e melhorar seu escore para voltar ao *feed*, se o produto continuar com mau desempenho será, gradualmente, rebaixado no ranking e, por fim, removido do feed novamente. Essa propriedade será garantida pelo decaimento utilizado no *Multi-armed Bandit*.

### 3.6. *Keywords* em comparadores

Os algoritmos *Multi-armed Bandit* e *Knapsack Auctions* foram, inicialmente, propostos para se encontrar os melhores *keywords* em motores de busca como Google e Yahoo (RUSMEVICHIENTONG; WILLIAMSON, 2007). Em comparadores não é possível comprar termos de busca, mas o consumidor navega sobre categorias e produtos, considerando categorias e produtos como *keywords* predefinidas, em que não é possível escrever, mas é necessário escolher as existentes, está-se falando de um subconjunto mais simples, em que é possível utilizar os mesmos algoritmos com os mesmos resultados e um custo computacional ainda menor, pois além de reduzido o conjunto é previamente conhecido.

### 3.7. **Resumo do capítulo**

Neste capítulo é apresentado o funcionamento do método proposto e conceitos que foram adaptados para o problema de otimizar comparadores. Os desenvolvimentos mais importantes são o *Ranking* de Produtos e o indicador Cliques entre Aquisições, que foram a base para pesquisa e aplicação dos demais conceitos e algoritmos.

## **4. AVALIAÇÃO**

Diversos algoritmos que poderiam contribuir com a melhoria de desempenho foram avaliados, entre eles o ARIMA para séries temporais, análise de elasticidade de preços e um algoritmo de Ranking baseado em teoria dos jogos. A seguir é feita uma descrição dos resultados obtidos.

### **4.1. Séries temporais**

O principal problema encontrado com a utilização do método ARIMA foi a necessidade de relacionar como diferentes métricas se relacionam com a conversão, mesmo oferecendo um ótimo desempenho para prever indicadores em uma série temporal, que se mostrou mais válido neste trabalho na utilização de teoria de jogos estocástico, em que os jogadores concorrentes precisam avaliar recompensa e riscos. A avaliação de eventos históricos acabou prejudicando a tomada de decisões, novos estados apresentam comportamentos não sequenciais e requerem reações, que não estão relacionadas com o passado.

### **4.2. Reação por correlação**

Combinado com o mecanismo de decaimento, a correlação melhorou a descoberta de produtos, pois os produtos com baixo desempenho não precisam ficar totalmente escondidos do público, anunciando em poucos lugares se reduz o custo global com o produto e é utilizado como uma rede de reconhecimento, que coloca o produto em exposição, quando ocorrem eventos externos que, potencialmente, representam boas oportunidades, mas com janela de tempo limitada e precisam ser exploradas imediatamente.

### **4.3. Elasticidade de preços**

A elasticidade de preço apresentou alguns problemas, o primeiro é que a comparação de preços realizada pelos serviços disponíveis compara exatamente o mesmo produto em diferentes concorrentes, mas o consumidor analisa também produtos concorrentes de diferentes marcas, neste caso não existe uma relação direta entre os preços, uma vez que o consumidor pode aceitar pagar



mais por determinada marca, assim mesmo com um preço menor é possível que o produto ofertado não seja a melhor opção.

Mesmo no caso em que se compara, exatamente, o mesmo produto ocorre o problema do frete, mesmo com um preço acima da média, as diferenças no valor do frete agregam o valor total do produto e acabam distorcendo a análise de preços.

Por estas razões se decidiu trabalhar apenas com variações do próprio preço, sem analisar concorrentes, assumindo que quando um produto tem seu preço reduzido, o produto já passou por uma análise de concorrência e o novo preço está competitivo, então quando um produto reduz de preço seu ranking é imediatamente melhorado para garantir sua exposição.

#### 4.4. Ranking

Os produtos listados no apêndice A são o *ranking* de produtos baseados em uma semana de coleta, resultado da interação entre cliques e conversões do *e-commerce* analisado. A lista é composta por 198 produtos, que foram divididos para análise em 5 grupos de 40 produtos, segundo sua qualificação no *ranking*. O modelo utilizado na formação no *ranking* foi capaz de ordenar, corretamente, os produtos com melhor desempenho nas primeiras posições. Um corte agressivo seria possível reduzir custos de 23% com 3% de redução na receita. Um corte conservador de 8% no custo não teria nenhuma perda na receita.

Tabela 5.1 – Eficiência de produtos por grupos de ranking

	Eficiência	Custo	Conversões
Melhores	4,99%	1483	74
Sub-ótimos	2,82%	5184	146
Intermediarios	0,72%	697	5
Baixo desempenho	0,44%	675	3
Piores	0,00%	648	0

Figura 5.1 – Eficiência de produtos em função da posição no *ranking*



A ordenação não é baseada na eficiência, mas na probabilidade de conversão a partir de eventos recentes, a correlação é suficiente para demonstrar que os produtos escolhidos com maior probabilidade acabam apresentando maior eficiência. Permitindo que sob condições de orçamento restrito seja possível ajustar a eficiência até que a restrição de orçamento seja satisfeita. Metas de desempenho podem, eventualmente, serem restritas, em que acima ou abaixo do especificado são considerados resultados ruins, nesses casos a meta deve ser atingida com o valor exato, pois uma eficiência muito alta resulta em baixo volume de vendas. Tanto em metas restritas quanto livres podem ser atingidas aumentando ou reduzindo a margem de corte dos produtos no ranking.

#### 4.5. Resumo do capítulo

Das abordagens avaliadas, as que apresentaram os melhores resultados e viabilidade foram: a reação por correlação, em que o produto é acompanhado em veículos e o sistema de *ranking* que ordena os produtos por maior probabilidade de compra e menor custo de aquisição ao mesmo tempo, foi necessário um período de coleta de dados inicial para permitir um bom espalhamento no ranking e que produtos importantes fossem reconhecidos.

Durante esta avaliação foram percebidas diversas possibilidades de melhorias, que permitirão a escolha de diferentes metas, como por exemplo, volume, orçamento, ou eficiência, todos utilizando o mesmo mecanismo de aumentar ou diminuir a quantidade de produtos removidos.

## 5. CONCLUSÕES

Este trabalho apresentou uma solução para auxiliar varejistas a melhorarem seus resultados em sites de comparadores, selecionando os produtos com maior probabilidade de venda com um algoritmo adaptativo baseado no problema do *multi-armed bandit*, que trata as duas fases do problema, descoberta (*exploration*) e utilização (*exploitation*) simultaneamente, atualizando constantemente a melhor opção a cada novo evento. O algoritmo possui uma complexidade inferior aos estudos semelhantes, pois não adota uma abordagem genérica, mas propõe uma solução específica, utilizando características conhecidas do problema analisado.

A principal dificuldade encontrada foi com a abordagem inicial, baseada em análise de dados complexos, comuns no meio de marketing, que dispõe de muitas informações sobre o consumidor e seu comportamento. No entanto, estudando o caso de comparadores se verificou que se trata de um problema de natureza estocástica, em que os estados possíveis apresentam pouca probabilidade de se repetirem ou de formarem ciclos de sazonalidade, em que informações históricas oferecem pouco suporte para tomada de decisões. O problema em questão apresentou diversas afinidades com teoria dos jogos e cadeias de Markov, em que existe a competição entre jogadores e cada novo estado está relacionado a um conjunto diferente de decisões.

Outra dificuldade está relacionada com indicadores utilizados no meio do marketing, a tentativa de maximizá-los levava a resultados que não se convertiam em ganho, pois existe um comportamento empírico entre os analistas de marketing difícil de extrair em entrevistas, na verdade, os indicadores utilizados servem apenas de intuição e comparação, mas suas decisões são baseadas em experimentação e são particulares para cada caso.

Por fim, abandonar abordagens comuns ao domínio de origem, o marketing, como a forma de analisar os dados, forma de medir resultados e as tomadas de decisão se mostraram importantes para conclusão deste trabalho, que se mostrou alinhado com soluções semelhantes desenvolvidas com abordagem de aprendizado de máquina para problemas de publicidade digital.

## REFERÊNCIAS

- SCOTT Steven L., **A modern Bayesian look at the multi-armed bandit** In: APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY, 26., Storrs, CT: John Wiley & Sons, Ltd., 2010. p. 639–658
- AGGARWAL G.; HATLINE J.D. **Knapsack Auctions**. In: 1st Workshop on Sponsored Search Auctions in conjunction with the ACM Conference on Electronic Commerce (EC'05). 5., Vancouver, BC, 2005. p.1083-1092
- CHAKRABARTY, D.; ZHOU, Y.; LUKOSE , R. **Budget Constrained Bidding in Keyword Auctions and Online Knapsack Problems**. Palo Alto CA: 2007
- MCSTAY Andrew. **Digital Advertising** . 1st ed., LONDON, N1: Palgrave Macmillan, 2009
- RUSMEVICHIENTONG, Paat; WILLIAMSON, DAVID P. **An Adaptive Algorithm for Selecting Profitable Keywords for Search-Based Advertising Services**. 1st ed., Ithaca, NY: Cornell University, 2007
- TRAN–THANH, Long ; CHAPMAN, Archie; ROGERS, Alex; JENNINGS, Nicholas R. **Knapsack Based Optimal Policies for Budget–Limited Multi–Armed Bandits** In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. 26., Southampton, SO17 1BJ: Association for the Advancement of Artificial Intelligence , 2012. p. 1134-1140
- Stokes, Rob **eMarketing - the essential guide to online marketing**, 1st ed, London SE1 0ES: Quirk eMarketing (Pty) Ltd adress, 2008. ISBN: 978-0-620-41135-6 2008
- HU, Yu; SHIN, Jiwoong; TANG, Zhulei **Performance-based Pricing Models in Online Advertising: Cost per Click versus Cost per Action**, New Haven, CT: Yale University 2013
- PERLICH, Claudia; DALESSANDRO, Brian; HOOK, Rod; STITELMAN, Ori ; RAEDER, Troy Bid Optimizing and Inventory Scoring in Targeted Online Advertising in: **Knowledge Discovery and Data Mining**. 12., Beijing: Media6Degrees, 2013. p. 804-812
- JEFFERY , Mark **Return on Investment Analysis for E-business Projects**, Evanston, IL: Northwestern University, 2003
- TANENBAUM, Andrew S. **MODERN OPERATING SYSTEMS**, 3rd ed., HV Amsterdam: Pearson, 2007 ISBN-13: 978-0136006633

AUER, Peter; CESA-BIANCHI, Nicolò; FISCHER, Paul **Finite-time Analysis of the Multiarmed Bandit Problem**. In: Machine Learning, 47, Dordrecht: Kluwer Academic Publishers, 2002. p. 235-256

NGO, Theresa Hoang Diem **The Box-Jenkins Methodology for Time Series Models**. In: SAS Global Forum 2013. Burbank, CA: Warner Bros. Entertainment Group. 2013.

MAKRIDAKIS, Spyros; WHEELWRIGHT, Steven; HYNDMAN, Rob J. **Forecasting: methods and applications**. 1st ed., New York: John Wiley & Sons. 1998. ISBN 0-471-53233-9

## APÊNDICE A - COLETA DE DADOS

Ranking de produtos baseado em coleta de dados realizada no período de 16/6/2015 a 22/6/2015

sku	escore	total clicks	total conversions
BO360A	0	55	5
BLB06A	0		
CWG11A	0	140	8
BMW20A	0	0	1
BWB08A	0	21	1
CBV12D	0	0	1
BYS5CA	0	1	
R35211	0		1
BI902A	0		1
BDH30A	1	14	1
CF376B	1	8	0
CMA30A	1	67	3
CWK12A	2	11	1
CHB53C	2	171	5
CF476B	3	5	0
BYS4GA	4	28	2
BA190A	4	10	1
BSX10A	4	4	0
BYS5VA	4	22	2
BFS5TA	5	12	1
BLF08A	5	162	8
BDD85A	5	79	2
BFS4GA	5	5	0
BVR28H	6	59	2
CWE11A	6	41	2

BMK45A	6	52	3
BYS5QA	7	13	0
BSR10A	7	48	4
BDD62A	7	21	0
BWL09B	7	74	2
CRM51A	8	39	2
CHB42D	8	21	1
BDD61A	9	45	4
CHA22D	9	10	0
CVU18G	9	33	1
BRE51N	9	38	2
CO060A	9	101	5
BFD5QA	10	60	1
CME20A	10	13	1
CRB36A	10	21	1
BLF12A	10	169	6
CD060A	10	24	1
BRE80A	11	222	6
BAI91B	11	13	0
CF576B	12	33	1
BFD5VA	12	21	0
CF676A	13	29	2
BRA08A	13	67	0
CF350A	14	17	1
BFO4TA	15	33	3
CWC08A	15	18	1
BWU11A	15	44	0
CRM55A	16	76	2
CRD37E	16	28	0
BWL11A	17	917	43
BZC12B	18	33	1
BWG12A	18	60	1

BRS62C	19	39	0
BDD75A	24	76	3
BAI60B	27	35	0
BRM48N	29	459	10
CZD12A	29	188	3
BOA61A	30	111	2
BYO4TA	34	118	8
BRM50N	34	514	14
BWG11A	35	169	1
BRE50N	36	752	24
BFS5VA	37	47	1
CRD36G	37	61	1
CWE08A	38	73	3
BRM42E	40	167	4
CHA31C	42	75	2
BRM39E	42	200	0
BMA30A	47	40	0
CD075A	49	53	0
BNS10A	53	12	0
BRO80A	57	84	1
BMT45B	57	86	0
BMS45B	59	14	0
BMJ38A	62	12	0
BWC10B	64	144	2
BO260A	65	25	0
CRC30H	65	94	3
CA090B	66	5	0
BOB61A	68	29	0
BLF10A	69	1	0
CF550B	69	53	0
BMS26A	70	11	0



BWH15A	71		
CI901A	75	3	0
BO160A	77	22	0
BZB31A	77	4	0
CF350B	77	5	0
BFS5QA	79	35	0
C7W07A	80	11	0
BSI10A	80	17	0
CRC08A	81	1	0
BZA08A	81	1	0
BAA60E	82	13	0
BLB14C	82	2	0
CA060B	84	8	0
BFO4NA	84	30	0
BA290A	85	13	0
CMD20A	88	3	0
BDG30A	88	7	0
CF150A	88	8	0
BMD35A	89	14	0
CRM52A	89	21	0
BDT85A	89	5	0
C1F07A	89	8	0
CRC12C	89	11	0
BF150A	91	11	0
BMJ23A	91	5	0
BFS4NA	91	6	0
BA390A	91	9	0
CVU26E	92	29	0
C1R07A	92	7	0
CVU30E	93	18	0
BRX50C	93	3	0

BFS6NA	94	18	0
CMW20A	95	6	0
CRC08C	95	13	0
CWE10A	96	10	0
BMO40A	96	11	0
CRB39A	97	37	0
BMF45A	97	1	0
BWP11A	97	3	0
BA890B	97	10	0
BFD4NA	97	2	0
CRM33E	97	400	3
BFS4TA	97	2	0
BFD4TA	100	10	0
CF475A	100	8	0
CF750A	100		
CWG12A	100	51	0
BAT80A	100	5	0
BVG24H	100	18	0
BAA80E	101	6	0
BFS5CA	101	1	0
BF775B	101	6	0
CMP25A	101		
BF760A	101		
BWI01A	102		
BDJ80A	102	2	0
BF375A	102	1	0
CFN50C	102	2	0
BF976B	102	1	0
BYS5TA	102	2	

BDF30A	102	2	
BF076C	102	2	0
BF060A	102	7	0
CWI07A	102	2	0
C1R06A	103	1	
BRN80A	103	2	
BDJ30A	104	8	0
CF575A	104	4	0
BNQ11D	104	4	0
BFD5SA	104	1	0
CVT10B	105	5	
BFS5NA	105	29	0
BZA12A	106	6	0
CM020B	106	40	0
BYS6NA	106	12	0
CVU20G	106	12	0
CWN07A	106	3	0
BWC09A	106	3	0
CWI06B	106	33	0
CWK11A	106	24	0
BRB39A	107	5	0
CRG36A	107	3	0
CMK38A	107	12	0
CMS26A	108	6	0
BOG40A	108	13	0
CF450B	108	17	0
CRC28F	109	37	0
BFT60A	110	8	0
CFU50A	110	7	0
BMC20A	111	49	0

BRS70H	111	21	0
CMY34A	111	43	0
CRP28C	111	10	0
CWL75A	111	8	0
BRF36G	111	11	0
CFC50D	112	12	0
BRV80A	112	18	0
BA790B	112	12	0
BWC08A	112	22	0
CRA30F	112	12	0
CRD48F	113	19	0
CRM37E	113	36	0
CWC10A	113	29	0
BRK50N	114	20	0
BRC12X	115	14	0
BAT60A	115	14	0
BMF45B	118	18	0
BLF06A	118	11	0
CFS50A	127	27	