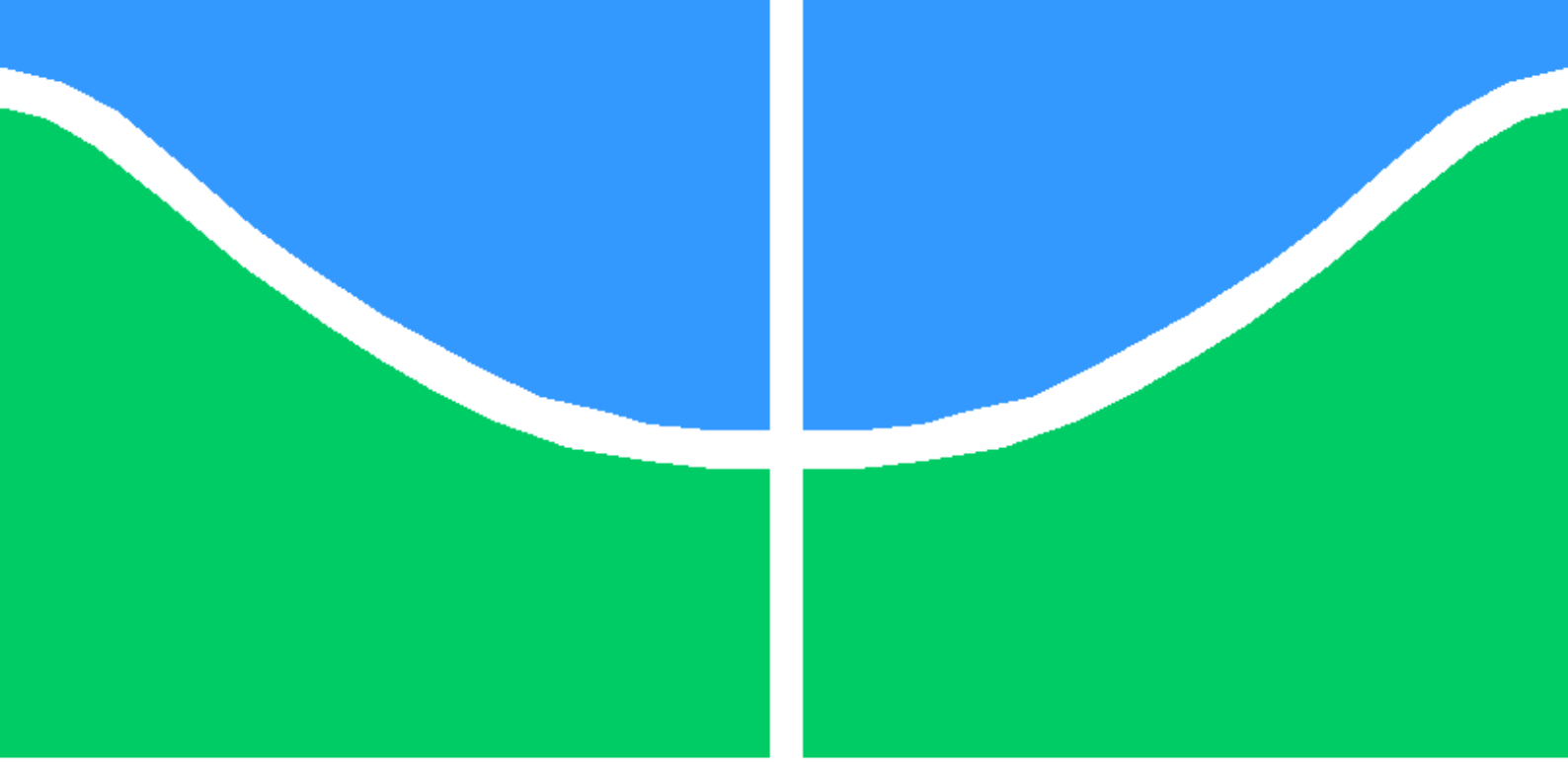

TRABALHO DE GRADUAÇÃO

APRENDIZADO POR REFORÇO APLICADO AO MERCADO FINANCEIRO

Breno Rodrigues Brito

Dyego Soares de Araújo

Brasília, Dezembro de 2013

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASILIA
Faculdade de Tecnologia

TRABALHO DE GRADUAÇÃO

APRENDIZADO POR REFORÇO APLICADO AO MERCADO FINANCEIRO

Breno Rodrigues Brito

Dyego Soares de Araújo

Relatório submetido ao Departamento de Engenharia Elétrica como requisito parcial para obtenção do grau de Engenheiro Eletricista

Banca Examinadora

Prof. Alexandre Romariz, ENE/UnB _____
Orientador

Bruno Guimarães, ENE/UnB _____
Co-orientador

Prof. Ricardo Zelenovsky, ENE/UnB _____
Examinador interno

Dedicatórias

Dedico este trabalho a todas as pessoas que, para o bem ou para o mal, me tornaram a pessoa que sou hoje.

Dyego Soares de Araújo

Dedico este trabalho a todos os meus amigos que, de uma forma ou de outra, estão constantemente reclamando que não tenho mais tempo para eles, em especial à minha namorada que faz meus outros trabalhos para me concentrar neste. Amo vocês.

Breno Rodrigues Brito

Agradecimentos

Gostaria de agradecer primeiramente a meus pais e minha família, motivos de eu estar aqui hoje e poder aproveitar todas as oportunidades de que tirei proveito. Em seguida, agradeço à minha namorada, que sempre me apoia com uma dedicação inabalável. Agradeço também a meus amigos, fonte inesgotável de diversão e aprendizado.

Impossível deixar passar os agradecimentos aos mestres, em especial ao professor Romariz, meu orientador, e ao professor Marco Cézar da Física, que considero como amigo. Junto a isso, agradeço também aos funcionários da secretaria de Engenharia Elétrica da UnB que sempre foram extremamente prestativos.

Ainda, gostaria de agradecer a DX Investimentos e ao Bruno Guimarães por apoiar meu trabalho e me fornecer os dados para análise, além de insights importantes do mercado financeiro e livros emprestados que abusamos para a realização desta monografia.

Por último, mas não menos importante, muito pelo contrário, gostaria de agradecer ao Dyego, meu colega e amigo que fez este trabalho comigo. Um trabalho tão vasto como este, ousado e tão novo para nós, alunos de graduação da engenharia elétrica, não poderia acabar tão bem sem a nossa dedicação e cooperação mútua.

Breno Rodrigues Brito

É impossível concluir essa etapa da minha vida sem ter muitas coisas pelas quais agradecer, e muitas pessoas para receber esses agradecimentos. Sem mais delongas, inicio o árduo processo de conferir mérito àqueles que muito o merecem por tudo o que fizeram por mim.

Inicio com minha família por motivos bem óbvios. Eles são as pessoas que me amam desde que eu nasci e que, em boa parte, me fizeram ser o que eu sou hoje. Agradeço aos meus pais Benjamim e Rosimary que com tanta paciência me criaram. Deixo um agradecimento especial para o meu irmão Thiago, uma vez que para o bem ou para o mal ele é responsável por uma parte incalculável da minha personalidade e do meu jeito de ser.

A minha próxima leva de agradecimentos é dedicada aos meus amigos. Estes constituem de fato uma segunda família para mim e eu não consigo imaginar a minha vida sem eles. Pela paciência infinita e por lidar com minha loucura (literalmente) e por incontáveis horas de DoTA, devo agradecimentos especiais a meu primo Marcos Henrique e meus amigos Gabriel Aleixo e Luis Cesar.

Impossível deixar de agradecer a todos os meus amigos que conheci ao cursar elétrica. Deixo um agradecimento especial para toda a minha turma. Quero que saibam que eu me considero a pessoa mais sortuda do mundo por ter tido o privilégio de conhecer vocês. Também gostaria de agradecer ao pessoal do CAFIS que em pouco tempo se tornaram amigos valorosos e boas fontes de opinião. Em especial, deixo um agradecimento ao Lucas Camacho pelos comentários valiosos efetuados no segundo capítulo.

Uma terceira leva de agradecimentos é destinada a meus mestres, que por meio de seus ensinamentos me imbuíram de habilidades técnicas para ser um engenheiro, e também de habilidades humanas para ser uma pessoa melhor. Alguns professores, no entanto, merecem agradecimentos especiais. Agradeço ao Prof Anésio pela excelente oportunidade de trabalhar em seu laboratório e pelas inúmeras experiências adquiridas neste ambiente. Agradeço também professor Romariz pela paciência e orientação fornecidos para este trabalho. Ao Prof. Célius Antônio pelas inúmeras horas de monitoria, pelas excelentes conversas e pela carta de recomendação que mudou meu futuro. Também um agradecimento especial ao Prof. Franklin que me apoiou durante a época mais complicada da minha vida.

Por fim, devo agradecer imensamente ao Breno pelas horas de trabalho que compartilhamos. Ele é o melhor parceiro de trabalho final que alguém poderia querer, além de um excelente amigo. Nós alcançamos um nível de cooperação e amizade sem igual. Acredito que eu seja uma pessoa de muita sorte por te-lo como amigo pessoal

Dyego Soares de Araújo

RESUMO

Este trabalho visa mostrar como um sistema inteligente de aprendizagem por reforço pode se utilizar de indicadores financeiros clássicos para superar várias estratégias clássicas de investimento na bolsa de valores. Devido à natureza não-linear, aleatória e não-estacionária dos mercados financeiros, várias estratégias clássicas deixam brechas onde poderiam comprar instrumentos financeiros e lucrar com eles. Para este fim, foi construído um sistema onde a estratégia de compra usa o algoritmo SARSA de aprendizagem por reforço enquanto a de venda é feita através de várias estratégias clássicas. Para testá-lo, foi feito um sistema que usa uma estratégia de venda idêntica enquanto a de compra é baseada em um indicador clássico. Resultados mostraram que o algoritmo inteligente conseguiu no final retornos mais estáveis e até seis vezes maiores que os das estratégias clássicas testadas.

Palavras-chave: Aprendizagem por Reforço, Aprendizagem de Máquina, Bolsa de valores, SARSA, day trade

ABSTRACT

This work aims to show how an intelligent system based on reinforcement learning can benefit of classical financial indicators to overcome classic trading strategies in the stock market. Due to the non-linear, random and non-stationary nature of the financial markets, various classic strategies fail to benefit from all opportunities where they could be making profit. In order to achieve that, a system was built where only the buying strategy uses reinforcement learning SARSA algorithms while the sell is made through various classic strategies. To test it, a system that uses an identical selling strategy was constructed, where the purchase was based on a classic indicator. Results show that the intelligent algorithm achieved more stable returns, up to six times higher than the ones from tested classical strategies.

Keywords: Reinforcement Learning, Machine Learning, Stock Market, SARSA, day trade

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SÍMBOLOS

Acrônimos Financeiros

HME	Hipótese do Mercado Eficiente
MMA	Média Móvel Aritmética
MME	Média Móvel Exponencial
OBV	On Balance Volume
$A\backslash D$	Accumulation\Distribution Line
MFM	Multiplicador de Fluxo Monetário
FV	Fluxo de Volume
TR	True Range
ATR	True Range Médio
DM_+	Movimento Direcional Positivo
ADM_+	Movimento Direcional Positivo Médio
DM_-	Movimento Direcional Negativo
ADM_-	Movimento Direcional Negativo Médio
DI_+	Índice Direcional Positivo
ADI_+	Índice Direcional Positivo Médio
DI_-	Índice Direcional Negativo
ADI_-	Índice Direcional Negativo Médio
DX	Índice Direcional
Bol^+	Banda de Bollinger Positiva
Bol^-	Banda de Bollinger Negativa
$MACD$	Moving Average Convergence Divergence
$HiLo$	HiLo Activator
S	Sharpe Ratio

Notação Matemática

a	Ação Presente
a^+	Ação Futura
s	Estado Presente
s^+	Estado Futuro
$Q(s,a)$	Valor Estimado da Função Q no Estado s tomada a ação a
$Q^*(s,a)$	Valor Real da Função Q no Estado s tomada a ação a
r_j	Recompensa recebida no instante j
$\alpha(k)$	Tamanho do Passo de Aprendizagem no instante k
α	Tamanho Constante do Passo de Aprendizagem
R_j	Retorno associado ao instante j
γ	Fator de Desconto Temporal na Aprendizagem
$Pr\{E_j = e \vert E_{j-1} = e_0\}$	Probabilidade que E_j seja igual a e, dado que E_{j-1} é igual a e_0
$P^a_{ss'}$	Probabilidade de se chegar ao estado s' partindo do estado s tomando a ação a
$R^a_{ss'}$	Retorno esperado ao se chegar ao estado s' partindo do estado s tomando a ação a
$E\{R \vert E = e_0\}$	Esperança do Valor R, dado que E é igual a e_0
π	Política
$\pi(s,a)$	Probabilidade de se Tomar a ação a no estado s seguindo a política π
$Q^\pi(s,a)$	Valor da Função Q no Estado s tomada a ação a, seguindo a política π
ϵ	Fator de Aleatoriedade da Política ϵ-Gananciosa
τ	Fator de Temperatura na Política SoftMax
λ	Coeficiente de Elegibilidade
$TD(\lambda)$	Diferenças Temporais com Fator λ
$e(s,a)$	Elegibilidade do par Estado-Ação (s,a)

Capítulo 1

Introdução

"A força mais poderosa do universo é o juro composto" - Albert Einstein

1.1 Contextualização

Diz o Dr. Alexander Elder[1] que investir no mercado de ações é a segunda empreitada humana mais perigosa, perdendo apenas para a guerra[2]. De fato, ao se investir de maneira descuidada, sem compreender de maneira adequada a dinâmica do mercado, corre-se o risco de perder muito dinheiro. Por esse motivo, qualquer vantagem que puder ser utilizada para auxiliar o processo de tomada de decisões relativas a compra e venda de ativos deve ser empregada tanto quanto for possível.

Em busca destas vantagens, diversos investidores desenvolveram técnicas de observação das tendências do mercado. Algumas eram gráficas e visuais, tais como traçar retas de suporte e resistência em um gráfico de um dado ativo, enquanto outras empregavam algum grau de matemática e estatística, os chamados indicadores. De posse destes indicadores, diversas estratégias foram desenvolvidas. Essas estratégias se mostram vantajosas, desde que se possua disciplina para empregá-las de maneira adequada e empregue-as conjuntamente com um bom sistema de gerência de riscos.

Entretanto, com o desenvolvimento da computação e o crescente melhoramento de sistemas inteligentes, existe uma tendência atual da utiliza-los aplicados ao Mercado Financeiro. Uma tipologia interessante é a de sistemas de Aprendizagem por Reforço. Sua principal capacidade é aprender a tomar decisões corretas em um ambiente carregado de incertezas, como o mercado de ações. Além disso, é capaz de incorporar dentro de si a dinâmica do problema, sem necessitar de uma modelagem específica. Seu grande poder de adaptação faz dos sistemas de Aprendizagem por Reforço candidatos ideais para o emprego na bolsa de valores.

[1] O Dr. Alexander Elder é um psiquiatra russo que se mudou para os Estados Unidos durante a Guerra Fria. Nessa época, tornou-se um investidor e passou a estudar a psicologia dos mercados, escrevendo vários livros a respeito do tema.

1.2 Definição do problema

Um *trader* [2] em um dado mercado, seja ele um mercado de ações[3], Mercado Futuro[4] ou FOREX[5], sempre possui dois objetivos primários: SOBREVIVÊNCIA e ENRIQUECIMENTO. SOBREVIVÊNCIA diz respeito a quanto ele suporta perder. Nesse sentido, é necessário um controle rigoroso do volume de capital investido e do risco presente em cada ação tomada dentro do mercado, o chamado *Money Management*. Quanto ao ENRIQUECIMENTO, este se dá na forma de resultados consistentemente vitoriosos.

O mercado é operado por meio de um sistema de decisões. A todo momento deve-se decidir se a atitude correta deve ser entrar no mercado ou manter-se de fora. Ao decidir entrar, deve-se decidir quanto do capital disponível será investido, e deve-se avaliar qual o risco que se corre ao entrar no mercado. Uma vez posicionado, deve-se sempre reavaliar a posição e escolher o melhor momento para desmontá-la. O sistema inteligente deve auxiliar portanto este processo de tomada de decisões.

1.3 Objetivos do projeto

Uma vez que o Mercado Financeiro é um ambiente bastante complexo, optou-se pelo desenvolvimento de um sistema inteligente mais simples. Seu objetivo é a indicação de bons pontos de entrada no mercado. A saída ficou a cargo de estratégias clássicas, baseadas em indicadores técnicos financeiros.

Para se medir a efetividade da estratégia traçada pelo sistema, um objetivo auxiliar é a implementação de estratégias de investimento clássicas com a mesma estratégia de venda, a fim de comparar os resultados e justificar a utilização do Sistema Inteligente. Para isso, serão exploradas estatísticas de comparação entre estratégias operando sob um mesmo mercado.

1.4 Apresentação do manuscrito

No capítulo 2 é feita uma familiarização do leitor com o universo do Mercado Financeiro, explicando as principais teorias correntes e a terminologia técnica utilizada. Em seguida, no capítulo 3 é feita uma revisão teórica a respeito de sistemas de Aprendizagem por Reforço. O capítulo 4 trata do o desenvolvimento dos sistemas empregados no trabalho e das estatísticas

[2] Um indivíduo que realiza trocas de ativos em mercados financeiros. A principal diferença entre um *trader* e um investidor é o tempo de manutenção do ativo. Investidores tendem a manter o ativo por um período de tempo maior, enquanto *traders* realizam operações mais imediatas. [1]

[3] Mercado onde se negociam ações. A ação é a menor fração do capital social de uma empresa e é um dos tipos de Instrumentos Financeiros.[1]

[4] Mercado onde se negociam contratos futuros. Os Instrumentos Financeiros chamados de Contratos Futuros são contratos de compra ou venda, onde as partes compradoras e vendedoras se comprometem a negociar determinada quantidade de instrumentos financeiros ou bens físicos.[1]

[5] Um acrônimo da expressão *Foreign Exchange*, Mercado de Câmbio[1]

analisadas, seguido do capítulo 5 que mostra os resultados obtidos por cada estratégia. Por fim, o capítulo 6 encerrará o trabalho com as conclusões, comentários finais e propostas para pesquisas futuras.

Capítulo 2

Análise Clássica do Mercado de Ações

"Be fearful when others are greedy, be greedy when others are fearful" - Warren Buffet

2.1 Introdução

Ao construir um sistema capaz de apoiar decisões Mercado Financeiro, deve-se primeiramente estar familiarizado com este meio, compreendendo as dinâmicas a ele inerentes. Para tal fim, este capítulo será dividido em partes.

Na primeira será explicada a Hipótese do Mercado Eficiente. Na segunda, as operações básicas disponíveis a um agente financeiro. Em seguida, será explicitado um pouco da dinâmica do mercado, além das principais escolas de pensamento que trabalham profissionalmente com gerência de riscos. Serão elucidadas teorias a respeito da movimentação de preços e da formação e extinção de tendências. Por fim, serão desenvolvidos e explicados os principais parâmetros utilizados por agentes financeiros do mundo inteiro ao realizarem seus negócios.

2.2 Hipótese do Mercado Eficiente

Em 1900, o matemático francês Louis Bachelier, em sua Tese de Doutorado "A Teoria da Especulação"[3], criou um modelo de preços aleatório da bolsa de valores, sendo este a maior base para a Hipótese do Mercado Eficiente (HME). Esta hipótese foi formulada nos anos 60 pelo professor Eugene Fama da Universidade de Chicago Booth School of Business[4], o que lhe rendeu o prêmio Nobel de Economia em 2013, conjuntamente com Robert J. Schiller e Lars Peter Hansen.

A Hipótese do Mercado Eficiente postula que os mercados financeiros são eficientes em termos de informação, ou seja, toda informação nova que um indivíduo tem sobre o mercado já se encontra refletida nos preços. Deste modo, é impossivel alcançar consistentemente retornos superiores à média dos retornos do mercado ajustados em relação ao risco, considerando as informações disponíveis no momento em que o investimento é realizado.

Uma analogia interessante pode ser traçada entre um Mercado Eficiente e uma Fila de Supermercado. Suponha que uma pessoa frequentemente vá ao mesmo mercado, e com o tempo perceba

que o caixa 5 é o mais rápido. Portanto, essa pessoa tentará diminuir o seu tempo de espera no supermercado indo consistentemente no caixa 5. Entretanto, em um Mercado Eficiente, a informação "O caixa 5 é o mais rápido" está disponível para todos, de modo que esta pessoa não será a única a perceber este fato. Deste modo, todas as pessoas irão para o caixa 5 eliminando assim qualquer chance de reduzir sistematicamente o tempo de espera no supermercado.

Durante muito tempo essa foi a visão dominante no mundo das finanças. Muitas das maiores ferramentas financeiras encontradas hoje em dia assumem que as mudanças nos preços são uma variável aleatória em movimento Browniano com uma distribuição normal e independência de caminho. Essa visão tem recebido críticas cada vez mais fortes, especialmente após a década de 90.

Um dos maiores críticos da HME foi Benoit Mandelbrot, um matemático Polonês radicado na França e em seguida nos Estados Unidos da América. Ele consegue mostrar em seu livro[5] como os modelos que usam passeio aleatório com movimento Browniano não representam suficientemente bem o movimento do mercado. Sua contribuição mais importante para este trabalho é:

- Mostrar a interdependência dos preços. Em um mercado em que os preços são independentes e puramente aleatórios, toda tentativa de uso de dados históricos na obtenção de lucros de maneira sistemática é vã;
- Mostrar a não-estacionariedade da mudança de preços, pois isto apoia a escolha de um algoritmo de aprendizagem por reforço treinado dinamicamente ao invés de uma ferramenta mais simples e estática.

2.3 Operações Básicas

Um bom ponto de partida para qualquer indivíduo que deseje ingressar no Mercado Financeiro é a correta compreensão do *Bid-Ask Spread* (diferença de cotação nas ofertas de Compra e Venda) [1]. Por meio deste tópico, é possível explorar com bastante riqueza quais operações estão disponíveis para um agente financeiro no mercado e como as suas ações se relacionam com as de outros agentes financeiros no chão da Bolsa de Valores.

Antes de iniciarmos, vale uma pequena ressalva: a bolsa de valores atrai diversos participantes. Dentre eles, temos os *traders*, os investidores, os administradores de fundos multi-mercado (*Hedge Fund Managers*), os ancos e outros. Essas figuras tem diferentes motivações e objetivos. Por simplicidade, todos eles serão referidos como agentes financeiros no decorrer deste trabalho.

A bolsa de valores é, antes de tudo, um mercado. Nesse mercado negociam-se ativos sendo estes representados por instrumentos financeiros. Um instrumento pode ser uma ação de uma determinada companhia, um contrato futuro, uma opção, dentre vários outros. No Mercado há agentes financeiros livremente comprando e vendendo ativos.

Considere um mercado hipotético em que existam apenas dois agentes financeiros, e que um deles deseje vender 100 ações. Ele possui em suas mãos duas opções:

- Observar o preço sob o qual os negócios estão sendo realizados e vender para quem quiser comprar. (Chamado *Market Order*, ou Ordem a Mercado)
- Ser um pouco mais exigente e pedir o preço pelo qual deseja vender. (Chamado *Limit Order*, ou Ordem Limitada)

De maneira análoga, suponha que o outro agente decidiu comprar 300 ações. Novamente existem duas opções:

- Observar o preço sob o qual os negócios estão sendo realizados e comprar pelo preço corrente. (Chamado igualmente de *Market Order*, ou Ordem a Mercado)
- Ser um pouco mais exigente e exigir um preço de compra. (Chamado igualmente de *Limit Order*, ou Ordem Limitada)

Ambas possuem vantagens e desvantagens. Comprar ou Vender com o *Market Order*, em geral, garante que sejam efetuadas todas as operações que se deseja de maneira instantânea. Entretanto, abre-se mão do controle do preço sob o qual essa compra ou venda será efetuada. Por outro lado, com o uso de *Limit Orders* é possível um melhor controle do preço sob o qual a operação será realizada, perdendo-se no processo a garantia de que haverá compradores ou vendedores dispostos a atender tais critérios. Isto porque o sistema da bolsa ranqueia as *Limit Orders* de venda, atendendo primeiro aquelas que possuírem o preço mais barato e ranqueia as *Limit Orders* de compra, atendendo primeiro aquelas que oferecem o maior preço.

Suponha que ambos os agentes financeiros se sentiram confiantes e optaram por *Limit Orders*. O primeiro agente financeiro não aceita vender por menos de R$ 21,30 (este preço é chamado de *Ask*), enquanto o segundo não aceita comprar por mais de R$ 21,00 (este preço é chamado de *Bid*). Como este mercado hipotético possui apenas esses dois agentes, e nenhum deles aceita as condições do outro, então nenhuma operação acontece. Neste cenário, diz-se haver um *Bid-Ask Spread* de R$ 0,30. Qualquer um deles que deseje que o negócio se concretize deverá cruzar essa barreira para que a operação aconteça.

2.3.1 Batalha entre Touros e Ursos

Existe uma analogia do Mercado, sendo este tratado como uma batalha entre Touros e Ursos[6]. Um Touro ataca por meio da chifrada, um golpe ascendente. Um Urso, por sua vez, ataca com a patada, um golpe descendente. Os compradores são comparados aos Touros (que, ao comprar, forçam o preço a subir) e os vendedores são comparados aos Ursos (que, ao vender, forçam o preço a descer). Assim, na literatura, é comum ler-se a respeito de Mercado de Touros (*Bull Market*) quando é um mercado em tendência de alta e Mercado de Ursos (*Bear Market*) quando se fala de um mercado em tendência de baixa.

Cada operação individual pode ser interpretada como um "golpe" no mercado. Em um Mercado de Touros, por exemplo, os compradores estão se sentindo confiantes com relação às compras que estão efetuando, enquanto os vendedores se encontram temerosos de realizar a venda.

Isso ocorre, por exemplo, quando o preço dos ativos está em ascensão. Um comprador, experimentando a confiança transmitida pela tendência, aceita pagar um pouco mais caro para possuir aquele ativo. De maneira análoga, um vendedor dessa mesma ação, ao notar a tendência de subida, fica temeroso de vender seu ativo e, com isso, deixar de lucrar. Assim, ele exige um valor um pouco mais elevado pela mesma ação. Dessa forma, os *Bid-Ask Spreads* serão cruzados na direção do vendedor, fazendo com que o preço sob o qual se realizam as operações seja elevado.

Em contrapartida, suponha um Mercado de Ursos, no qual a tendência é a queda nos preços dos ativos. Um vendedor, de posse de um ativo, desejará se livrar dele o mais rápido possível, aceitando para isso receber um pouco menos. Já um comprador pensará duas vezes antes de comprar aquele ativo em queda, exigindo preços mais baratos que compensem o risco que ele está assumindo. Note que de maneira análoga ao Mercado de Touros, os *Bid-Ask Spreads* serão cruzados na direção do comprador, fazendo com que os preços correntes sejam diminuídos.

Em mercados sem tendências formadas, cada vez que o *Bid-Ask Spread* é cruzado na direção da venda, diz-se que os Touros (associados portanto a figura dos compradores) estavam confiantes e derrotaram os Ursos (associados a figura dos vendedores). Analogamente, quando o *Bid-Ask Spread* é cruzado na direção de compra, diz-se que a força dos Ursos superou a dos Touros.

2.3.2 Entrando no mercado: *Short* vs *Long*

Existem duas maneiras de operar instrumentos financeiros (entrar no mercado) como um agente financeiro. A primeira maneira consiste em comprar um determinado ativo, sob a expectativa de que seu preço venha a subir. De fato, esta é a maneira mais intuitiva de se entrar no jogo do Mercado Financeiro. Note que, conforme explicado anteriormente, essa operação de compra está intimamente ligada à figura do Touro. A essa operação dá-se o nome de *Long*[6].

Por outro lado, caso o mercado apresente tendências de queda, existe outra operação disponível para auferir lucro do mercado. É possível tomar ativos emprestados de outros agentes financeiros e vendê-los. Após um determinado período de queda, deve-se recomprar estes ativos e devolvê-los aos seus donos originais pagando-se uma taxa de aluguel. Ao fim desta operação, o lucro consistirá na diferença entre o preço de venda e o de recompra (supostamente menor que o de venda), subtraindo-se as taxas. Ao processo de recompra, dá-se o nome Cobertura. Existe um controle legal sob esta operação, embora ela seja bastante utilizada por diversos agentes financeiros. Note que entrar no mercado com esperança de queda caracteriza bastante a figura do Urso descrita anteriormente. Essa operação recebe o nome de *Short*[6].

2.3.3 Saindo do mercado: *Stop Loss*, *Stop Gain* e *Trailing Stop*

Ao sair do mercado (fechar a posição), um agente financeiro pode se deparar com duas situações: realizando lucro ou assumindo prejuízo. Tipicamente, as estratégias imponhem limites máximos sobre as perdas para o caso em que o mercado não segue a tendência estimada. Uma das maneiras pelas quais uma estratégia pode limitar o seu risco de perdas é fazendo uso de *Stop Loss*.

Em sua forma mais simples, o *Stop Loss* consiste em um limite inferior de preços estabelecido no momento da compra. Caso este limite seja ultrapassado, efetua-se de maneira automática uma ordem de venda do ativo a mercado, protejendo-se assim de uma possível perda ainda maior.

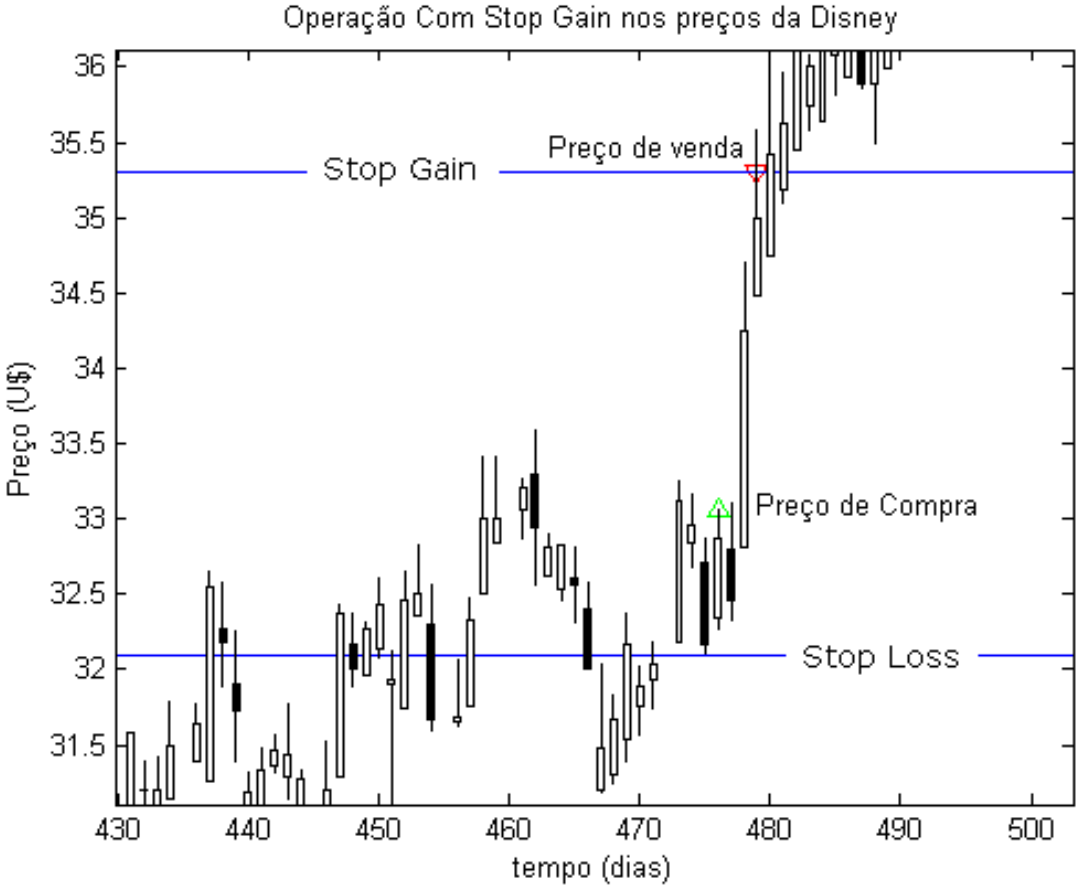

Figura 2.1: Stop Loss

Da mesma forma, existe o *Stop Gain*, que limita o ganho máximo. A princípio, não faz sentido cortar voluntariamente o próprio lucro. Entretanto, não existe um meio de se saber com exatidão o comportamento do mercado. São bastante frequentes as reversões de tendência, sendo que o *Stop Gain*, quando atingido, é capaz de capturar parte do lucro (ou mesmo sua totalidade) sem sofrer com esta reversão.

Assim como o *Stop Loss*, o *Stop Gain* em sua forma simples consiste na escolha de um limite. Quando os preços rompem esse limite uma ordem de venda é lançada no mercado a um preço pré-determinado (*Limit Order*) e, quando consumida, proporciona o lucro.

Tanto ao *Stop Gain*, quanto ao *Stop Loss*, está associado um *trade-off* a ser realizado na escolha do Limite. A escolha de valores elevados do limite no *Stop Gain* resulta em grandes possibilidades de lucro. Entretanto, isso aumenta proporcionalmente o risco de uma reversão de tendência que reduz as possibilidades de lucro. Por outro lado, a escolha de um limite baixo garante que, com uma probabilidade maior, a compra termine com um lucro. Entretanto, sacrifica-se boa parte do lucro possível ocasionado pelo abandono da tendência antes que esta se complete.

Existem também versões mais complexas e não estáticas dessas ferramentas. Um exemplo é o uso de *Stop Loss* com limite que se ajusta de acordo com o progresso do instrumento. Nisto consiste o *Trailing Stop*. Com o uso de *Trailing Stops* consegue-se uma proteção análoga ao caso do *Stop Loss*, sem com isso sacrificar as possibilidades de lucro da mesma maneira que o *Stop Gain*.

Apesar de seu apelo intuitivo, estas não são as únicas estratégias de saída disponíveis no mercado. Inclusive, existem agentes financeiros que advogam contra o uso destas formas de saída, para determinadas estratégias.

2.3.4 Custos Associados ao Trade

Em Teoria dos Jogos, todo jogo pode ser classificado em duas grandes categorias: Jogos de Soma Zero e Jogos de Soma Não-Nula. A distinção entre esses jogos está no fluxo de suas recompensas.

Em Jogos de Soma Zero, as perdas sofridas pelos jogadores perdedores fluem de maneira perfeita na forma de ganhos para os vencedores. Em outras palavras, caso se some o total dos ganhos dos vencedores e se subtraia do total das perdas dos perdedores o total deve ser Zero. Pode-se interpretar que os perdedores "pagam" os ganhos dos vencedores.

Já em Jogos de Soma Não-Nula, existe uma diferença entre o total de perdas dos perdedores e ganhos dos vencedores. Isto pode se dar de duas maneiras: Ou os ganhos dos vencedores se sobressaem às perdas dos perdedores, ou as perdas se sobressaem aos ganhos. No primeiro caso, pode-se interpretar que o próprio jogo "paga" a diferença nas recompensas, enquanto que no segundo o jogo "recebe" a diferença.

A bolsa de valores é um exemplo de um Jogo de Soma Não-Nula. Note que mesmo que o lucro de um agente financeiro esteja associado a perdas sofridas por outros agentes, existem diversos fatores a serem considerados. Isto se dá por uma série de motivos explorados a seguir.

Suponha, por exemplo que um agente financeiro decida que é hora de comprar um determinado ativo. Deve-se sempre lembrar que o ato de compra pressupõe a existência de outro agente financeiro que está disposto a vender. Caso a compra siga no sentido desejado, o agente financeiro efetuou a compra por um preço inferior ao de venda. A venda também pressupõe a existência de um comprador. Note que o agente financeiro conquistou uma recompensa do mercado. Tal recompensa foi paga tanto pelo agente financeiro que lhe vendeu o ativo (e assim deixou de ganhar aquela recompensa) quanto por aquele que comprou (e efetivamente pagou o valor da recompensa).

Assim, a recompensa de uma operação é paga pelos agentes financeiros que estão do outro lado das mesmas operações. Como a recompensa é paga pelos outros jogadores, pode-se em uma primeira observação interpretar a bolsa como um Jogo de Soma Zero. Entretanto, o próprio jogo (a bolsa) cobra de seus jogadores uma taxa por cada operação realizada. Assim, o prejuízo sofrido pelos agentes financeiros é ainda mais acentuado, pois além de pagar a recompensa do vencedor, eles tem de pagar as taxas para participar do jogo, as chamadas Taxas de Corretagem. De maneira análoga, uma parte da recompensa do agente financeiro vencedor é utilizada para pagar estas taxas.

Existe ainda outro custo associado ao mercado. Uma vez que se emita uma ordem de compra

ou de venda no mercado, existe uma diferença entre o preço esperado e o preço real pelo qual se conseguirá efetuar a compra ou a venda. Essa diferença ocorre por diversos fatores. Dependendo da velocidade de transmissão da informação, o *Home Broker*[1] pode mostrar um valor que não corresponde mais ao valor real sob o qual se negociem as ações naquele instante. Ou, ainda que o valor esteja atualizado, uma *Market Order* muito grande pode encontrar uma demanda muito pequena para o preço no mesmo instante, realizando então só uma parte da ordem no preço original e o resto de acordo com o preço demandado pelo mercado. Tais diferenças são chamadas *Slippages*.

Deste modo, o objetivo no mercado de ações deve ser ajustado para contar com tais custos. Deve-se conseguir efetuar operações boas o bastante de modo que a recompensa conquistada seja suficiente para extrair lucro, mesmo considerando as Taxas de Corretagem e o eventual *Slippage*.

Existem outros fenômenos presentes na bolsa, que a distanciam de um Jogo de Soma Zero. Um agente financeiro de posse de uma ação de determinada companhia é, para todos os efeitos, um sócio desta. Deste modo, enquanto ele detiver o papel desta ação, este agente financeiro tem direito à parte dos lucros produzidos pela companhia. Estes lucros são ocasionados por Dividendos ou mesmo por Juros sobre o capital próprio, e podem ser interpretados, na visão da Teoria dos Jogos, como um prêmio pago pelo jogo ao possuidor destas ações. Outros fenômenos que devem ser levados em consideração são a Inflação e o Imposto de Renda, sendo este responsável por consumir boa parte dos lucros obtidos.

2.4 Análise Fundamentalista vs Análise Técnica

Uma vez conscientes do que é ou não possível dentro do mercado em termos de compras e vendas, pode-se começar a pensar teorias a respeito do funcionamento do mercado. De maneira geral, aqueles que se propõem analisar o mercado partem de dois pontos de vista que, embora pareçam opostos, muitas vezes se completam. Essas duas escolas de pensamento são chamadas de Análise Fundamentalista e Análise Técnica. Elas partem de hipóteses distintas e utilizam indicadores totalmente diferentes.

2.4.1 Análise Fundamentalista

A Análise Fundamentalista analisa os fundamentos de uma empresa, ou de um determinado contexto no mercado como Renda Fixa ou *commodities*. O objeto de estudo primordial nesse tipo de análise é tipicamente a companhia por trás do papel. Informações tais como o lucro anual dessa companhia, o nível de dívidas, a taxa de crescimento anual, histórico dos dividendos, entre outras informações são extremamente relevantes para uma análise bem apurada. A hipótese primordial da Análise Fundamentalista, que a contrapõe à Análise Técnica, é que **nem toda a informação disponível a respeito da companhia está contida nos preços**. Em outras palavras, analisando os fundamentos por trás de uma empresa, é possível descobrir uma companhia

[1] *Broker* é o indivíduo ou empresa que, dentre outras atividades, executa ordens de compra e venda por um terceiro.[1] *Home Broker* é o software usado para substituir e automatizar o *Broker*.

com papéis na bolsa baratíssimos, embora eles sejam extremamente valiosos no futuro, ou mesmo no presente. Deste modo, além do preço atual sob o qual o papel é negociado, outras informações da companhia também são relevantes, uma vez que o preço irá "incorporar" estas informações futuramente, seja aumentando ou diminuindo.

Um dos principais conceitos dentro da Análise Fundamentalista é o de Valor Intrínseco. Esse valor representaria o valor real de uma companhia. Ele costuma ser um valor distinto daquele que está sendo negociado nas bolsas, por ser um valor mais estável. De posse de uma estimativa adequada de Valor Intrínseco, teoricamente, é possível efetuar compras de ações cujo Valor Real seja inferior ao valor intrínseco, na esperança de que o Valor Real se equipare ao Intrínseco com o passar do tempo. De maneira análoga, é possível abandonar ou mesmo vender em *Short* papeis cujo Valor Real seja superior ao Valor Intrínseco, valendo-se da hipótese de que o valor de mercado eventualmente tenda ao Valor Intrínseco.

Existem várias críticas a respeito da Análise Fundamentalista. Uma delas advém da estimativa do Valor Intrínseco. Sendo uma estimativa baseada em fatores qualitativos, tais como nível de organização e reconhecimento da empresa, tal valor não possui uma maneira exata de ser calculado, dependendo bastante do analista. Isto dificulta que a Análise Fundamentalista seja submetida a um teste estatístico que verifique a sua validade. Uma segunda crítica vem da Hipótese do Mercado Eficiente. Segundo ela, toda a informação disponível sobre a empresa já está refletida no histórico de preços. Deste modo, uma situação similar àquela da fila de supermercado se forma. Se os investidores começassem a comprar ações cujos Valores Intrínsecos são inferiores aos valores reais, então o Valor Real naturalmente subiria até alcançar o Valor Intrínseco.

Uma outra crítica advém da estacionariedade presente nas avaliações. Por exemplo, suponha um investidor avaliando duas companhias na década de 80, uma produtora de máquinas de escrever e outra produtora de computadores pessoais. Utilizando os dados disponíveis naquela época, o investimento na companhia produtora de máquinas de escrever seria extremamente razoável, uma vez que esta provavelmente possui patrimônio maior e possui um maior valor agregado a si quando comparada com as empresas que produzem computadores, que neste momento ainda estão em seu estágio inicial. Entretanto, 30 anos mais tarde, vemos que a máquina de escrever se tornou peça de museu enquanto o computador se tornou quase uma necessidade básica.

A Análise Fundamentalista é excelente para justificar tendências passadas, mas possui diversas complicações ao ser aplicada para projeções futuras. Por esse motivo, embora ela seja uma fonte valiosa de informações a respeito de um dado ativo, deve sempre ser utilizada de maneira cautelosa.

2.4.2 Análise Técnica

A outra escola de pensamento do Mercado Financeiro é a Análise Técnica. Os principais objetos de estudo deste tipo de análise são os gráficos e os dados históricos do papel em questão. A principal hipótese por trás de toda a Análise Técnica é a de que **o histórico de preços contém toda informação pertinente sobre a companhia**[1]. Deste modo, um agente financeiro utilizando-se de Análise Técnica opta por concentrar seus esforços na correta compreensão dos gráficos de

preços e dos dados históricos disponíveis, tais como volume monetário movimentado e quantidade de negócios fechados nos períodos anteriores. Torna-se irrelevante para o analista técnico qual o papel que se está negociando, importando mais a tendência e a dinâmica dos preços, conjuntamente com as estatísticas de participação dos agentes no mercado.

A Análise Técnica possui algumas hipóteses centrais. Uma destas hipóteses é a da existência de tendências. Parte-se do pressuposto de que um crescimento no preço em geral é seguido de outros crescimentos e que o decrescimento, em geral, é seguido por outros decrescimentos. Deste modo, existem diversas técnicas e indicadores desenvolvidos para melhor compreender e aproveitar uma tendência. Ao longo da história, diversos indicadores foram inventados com o objetivo de retratar, com riqueza de detalhes, o que ocorre nos gráficos.

Um outro princípio bastante utilizada é a de Retorno à Média. Esta princípio se assemelha bastante àquela da Análise Fundamentalista, que associa a ação um Valor Real. Neste caso, associa-se a ação um Valor Médio. Acredita-se que as variações de preço, por mais que se distanciem deste Valor Médio, eventualmente tenderão a retornar a ele. Toda uma categoria de indicadores (os chamados osciladores) faz uso desta hipótese fornecendo medidas para este distanciamento da média como indicativos de compra ou de venda.

Até pouco tempo, a Análise Técnica não era vista com bons olhos pelos altos investidores de Wall Street. Isto se deu por diversos fatores. Entre eles está o fato de que, embora seja possível, não existem estudos estatísticos amplamente divulgados que justifiquem o uso de tais técnicas. Outra crítica pertinente advém da análise do formato dos gráficos, bastante utilizada inicialmente. A maneira como ela era apresentada se assemelhava bastante a uma pseudociência, o que cristalizou a visão de que a Análise Técnica era uma "Astrologia Financeira". Outra crítica importante vem da Hipótese do Mercado Eficiente. Se os agentes financeiros operam seguindo uma vantagem disponível nos gráficos, naturalmente o mercado incorporará esta informação, fazendo com que a vantagem inicial percebida seja destruída.

2.5 Principais Indicadores

Este trabalho está calcado nos princípios da Análise Técnica. Logo, é conveniente analisar seus indicadores, visto que foram projetados para extrair mais informações a respeito dos mercados do que os simples valores dos ativos. Essa seção é dedicada a apresentar os principais indicadores utilizados por agentes financeiros que se baseiam em Análise Técnica ao efetuar suas operações.

2.5.1 Médias Móveis

Uma intuição clássica do mercado dita que os agentes financeiros no mercado encontram-se frequentemente sob um determinado grau de otimismo ou pessimismo, o que os leva a tomar determinadas decisões de maneira aparentemente irracional. Tais decisões acabam por induzir sobre o preço analisado um elevado grau de ruído sobre o que seria o "preço real" de um dado instrumento. Uma forma de se contornar este problema se dá por meio do uso de uma média

amostral aritmética, calculada com base no preço de um dado número de dias passados.

Uma média móvel aritmética (MMA) com N períodos pode ser calculada por:

$$MMA = \frac{1}{N}\sum_{j=1}^{N}(P_j) \tag{2.1}$$

Uma vez calculado o primeiro valor de uma média móvel, ele pode ser atualizado de maneira computacionalmente mais eficiente por meio da fórmula recursiva:

$$MMA_t = MMA_{t-1} + \frac{P_t}{N} - \frac{P_{t-N}}{N} \tag{2.2}$$

Onde P_t é o preço atual e P_{t-N} é o preço de N períodos atrás. Sob a perspectiva da Teoria de Sinais Discretos, essa equação pode ser interpretada como um filtro passa-baixas de Resposta Finita ao Impulso (FIR), que elimina as flutuações bruscas do mercado.

Deste modo, uma interpretação razoável para a média móvel seria a seguinte: caso o valor de um preço se encontre acima da média móvel, é esperado que esta diferença eventualmente desapareça, seja pela elevação da média móvel, seja pela diminuição do preço. De maneira análoga, caso os preços se encontrem abaixo da média móvel, eventualmente eles deverão subir para acompanhá-la ou ela deverá descer para encontrá-los. Uma forma simples, então, de traçar tendências é observar a relação entre os preços e suas médias móveis. Se um preço sobe acima de sua média, significa que o instrumento está se valorizando em relação ao consenso de longo prazo, ou seja, existe uma possível tendência de alta. No caso em que os preços cruzem a média para baixo, mostraria uma possível tendência de baixa.

Existe, no entanto, um problema grave associado a esse tipo de cálculo, quando efetuado desta forma. Suponha que ocorra uma mudança brusca em um dado dia excepcionalmente volátil, em que, por exemplo, as compras foram efetuadas sob preços anormalmente elevados. Suponha que este dia seja seguido por dias de calmaria, em que o preço não se eleve nem caia de maneira significativa. A média móvel aritmética, no entanto, sofrerá uma queda brusca no momento em que tal dia deixar de ser levado em conta em seu cálculo[6]. Note que os dias passados possuem uma capacidade de interferência significativa no cálculo do valor atual, conforme a equação 2.2.

Para se superar esse problema geralmente se usa uma Média Móvel Exponencial (MME). Ela é uma média ponderada que, diferentemente da MMA, reage apenas a novos preços. Além disso, em seu cálculo, o valor do preço atual possui um peso exponencialmente maior que o preço do dia anterior. Seu cálculo se dá por:

$$MME_t = P_t \cdot K + MME_{n-1} \cdot (1 - K) \tag{2.3}$$

Onde $K = \frac{2}{N+1}$, N é o número de períodos da média móvel, escolhido pelo agente financeiro, P_t é o preço atual e MME_t e MME_{t-1} são as médias móveis exponenciais do período em questão e do período anterior, respectivamente.

2.5.2 OBV - On Balance Volume

O objetivo deste parâmetro é indicar o fluxo de agentes financeiros comprando ou vendendo um determinado ativo, além de estimar a quantidade de pessoas atualmente em posse de um ativo. Foi desenvolvido por Joseph Granville em 1963 e exposto em seu livro "Granville's New Key to Stock Market Profits"[7]. Segundo Granville, o volume de negócios é a principal força que move os preços no mercado.

Figura 2.2: Variação brusca nos preços acompanhada de Volume elevado[1]

Se um ativo possui uma procura grande, a tendência é um aumento em seu preço. De maneira análoga, uma diminuição em sua demanda será seguida de uma diminuição em seu preço. O OBV é uma maneira de quantificar a demanda por um ativo no tempo. Ele é calculado da seguinte forma: Em cada período, verifica-se se houve uma elevação ou uma queda nos preços. Para cada elevação no preço, o volume de negociações daquele período é somado ao valor anterior de OBV. A cada queda, o volume daquele intervalo é subtraído do OBV. Caso o preço se mantenha o mesmo, não se altera o OBV. Em outras palavras:

$$OBV_{n+1} = \begin{cases} OBV_n + Volume & \text{se } Open < Close \\ OBV_n & \text{se } Open = Close \\ OBV_n - Volume & \text{se } Open > Close \end{cases} \tag{2.4}$$

O OBV é um parâmetro que captura a intensidade de uma tendência. Em outras palavras, se há um aumento de preços, mas o volume de negociações não justifica tal aumento, significa que não há uma tendência de negociações neste sentido. De maneira análoga, se há uma diminuição brusca de preços, mas o OBV não acompanha este decréscimo, significa que não está se formando uma tendência de baixa, e sim poucos agentes financeiros negociando sob aquele regime de preços.

Deste modo, a principal utilidade do OBV é a confirmação de tendências. Uma tendência só se confirma caso o OBV a acompanhe.

2.5.3 A\D Line - Accumulation\Distribution Line

Um dos problemas do OBV é que ele atribui todo o volume de negociações de um período a apenas uma tendência, seja ela crescente ou decrescente. Entretanto, sabe-se que nem todo o

volume de negociações deve ser atribuído à mesma tendência. Assim, Larry Williams desenvolveu o indicador A\D Line[8], que consiste basicamente em um refinamento do OBV. Abaixo segue como é efetuado seu cálculo:

- Calcula-se inicialmente o Multiplicador de Fluxo Monetário (MFM). Este número está no intervalo $[-1, 1]$ e indica a direção e a intensidade da tendência. Quanto maior o valor absoluto de MFM, mais intenso é o fluxo. Valores positivos de MFM indicam tendência de aumento de preços enquanto valores negativos indicam tendência de redução. Ele é dado por:

$$MFM = \frac{(Close - Low) - (High - Close)}{High - Low} \tag{2.5}$$

- Uma vez de posse do Multiplicador de Fluxo Monetário, procede-se calculando o Fluxo de Volume (FV) do período correspondente. Ele é obtido, simplesmente, pelo produto entre o Multiplicador de Fluxo Monetário pelo Volume do dia.

$$FV = MFM \cdot Volume \tag{2.6}$$

- Uma vez de posse do Fluxo de Volume do dia, procede-se finalmente ao cálculo do A\D Line. Sendo um parâmetro cumulativo, seu cálculo é recursivo e incremental, tal qual o OBV. Ele é dado por:

$$A\backslash D_{n+1} = A\backslash D_n + FV \tag{2.7}$$

2.5.4 Sistema Direcional

Em 1978, um engenheiro mecânico chamado J. Welles Wilder Jr. lançou um livro chamado "New Concepts in Technical Trading Systems"[9]. Nele figuravam diversos indicadores inovadores para a época, que inclusive são bastante utilizados ainda hoje. Vários destes indicadores foram projetados para serem utilizados em conjunto, na forma de *Trading Systems*. Em particular, um destes conjuntos desenvolvidos por Wilder é o chamado Sistema Direcional.

Para corretamente compreender o Sistema Direcional, é necessário compreender cada um de seus indicadores, e como este se relaciona com os demais. Para isso, será dedicada uma breve seção para cada um deles.

2.5.4.1 True Range

O True Range é um indicador da volatilidade de uma ação. Quando foi projetado por Wilder, este tinha em mente o uso de preços diários de *commodities*. Usualmente a volatilidade era calculada pela medida da amplitude entre o maior e o menor preço sob os quais se realizaram operações. Entretanto Wilder, ao criar o True Range, optou por uma análise diferente sobre a volatilidade. O cálculo do True Range se dá da seguinte maneira:

$$TR_n = max \begin{pmatrix} |High_n - Low_n| \\ |High_n - Close_{n-1}| \\ |Low_n - Close_{n-1}| \end{pmatrix} \tag{2.8}$$

Existem 3 maneiras de se efetuar o cálculo. A primeira consiste em observar o *gap* existente entre o máximo e o mínimo do dia. A segunda maneira consiste em comparar o *gap* entre o máximo do dia e o fechamento do dia anterior. A terceira consiste em tomar-se o mínimo do dia e comparar-se com o fechamento do dia anterior. O valor do True Range será, dentre estes 3, o maior. Observe que sempre é tomado o valor absoluto, uma vez que este não se trata de um indicador de direções, e sim de volatilidade. Assim, o True Range sempre resulta em um valor superior a zero.

Em geral, este valor não é tomado de maneira simples. Os indicadores desenvolvidos por Wilder costumam utilizar com frequência suavizações por meio de Médias Móveis. Em particular, o True Range foi projetado para ser utilizado com uma suavização por uma Média Móvel que leva em conta 14 períodos anteriores. Após sofrer a suavização, o indicador passa a ser conhecido como Average True Range. A média móvel utilizada é a exponencial e é dada pela seguinte fórmula de recorrência:

$$ATR_n = \frac{13 \cdot ATR_{n-1} + TR_n}{14} \tag{2.9}$$

Sendo que o primeiro ATR é calculado usando a média aritmética:

$$ATR = \frac{1}{14} \sum_{i=1}^{14} TR_i \tag{2.10}$$

2.5.4.2 Movimentos Direcionais

Note que o True Range é um indicador projetado para calcular volatilidade. Entretanto no sistema imaginado por Wilder outros dois indicadores também tem papel central: os assim chamados Movimento Direcional Positivo (DM_+) e Movimento Direcional Negativo (DM_-). Estes são responsáveis por calcular a intensidade das tendências, tanto de crescimento (DM_+) quanto de decrescimento (DM_-). Voltando para a analogia da batalha entre Touros e Ursos, o DM_+ seria um valor representativo do poder dos Touros em um dado instante e o DM_- um valor que representa o poder dos Ursos. Sendo uma batalha, apenas um deles pode obter a vitória em um dado dia. Portanto, em cada dia apenas um destes, o maior, poderá ser positivo, sendo o outro considerado zero.

$$DM_+ = \begin{cases} High_n - High_{n-1} & \text{se } High_n > High_{n-1} \\ 0 & \text{Caso Contrário} \end{cases} \tag{2.11}$$

$$DM_- = \begin{cases} Low_{n-1} - Low_n & \text{se } Low_n < Low_{n-1} \\ 0 & \text{Caso Contrário} \end{cases} \tag{2.12}$$

Note que podem existir dias que não possuem nenhum destes movimentos direcionais (são chamados "*Inside Days*", ou "Dias Internos"). Nestes casos, tanto o movimento direcional positivo quanto o negativo são nulos. Caso haja a situação oposta, existindo movimentos direcionais nos dois sentidos ("*Outside Days*", ou "Dias Externos"), diz-se existir apenas o maior deles, sendo o outro nulo. Veja na figura abaixo exemplos das 4 situações:

Figura 2.3: Possíveis Movimentos Direcionais[1]

Assim como o True Range, estes também costumam ser suavizados com o uso de Médias Móveis Exponenciais de 14 períodos. (ADM_+ e ADM_-)

2.5.4.3 Índice Direcional

Observe que o Sistema Direcional de Wilder já conta com um mecanismo que mede a volatilidade de uma determinada ação e outros mecanismos de medida de direção de tendência. Wilder, no entanto, achou necessário incluir ainda um último indicador, sendo este responsável por medir a intensidade da tendência. Este indicador é muitas vezes utilizado de maneira independente em outros *Trading Systems* como meio de detecção de formação e medida de intensidade de tendências.

Para seu cálculo, existem duas etapas. A primeira consiste em calcular os indicadores previamente explicitados, o Average True Range (ATR) e ambos os Movimentos Direcionais suavizados, (ADM_+ e ADM_-). Em seguida, são calculados os Índices Direcionais Positivos e Negativos. Tais

índices consistem em um ajuste dos Movimentos Direcionais com relação a volatilidade. Portanto, sua forma de cálculo é dada por:

$$DI_+ = \frac{ADM_+}{ATR} \tag{2.13}$$

$$DI_- = \frac{ADM_-}{ATR} \tag{2.14}$$

Novamente, tal qual todos os anteriores, Wilder sugere uma suavização destes índices com a mesma Média Móvel Exponencial de 14 períodos citada anteriormente, obtendo ADI_+ e ADI_-. De posse desses valores, o Índice Direcional (DI) é calculado utilizando a seguinte fórmula:

$$DI = \frac{ADI_+ - ADI_-}{ADI_+ + ADI_-} \tag{2.15}$$

O sistema direcional foi pensado para ser utilizado como um todo. Entretanto, nada impede que um agente financeiro faça o uso deste sistema conjuntamente com outros, ou mesmo de partes deste sistema. O próprio Wilder inventou diversos outros sistemas, também descritos em seu livro[9].

2.5.5 Bandas de Bollinger

Até o momento, todos os indicadores analisados fazem parte de uma categoria dos chamados *Trend Followers* ou seguidores de tendências. Estes se aproveitam da existência de tendências e exploram ao máximo este conceito. Existe, no entanto, uma segunda categoria de indicadores que se baseia em um princípio completamente diferente. Eles são chamados osciladores. Eles se aproveitam do princípio do Retorno à Média, explicado anteriormente. Deste modo, eles estimam o quão distante da média uma determinada ação se encontra aproveitando para efetuar a compra em um momento de maior distância, que sob a hipótese de Retorno à Média marcará o ponto onde haverá uma reversão de tendência.

As bandas de Bollinger é um dos osciladores mais populares. Eles utilizam a hipótese de que as oscilações de preço costumam seguir uma distribuição gaussiana, de modo que oscilações muito bruscas devem ser extremamente raras. Deste modo, em cada ponto, calculam-se médias móveis, que teoricamente representam o ponto de maior probabilidade para o próximo movimento do mercado, conjuntamente com o desvio padrão móvel, que delimita toda uma zona na qual existe elevada probabilidade em que o preço seguinte se encontre.

Em alguns dias, no entanto, os preços rompem as bandas de Bollinger indicando que aquele dia sofreu um movimento anômalo. Neste ponto, os investidores tendem a utilizar a hipótese de Regressão à Média. Supõe-se que existe uma certa inércia associada ao preço de um dado ativo. Deste modo, se ele se eleva de maneira brusca em um determinado instante, isto se deveu a peculiaridades do mercado naquele instante de tempo, e não a um crescimento real de valor. Assim, espera-se que quando esse fenômeno terminar, o preço retorne a um valor mais razoável, aqui estimado pela média.

Existem 3 curvas associadas às bandas de Bollinger: a média móvel, a banda superior e a banda inferior. Para o correto cálculo das bandas sobre K períodos, procede-se da seguinte maneira:

$$MA_n = \sum_{j=n-K}^{n} P_j \tag{2.16}$$

$$\sigma = \sqrt{\sum_{j=n-K}^{n} \frac{(MA_j - P_j)^2}{K-1}} \tag{2.17}$$

$$Bol_n^+ = MA_n + \sigma \tag{2.18}$$

$$Bol_n^- = MA_n - \sigma \tag{2.19}$$

Veja abaixo ilustradas as bandas de Bollinger:

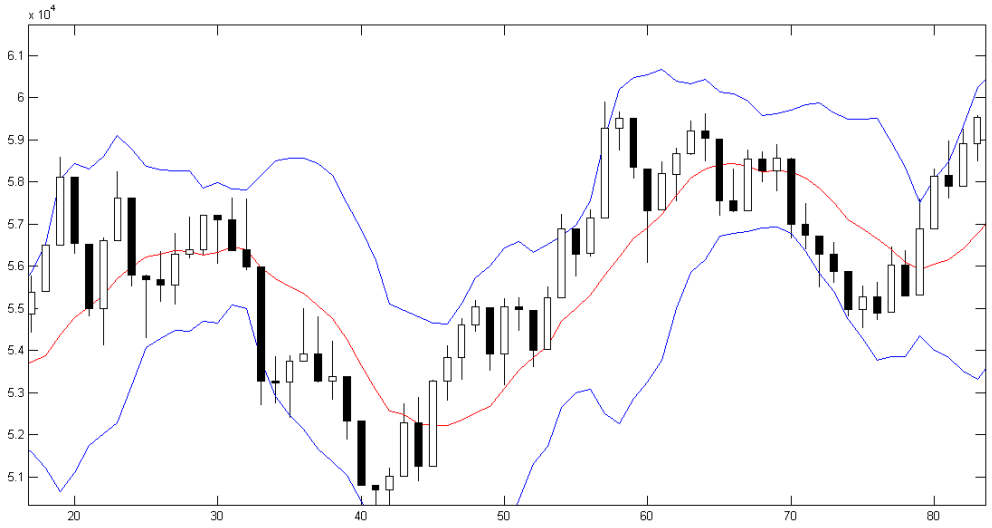

Figura 2.4: Bandas de Bollinger

2.5.6 MACD - Moving Average Convergence/Divergence

A Média Móvel Convergência/Divergência, o MACD, foi criada por Gerald Appel[2] e é usada com o intuito de identificar direção, momento, força e duração de uma tendência de preços de um instrumento. O sistema é composto 3 sinais, geralmente usando o preço de fechamento para seu cálculo.

- A linha MACD, ou linha rápida, é calculada através da diferença entre 2 médias móveis exponenciais, uma de 12 períodos menos uma de 26.
- A linha chamada de sinal, gatilho ou linha lenta, é calculada fazendo a média móvel de 9 períodos da linha MACD.

Os valores 12, 26 e 9, se tornaram com o tempo valores padrão e são usados desta forma na maioria dos softwares. Apesar de ser possível muitas vezes customizar isso, faz pouca diferença, segundo Alexander Elder, a não ser que se dirtorça enormemente um dos valores sem mexer nos outros[6]. Desta maneira, assim que a linha MACD cruza a linha de sinal para cima, é um gatilho para compra. Caso seja para baixo, é um gatilho para a venda. Segundo Elder, a linha rápida reflete um consenso de valor de curto período, enquanto a linha lenta representa um consenso a longo prazo.

Outra forma muito utilizada é o Histograma MACD. Seu cálculo é simples: basta subtrair da linha rápida do MACD a linha lenta. O Histograma MACD mede a diferença entre o consenso de curto e longo prazo e é plotado num gráfico de barras, como um histograma. Com o Histograma MACD é muito mais simples ver as diferenças de distância entre as duas curvas e assim se pode ver pela inclinação nas barras do histograma a crescente força dos Ursos ou Touros.

Figura 2.5: Análise da construção do MACD nos preços da OGXP3

Na figura 2.5, pode ser observado como funciona este indicador. É o histórico diário de preços da OGXP3 visualizado no programa ProfitChart©. A linha roxa é a MME de 26 períodos, a amarela é a de 12 períodos. Nota-se que, como esperado, elas se cruzam na linha no mesmo instante em que a linha MACD está no zero, representada no gráfico de baixo de cor laranja. Nota-se também que as barras do histograma seguem a tendência dos preços, ou seja, crescem em tendências de alta e decrescem em tendências de baixa.

Também pode ser explicitado um evento mais raro e muito importante de usabilidade do Histograma MACD que é a divergência. No final de agosto, na figura 2.5, pode se notar que a série temporal chegou a um mínimo local e no final de setembro chegou em outro mais fundo. O histograma, porém, cria um mínimo mais raso em setembro. Essa divergência entre o indicador e o histograma indica que os preços estão próximos a uma subida, como de fato aconteceu. Por

volta da metade de outubro, as ações da OGX sobem por volta de 100% no espaço de dois dias.

2.5.7 HiLo Activator

Criado por Krausz[10], o HiLo Activator usa a média aritmética dos preços máximos e mínimos de um período de tempo. Geralmente é representado num formato escada ao invés do formato suavizado normal de uma média móvel simples. O *Sell Stop* é o valor da média simples dos preços mínimos de um determinado período de tempo. Ele, no formato de escada verde na figura 2.6, é plotado abaixo dos *candles*, sinalizando tendência de alta, até que o instrumento fecha seu período com seu preço abaixo ao *Sell Stop*. A partir deste momento inicia-se a plotagem do *Buy Stop* acima dos preços em vermelho, que é a média simples dos preços máximos dos mesmo número de períodos passados. Analogamente, quando o instrumento, nesta tendência de baixa, fechar o período com seu preço acima do *Buy Stop*, este desaparece e dá lugar ao *Sell Stop* novamente.

Ele é um indicador muito interessante pois é excelente em seguir tendências e além dos pontos de compra e venda também fornece valores de *Stop Loss* na forma de um *Trailling Stop*.

Figura 2.6: HiLo

2.6 Sharpe Ratio

Inicialmente nomeado como *Reward-to-variability ratio* pelo seu criador, William Sharpe [11], é uma maneira muito utilizada para examinar a performance de um investimento em relação ao seu risco. Essa taxa mede a recompensa por unidade de desvio padrão numa estratégia de investimento.

O Sharpe Ratio S é definido por:

$$S = \frac{E[R_a - R_f]}{\sigma} = \frac{E[R_a - R_f]}{\sqrt{var[R_a - R_f]}} \tag{2.20}$$

Onde $E[R_a - R_f]$ é a esperança do retorno R_a sobre um retorno de comparação R_f, e $var[R_a - R_f]$ é sua variância. Originalmente, a comparação era feita com um retorno livre de riscos, sendo assim $\sqrt{var[R_a - R_f]} = \sqrt{var[R_a]} = \sigma$.

Existe, no entanto, algumas críticas que cercam o uso do Sharpe Ratio. A principal delas é o

Sharpe Ratio considerar que os retornos têm uma distribuição normal, o que muitas vezes não é verdade. Quando essa distribuição não é normal, o desvio padrão não tem a mesma efetividade e essa medida pode se tornar perigosa[1].

Capítulo 3

Revisão Teórica - Sistemas Inteligentes

"Knowledge, then, is a system of transformations that become progressively adequate."
Jean Piaget

3.1 Contextualização

Quando se levanta a questão da natureza do aprendizado, geralmente chegam a mente duas imagens: uma criança tendo aulas na escola e um bebê aprendendo a andar. O primeiro caso ilustra muito bem como funciona o aprendizado supervisionado: o professor ensina sua matéria e o aluno leva exercícios para casa. Após treinar repetidamente, no final do período, testa-se o que foi aprendido em uma prova final. Caso o aluno obtenha uma nota superior a um determinado valor, é considerado que ele aprendeu suficientemente bem o que lhe foi designado.

No caso do bebê tentando andar, o que acontece é uma infinidade de sequências de tentativa e erro até que ele consiga engatinhar. Ainda assim, quando engatinhar não é mais suficiente, ele inicia uma nova sequência de tentativas até conseguir se manter em pé e finalmente andar. Durante o processo, podem ocorrer eventuais "acidentes", que acabam por se tornar instrutivos. Neste caso não há uma comunicação, não existe professor. Por mais que se tente mostrar para o bebê como se anda, ele não consegue transformar essa informação em coordenação motora. É um caso típico de aprendizagem não supervisionada: as únicas coisas que existem são a interação do agente com o ambiente e os retornos que o ambiente dá ao agente. Esses retornos podem ser tanto em forma de custos, como a dor de quando o bebê cai e se machuca, quanto em forma de recompensas, como a satisfação do bebê ao conseguir alcançar o que queria.

3.2 Aprendizagem por Reforço

Esta ideia de um sistema que quer algo e adapta seu comportamento para maximizar um sinal especial de seu ambiente é chamada de Aprendizagem por Reforço. Segundo Sutton e Barto[12], a definição de Aprendizagem por Reforço não vem de alguma característica do algoritmo, mas com a caracterização de um problema de aprendizado. Qualquer algoritmo de aprendizado que sirva para resolver esse tipo de situação, na qual o agente precisa desbravar, sem guia, um ambiente

que lhe retorna uma recompensa (não necessariamente imediata) por suas ações é considerado um algoritmo de Aprendizagem por Reforço.

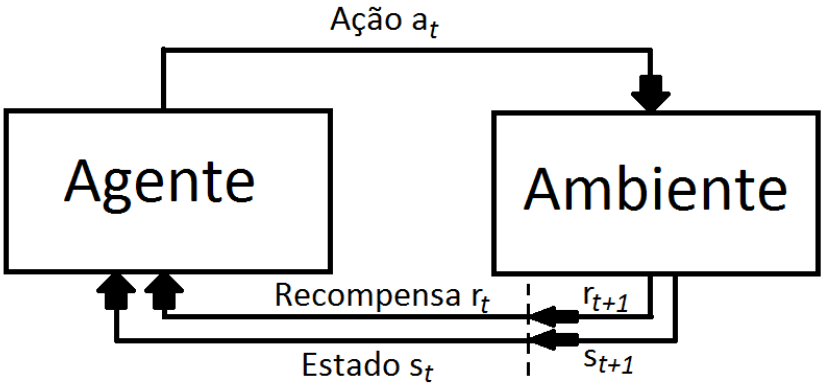

Figura 3.1: Modelo de um algoritmo de Aprendizagem por Reforço

A ideia básica é a representação de um problema real de um AGENTE que aprende interagindo com o AMBIENTE para atingir um objetivo. Naturalmente, o AGENTE precisa ser capaz de "sentir" o ESTADO do ambiente e de realizar AÇÕES que afetam o ESTADO. Ele também precisa ter pelo menos um objetivo em relação ao estado do ambiente. A intenção desta formulação é incluir apenas estes três aspectos – SENSAÇÃO, AÇÃO e OBJETIVO –, da maneira mais simples possível.[12]

Um exemplo simples de um problema de Aprendizado por Reforço é o jogo de xadrez. O AGENTE pode ver a posição das peças, ou seja, pode saber o ESTADO. Ele realiza ações no ambiente movendo as peças e faz isso com o intuito de ganhar o jogo. Note que, como seu objetivo é vencer sempre, o AGENTE em processo de aprendizado não sabe se suas jogadas foram boas ou ruins até acabar o jogo, quando ele recebe sua RECOMPENSA. Neste caso, cada jogo de xadrez constitui um EPISÓDIO. EPISÓDIOS são qualquer forma de interação repetida. Cada EPISÓDIO tem um ESTADO TERMINAL, que marca seu fim. Um novo EPISÓDIO começa no ESTADO INICIAL ou em uma distribuição padrão de Estados Iniciais. Uma interação AGENTE-AMBIENTE que não se separa em episódios pode ser tratada como um episódio infinito.

Resumindo, um algoritmo de Aprendizagem por Reforço trata seu problema com um AGENTE observando os ESTADOS do AMBIENTE e realizando AÇÕES sobre ele, de acordo com o aprendido em experiências passadas, com o objetivo de maximizar as RECOMPENSAS retornadas, como na figura 3.1.

3.2.1 Política e Função de Valor-Ação Q

O princípio básico que norteia as ações do Agente é aH Política. Uma Política (aqui denotada por π) consiste no método de tomada de decisões do Agente. Utilizando o conhecimento previamente aprendido, a Política deve associar a cada Ação possível em um Estado um valor de Probabilidade de que aquela Ação específica seja tomada, ou seja $\pi(s, a)$.

A maneira pela qual o Agente acumula o conhecimento adquirido está na idéia das Funções V

e Q. A Função V (chamada Função Valor-Estado) é uma função que associa a cada Estado um determinado Valor. Funções deste tipo são primariamente empregada em algoritmos preditivos. A Função Q (Chamada Função Valor-Ação) é uma função que associa a cada Ação possível dentro de um Estado um determinado Valor. Funções Q são mais frequentemente empregadas por algoritmos de controle. Neste trabalho, o foco primário será na Função Q.

Em geral, o Agente é codificado de modo que seu objetivo seja alcançado por meio da maximização de uma dada Recompensa obtida a longo prazo. Esta, por sua vez, deve ser obtida sem sacrificar muito aquela adquirida em curto prazo. Neste contexto, cada Ação tomada pode aproximar o agente de seu objetivo ou afastá-lo ainda mais. É evidente que uma Ação que o traga mais próximo deste objetivo deve possuir um Valor maior do que uma Ação que o distancie.

Define-se, portanto, $Q^{\pi}(s,a)$ como a Função Valor-Ação associada a Política π. Essa função associa a cada Estado s e cada Ação tomada a o Valor Esperado da recompensa a ser obtida:

$$Q^{\pi}(s,a) = E_{\pi}\left\{r_t | s_t = s, a_t = a\right\} \tag{3.1}$$

Parte-se do pressuposto de que a cada par Estado-Ação, dada uma política π, existe um valor real do qual o Agente deverá se aproximar. Para efetuar esta aproximação, deve-se utilizar Estimadores. Deste modo, lembrando que o Valor Real do par Estado-Ação sob a política π é denotado por $Q^{\pi}(s,a)$, a estimativa utilizada pelo Agente será denotada por $Q(s,a)$.

Um estimador bastante natural para este valor é a média amostral das recompensas passadas que foram recebidas quando a ação a foi tomada no estado s. Seja k o número de vezes que este evento ocorreu, resultando nas recompensas $r_1, r_2, r_3, ..., r_k$. O valor estimado é dado por:

$$Q(s,a) = \frac{r_1 + r_2 + r_3 + ... + r_k}{k} \tag{3.2}$$

Caso $k = 0$, inicializa-se $Q(s,a)$ com um valor arbitrário. A Lei dos Grandes Números garante que $\lim_{k\to\infty} Q(s,a) = Q^{\pi}(s,a)$[12]. Em outras palavras, dado um número suficientemente grande de passagens pelo par Estado-Ação (s,a), o Valor Estimado da função Q convergirá para seu suposto Valor Real segundo a política π. É relevante para este estudo definir maneiras de comparação entre políticas. Uma política π é definida como melhor que uma política π' se, e somente se, $Q^{\pi}(s,a) \geq Q^{\pi'}(s,a)$ para todos s e a. Sempre existirá pelo menos uma política que é melhor ou igual todas as outras políticas, que é a política ótima. Mesmo podendo haver mais de uma, todas as políticas ótimas são definidas por π^*. Todas elas têm a mesma função de valor, denotada por $Q^*(s,a)$, definida como:

$$Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a) \qquad \forall s, a \tag{3.3}$$

Observe que o cálculo da estimativa dada pela equação 3.2 requer um elevado custo computacional, pois eventualmente toda a memória do computador será ocupada com valores armazenados

de recompensas passadas. Todo esse custo pode ser radicalmente diminuído por meio de uma simples reformulação.

$$\begin{aligned} Q_{k+1} &= \frac{1}{k+1}\sum_{i=1}^{k+1} r_i \\ Q_{k+1} &= \frac{1}{k+1}[r_{k+1} + \sum_{i=1}^{n} r_i] \\ Q_{k+1} &= \frac{1}{k+1}[r_{k+1} + kQ_k + Q_k - Q_k] \\ Q_{k+1} &= \frac{1}{k+1}[r_{k+1} + (k+1)Q_k - Q_k] \\ Q_{k+1} &= Q_k + \frac{1}{k+1}[r_{k+1} - Q_k] \end{aligned} \tag{3.4}$$

A equação resultante 3.4 basicamente expressa:

$$\text{Estimativa nova} \leftarrow \text{estimativa velha} + \text{tamanho do passo}(\text{recompensa} - \text{estimativa velha}) \tag{3.5}$$

A expressão (recompensa − estimativa velha) consiste em uma estimativa de erro que está sendo diminuído a cada passo dado em direção ao "objetivo". Esse algoritmo converge com sucesso para sistemas estacionários. Entretanto, incorre-se em um problema caso seja usado em sistemas dinâmicos. Eventualmente o número de visitas a um par Estado-Ação (k) se torna tão grande que o tamanho do passo $\frac{1}{k+1}$ se torna desprezível e reduz-se drasticamente a capacidade de aprendizado.

Felizmente este problema possui correção simples. Como $\frac{1}{k+1}$ consiste no tamanho do passo em que é efetuada a correção do erro, podemos substituí-lo por uma dada função natural $\alpha(k)$, que associa o número de visitas ao par Estado-Ação ao tamanho do passo a ser dado. Se esta for uma função constante, $\alpha(k) = \alpha$, a função Q evoluirá dinamicamente com o sistema. Dessa forma, Q passa a ser uma média ponderada exponencial dos resultados passados e da estimativa inicial.

$$\begin{aligned} Q_k &= Q_{k-1} + \alpha[r_k - Q_{k-1}] \\ Q_k &= \alpha r_k + (1-\alpha)Q_{k-1} \\ Q_k &= \alpha r_k + (1-\alpha)\alpha r_{k-1} + (1-\alpha)^2 Q_{k-2} \\ Q_k &= \alpha r_k + (1-\alpha)\alpha r_{k-1} + (1-\alpha)^2 \alpha r_{k-2} + \\ &\qquad \ldots + (1-\alpha)^{k-1}\alpha r_1 + (1-\alpha)^k Q_0 \\ Q_k &= (1-\alpha)^k Q_0 + \sum_{i=1}^{k} \alpha(1-\alpha)^{k-i} r_i \end{aligned} \tag{3.6}$$

Um resultado clássico de teoria de aproximação estocástica garante que o algoritmo convergirá com Probabilidade 1 caso o tamanho do passo satisfaça essas duas condições:

$$\sum_{k+1}^{\infty} \alpha(k) = \infty \qquad \text{e} \qquad \sum_{k+1}^{\infty} \alpha^2(k) < \infty \tag{3.7}$$

3.2.2 Retorno

Um importante ponto a ressaltar é que não é tão interessante um sistema que garanta apenas uma recompensa momentânea. O objetivo deve ser maximizar a soma das recompensas recebidas à longo prazo. É possível que ações que maximizam a recompensa imediata sacrifiquem a obtenção de recompensas futuras. Assim, embora a recompensa atual tenha valor elevado, não se deve desconsiderar o valor das expectativas de recompensas futuras. Feita esta consideração, introduz-se um novo parâmetro, chamado fator de desconto (γ), que indica o grau de valoração das recompensas futuras. Deste modo, define-se o Retorno como:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{3.8}$$

Onde os r_j são as recompensas no tempo j e R_t é o retorno. Este valor possui uma significância maior do que apenas as recompensas individuais, pois ele considera as possíveis recompensas futuras. Vale ressaltar que $0 \leq \gamma \leq 1$. Caso $\gamma = 0$, a soma total considerada será apenas a próxima recompensa, o agente se tornará "míope". À medida que γ aumenta, o agente tenderá a "enxergar" melhor o futuro. Uma recompensa que será recebida k passos de tempo no futuro possuirá um impacto proporcional a γ^{k-1} no presente.

Deste modo, redefine-se a função Q^π para abarcar este novo conceito:

$$Q^\pi(s,a) = E_\pi \left\{ R_t | s_t = s, a_t = a \right\} \tag{3.9}$$

Rearranjando os termos e expandindo o cálculo da esperança, obtemos:

$$\begin{aligned}
Q^\pi(s,a) &= E_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a \right\} \\
Q^\pi(s,a) &= E_\pi \left\{ r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_t = s, a_t = a \right\} \\
Q^\pi(s,a) &= \sum_a \pi(s,a) \sum_{s'} P^a_{ss'} \left[R^a_{ss'} + \gamma E_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_{t+1} = s', a_{t+1} = a' \right\} \right]
\end{aligned} \tag{3.10}$$

Por fim, obtém-se:

$$Q^\pi(s,a) = \sum_a \pi(s,a) \sum_{s'} P^a_{ss'} \left[R^a_{ss'} + \gamma Q^\pi(s') \right] \tag{3.11}$$

3.2.3 Propriedade de Markov

Uma parte importante de se construir um sistema que usa Aprendizagem por Reforço é a escolha adequada na codificação dos estados. Esta escolha deve refletir um resumo compacto das

sensações passadas do agente de forma que todas as informações relevantes estejam disponíveis. Isto geralmente requer que um estado seja mais do que as sensações imediatas experimentadas pelo agente, mas nunca mais do que seu histórico completo de sensações. Um sinal de estado que obtenha sucesso em representar toda a informação relevante é dito ser Markov, ou ter a propriedade de Markov. Um exemplo é a posição e a velocidade de uma bola de canhão. De acordo com a Mecânica Clássica Newtoniana, posição e velocidade são as únicas informações relevantes para este problema, e portanto possuem a propriedade de Markov.

Considere um ambiente arbitrário respondendo em um instante de tempo $t+1$ a uma dada ação tomada no instante de tempo t. Em sua forma mais geral (sob hipótese de causalidade), essa resposta pode depender de todo o histórico de ações. Neste caso, a única alternativa para definir a dinâmica do sistema é a especificação da distribuição de probabilidades completa:

$$Pr\left\{s_{t+1}=s', r_{t+1}=r|s_j, a_j, r_j \quad \forall j=0..t\right\} \tag{3.12}$$

Em outras palavras, a probabilidade de um dado evento sempre deve ser dada por uma função de probabilidade condicional que dependa de todo o histórico do sistema. Entretanto, se o sinal do estado tem a propriedade de Markov, sua resposta no instante $t+1$ depende somente das representações de estado e ação no instante t. Desta forma, pode ser representado por:

$$Pr\left\{s_{t+1}=s', r_{t+1}=r|s_t, a_t\right\} \tag{3.13}$$

Em outras palavras, o estado possui a propriedade de Markov, se e somente se a equação 3.12 é igual à equação 3.13 para todos os valores possíveis de estado, ação e recompensa.

Todavia, mesmo que o estado não possua a propriedade de Markov, sob determinadas condições, é válido se valer de aproximações que o considerem como possuidor desta propriedade. Estados com esta propriedade possuem a vantagem de melhor prever estados subsequentes. Portanto, um estado Markov provê a melhor base de escolha de ações.

Um problema de aprendizagem por reforço que satisfaça a propriedade de Markov é chamado de Processo de Decisão de Markov (PDM). Se o espaço de estados e ações é finito, é então chamado de Processo de Decisão de Markov Finito (PDMF).

Num PDMF, dados quaisquer ações e estados a e s, a probabilidade de cada próximo estado s' é:

$$P^a_{ss'} = Pr\left\{s_{t+1}=s'|s_t=s, a_t=a\right\} \tag{3.14}$$

Da mesma forma, dados quaisquer estados e ações s e a presentes, junto com qualquer estado futuro s', o valor esperado da próxima recompensa é:

$$R^a_{ss'} = E\left\{r_{t+1}|s_t=s, a_t=a, s_{t+1}=s'\right\} \tag{3.15}$$

As equações 3.14 e 3.15 especificam os aspectos mais importantes da dinâmica de uma PDMF.

3.2.4 Políticas Frequentes

Até o momento, falou-se bastante a respeito da Função Q e sobre as técnicas para efetuar sua aproximação por meio de estimadores. Entretanto, ainda não se falou a respeito das Políticas que farão uso desta aproximação em seus algoritmos de tomada de decisão. Existem algumas políticas bastante frequentes no que tange o modelo de Aprendizagem por Reforço que serão explicadas a Seguir.

Gananciosa (*Greedy*)

A ação escolhida é aquela que possui maior Valor. Como o Valor será dado pela estimativa $Q(s,a)$ e não pelo valor real $Q^{\pi}(s,a)$, ao utilizar esta política corre-se o risco de ficar preso em estados de máximo local, pois existe pouco espaço para desbravamento (*exploration*), e foca-se apenas no aproveitamento do conhecimento possuido (*exploitation*).

ϵ-Gananciosa (ϵ-*Greedy*)

É gananciosa $(1-\epsilon)\%$ das vezes. A política vai escolher uma ação aleatória com probabilidade $\epsilon\%$ de chance. Com isso, garante-se que o Agente, caso encontre-se em uma situação de máximo local possa, com uma probabilidade ϵ optar por uma Ação que não lhe aparente a melhor, com o objetivo de desbravamento (*exploration*). Entretanto, $(1-\epsilon)\%$ das ocasiões, ela opta por aproveitar de seu conhecimento na obtenção de recompensas (*exploitation*).

SoftMax

A probabilidade de cada ação tomada segue uma distribuição de Boltzman:

$$P(a) = \frac{e^{Q(s,a)/\tau}}{\sum_{i=1}^{n} e^{Q(s,a_i)/\tau}} \tag{3.16}$$

Onde τ é um parâmetro de temperatura. Para temperaturas altas, todas as ações serão aproximadamente equiprováveis. Conforme a temperatura é reduzida, a probabilidade de se tomar a ação de maior valor estimado tende a 1. Por causa disso, com o tempo, as ações de maior valor são tomadas com maior frequência, enquanto as ações de menor valor se tornam mais raras, sem nunca se tornarem impossíveis. Deste modo, garante-se o desbravamento de novas ações (*exploration*) sem prejudicar o aproveitamento do conhecimento obtido (*exploitation*).

3.2.5 TD

A principal ideia por trás da Aprendizagem por Reforço é o aprendizado por meio de diferenças temporais, ou *temporal difference learning* (TD). O aprendizado TD não espera o final de um episódio para poder extrair informações, como num método *offline*[1]. Seu uso permite o aprendizado incremental no tempo, em parte, por meio de estimativas aprendidas anteriormente.

[1] Em um treinamento *online*, as atualizações são feitas durante o episódio, assim que o incremento é computado. Por outro lado, em um treinamento *offline* os incrementos são acumulados à parte e não são usados para se atualizar as estimativas de valor até o final do episódio.[12]

Para efeitos de comparação, considere o seguinte método *offline*. Trata-se de um método bastante simples. Primeiramente, espera-se a conclusão de um dado episódio e a consolidação de todos os Retornos. Por fim, atualiza-se simultaneamente (de maneira *offline*) toda a tabela simultaneamente. A sua equação de atualização da tabela Q é dada por:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[R_t - Q(s_t, a_t)\right] \tag{3.17}$$

Onde R_t é o retorno real recebido após o instante de tempo t e α é o parâmetro que denota o tamanho do passo. Este é um exemplo de método de uma classe maior conhecida como métodos Monte Carlo[2]. Métodos Monte Carlo precisam esperar até o final de um episódio para determinar os incrementos de $Q(s_t, a_t)$, quando R_t se tornar conhecido.

Métodos TD atualizam sua tabela Q em todos os instantes de tempo, sendo portanto passíveis de treinamento *online*. No instante de tempo $t + 1$, um método TD imediatamente faz uma atualização utilizando a recompensa r_{t+1} e a estimativa $Q(s_{t+1}, a_{t+1})$. O método TD mais simples, conhecido como TD(0), é dado por:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\right] \tag{3.18}$$

Essencialmente, um método Monte Carlo espera a consolidação do valor de R_t para se fazer a atualização. Já o método TD faz estimativas sucessivas que convergem para este valor ($r_{t+1} + \gamma Q(s_{t+1}, a_{t+1})$). Grosseiramente falando, métodos Monte Carlo usam uma estimativa da equação 3.9 no processo de atualização enquanto métodos TD usam uma estimativa da equação 3.11. Sob a hipótese de um valor α adequado, os métodos Monte Carlo e TD(0) possuem garantia de convergência pela Lei dos Grandes Números[12]. Entretanto, em geral, ambos convergem para valores distintos.

Deste modo, verifica-se que métodos Monte Carlo são de fato ótimos, porém de uma maneira limitada. Isto se deve à utilização do retorno obtido no passado como uma estimativa do retorno futuro. Isso os torna ótimos segundo a informação existente. Entretanto o interesse do agente reside em reduzir o erro com relação a informações futuras. Neste quesito métodos TD(0) efetuam a mesma tarefa de maneira mais adequada.

3.2.5.1 Rastros de Elegibilidade

Um problema comum associado à Aprendizagem por Reforço é o fato que, em geral, apenas à ação tomada no último estado é atribuída a recompensa. Isto é problemático na medida em que diversas outras ações foram tomadas apenas para que o agente pudesse chegar naquele estado. Deste modo, parece sensato recompensar os pares Estado-Ação intermediários que viabilizaram aquele

[2]Em geral, a expressão Método de Monte Carlo está associada a qualquer método que utilize simulações aleatórias para aproximar um dado resultado. No contexto de Aprendizado por Reforço, no entanto, Monte Carlo denota apenas uma classe específica de métodos.

estado. Um sistema pelo qual estes pares Estado-Ação passados recebam recompensas por terem aberto caminho para o estado presente é o assim chamado sistema de Rastros de Elegibilidade.

O sistema de Rastros de Elegibilidade associa a cada par Estado-Ação um valor que indica o quão "elegível" ele é a uma determinada recompensa. Em outras palavras, a elegibilidade de um par Estado-Ação representaria o quão relevante ele foi ao conduzir ao estado atual, e portanto qual a parte da recompensa que lhe cabe. Em geral, interpreta-se que ações recentes tem uma relevância maior do que ações passadas. Deste modo, por meio de um coeficiente de elegibilidade, λ, cada estado no passado se torna exponencialmente menos elegível às recompensas.

Este sistema também fornece uma ponte bastante elegante entre os métodos TD e Monte Carlo. Eles produzem um espectro de métodos chamados de TD(λ). Se de um lado tem-se TD(0), um algoritmo de diferença temporal simples que distribui a recompensa em apenas 1 passo de tempo, do outro está o TD(1), que é o próprio método de Monte Carlo codificado para um aprendizado *online*, que a cada passo recalcula todos os Retornos anteriores e, portanto, atualiza a tabela Q em todos os pontos passados. É interessante notar que métodos onde $0 < \lambda < 1$ geralmente se mostram bem melhores do que quando λ está em alguma das extremidades[12].

Em sua forma tabular, o sistema de rastros de elegibilidade consiste de uma tabela $e(s, a)$ do mesmo tamanho da tabela $Q(s, a)$. A cada passagem por um par (s_t, a_t), incrementa-se 1 na entrada correspondente da tabela, $e(s_t, a_t)$. Além disso, em todos os períodos de tempo, a tabela deve sofrer um decaimento de λ, na forma:

$$e(s, a) \leftarrow \gamma\lambda e(s, a) \tag{3.19}$$

Sendo $0 < \gamma < 1$ e $0 < \lambda < 1$. Note que, desta maneira, eventos distanciados no tempo possuem uma elegibilidade cada vez menor às recompensas atuais.

3.2.5.2 Algoritmo SARSA com rastros de elegibilidade

SARSA é um dos algoritmos de controle na área de Aprendizado por Reforço mais famosos e mais frequentemente utilizados. Suas raízes datam de 1994, quando foi primeiramente explorado por Rummery e Niranjan[13], chamando-o de *Q-Learning* modificado. O nome SARSA foi introduzido por Sutton[14] em 1996, sendo um acrônimo das entradas da fórmula 3.18: s, a, r, s' a'. O algoritmo completo na sua versão usando os rastros de elegibilidade está descrito abaixo.

Introduzido primeiramente em 1989 na tese de doutorado de Chris Watkins[15], o algoritmo *Q-Learning*, diferentemente do SARSA, escolhe o a_+ ótimo, ou seja, a ação futura de maior valor para constituir o δ. Esta é a única diferença entre os dois algoritmos, fazendo com que o *Q-learning* seja um algoritmo *off-policy*, o que significa que sua tabela Q é atualizada para um valor que não segue a Política adotada. Em outras palavras, emboras ambos convirjam para uma Política Ótima, o SARSA o faz por meio de atualizações utilizando valores dos pares Estado-Ação efetivamente visitados, enquanto o *Q-learning* o faz sempre empregando a Ação de maior Valor disponível no Estado. Caso o algoritmo siga uma Política Gananciosa, ambos são exatamente iguais.

Algorithm 1 SARSA com Rastros de Elegibilidade, a ser rodado por N episódios

Inicialize $Q(s, a)$ arbitrariamente;
$e(s, a) \leftarrow 0$ para todos os pares (s, a);
for j=1:N **do** ▷ Repetir para cada um dos N episódios
    Receba o estado inicial s;
    De acordo com a política π e o estado s, escolha a ação a ;
    while s $\neq$ Estado Terminal **do** ▷ Loop a ser rodado até atingir o estado final
        Realize a ação a;
        Receba a recompensa r;
        Observe o próximo estado s_+;
        De acordo com a política π e o estado s_+, escolha a ação a_+
        $\delta \leftarrow r + \gamma Q(s_+, a_+) - Q(s, a)$;
        $e(s, a) \leftarrow e(s, a) + 1$
        for todos os pares Estado-Ação possíveis (s', a') **do**
            $Q(s', a') \leftarrow Q(s', a') + \alpha\delta e(s', a')$;
            $e(s', a') \leftarrow \lambda\gamma e(s', a')$;
        end for
        $s \leftarrow s_+$
        $a \leftarrow a_+$
    end while
end for

Capítulo 4

Desenvolvimento

"Make everything as simple as possible, but not simpler" - Albert Einstein

4.1 Introdução

Este capítulo tratará sobre o processo de desenvolvimento do Sistema Inteligente e do *Benchmark* de comparação. Para melhor explanar os caminhos e decisões tomadas, este capítulo será dividido em partes. A primeira parte destina-se a explicar a codificação de estados empregada. A segunda trata da codificação das ações possíveis. A terceira parte visa esclarecer as decisões tomadas a respeito do sistema de recompensas. A quarta parte falará sobre o sistema utilizado como *Benchmark*. Por fim, a última parte se destinará à explicação dos parâmetros de comparação entre o sistema inteligente e os *Benchmarks*.

4.2 Considerações Iniciais

Geralmente a Bovespa abre seus mercados nos dias úteis às 10 da manhã e os fecha por volta de 17 horas. Após o fechamento, um agente financeiro ainda pode fazer ofertas de venda ou de compra, mas, mesmo que essas ofertas superem o *Bid-Ask Spread*, as operações só poderão ser realizadas na próxima abertura do mercado. Geralmente, um agente financeiro que costuma fazer suas operações *intraday*[1] prefere, por segurança, não segurar instrumentos de um dia para outro, pois durante a noite pode acontecer algo que abale os ânimos dos outros agentes e mude radicalmente os preços.

Na realização desse trabalho, no entanto, algumas simplificações foram feitas de modo a facilitar sua implementação. Uma abordagem mais realista trataria cada dia particular no mercado como um episódio. Entretanto, este trabalho optou por tratar todo o processo de decisão como um único episódio infinito, como se o mercado nunca fechasse. Deste modo, o sistema inteligente não enxerga objeções em comprar ações ao término de um dia de mercado para vendê-las apenas no dia seguinte. Além disso, todos os custos de corretagem e *Slippage* são desconsiderados.

Segundo Sutton[12], é crítico que a recompensa fornecida ao Agente realmente indique o objetivo a alcançar, e não tente indicar como este objetivo deve ser alcançado. À vista disso, surgiu

[1] Comprar e vender no mesmo dia

um problema sobre como codificar o sistema de recompensas. A primeira ideia seria recompensar a ação de compra com o lucro obtido. Entretanto, neste contexto, seria possível recompensar boas compras de maneira inadequada, caso nelas se efetuasse uma venda ruim. Analogamente, seria possível recompensar de maneira exagerada compras ruins, devido a uma venda bem feita. Para se contornar este problema, decidiu-se utilizar o Sistema Inteligente apenas como método de entrada, ficando a saída a cargo de um dado sistema fixado, baseado em estratégias clássicas.

4.3 Codificação de Estados

Com estas ressalvas em mente, parte-se para a explicação a respeito da codificação de estados. Embora houvesse a opção de se utilizar apenas as informações de preços de maneira "crua" (isto é, sem pré-processamento), verificamos que isto traria implicações graves ao sistema. Isto porque para obter estados mesmo que aproximadamente dotados da propriedade de Markov, seria necessário um número exorbitante de estados. Deste modo, optou-se por uma saída distinta.

Com base nos dois princípios fundamentais da Análise Técnica:

- O mercado tende sempre a voltar à média;
- O mercado usualmente segue tendências;

Optou-se pelo uso de indicadores técnicos que refletissem de maneira eficiente estes princípios e que estivessem aproximadamente contidos em um intervalo finito. Deste modo, mostrou-se necessário o uso de pelo menos 3 indicadores.

- O primeiro destinou-se a atender o problema de refletir a distância da média. Optou-se pelo uso de uma adaptação das Bandas de Bollinger como oscilador. Codificou-se o intervalo entre as duas bandas discretizando-o em 20 níveis. Acrescentou-se ainda um nível representante dos preços acima da banda superior e um representante dos preços abaixo da banda inferior, resultando em um total de 22 possíveis níveis. A localização do preço de fechamento determina o nível presente.
- Para indicar a direção das tendências, optou-se pelo uso do *HiLo Activator*, também adaptado. Optou-se por ignorar os sinais *Buy Stop* e *Sell Stop* e concentrar-se apenas na tendência em vigor. Codificou-se com 1 a tendência de descida e com 2 a tendência de subida.
- Por fim, utilizou-se um indicador de força de tendências. Para esta tarefa utilizou-se o ADX *(Average Directional Index)*, uma vez que este é bastante consagrado. Como este se trata de um valor que varia entre 0 e 100, cada intervalo de 5 foi codificado como um nível possível, totalizando 20 níveis.

Desta forma, existem $22 \times 20 \times 2 = 880$ possíveis estados.

4.4 Codificação das Ações

Para efeitos de simplificação da análise, o sistema considerado operará seus instrumentos apenas no modo *Long*, nunca em *Short*. Futuramente, pode ser criado um sistema idêntico treinado em *Short* para se operar em paralelo. Deste modo, como ponto de partida, considera-se sempre que o sistema se encontra fora do mercado, esperando o momento adequado de entrada. Suas ações disponíveis, neste sentido a cada instante de tempo são:

- Manter-se fora do mercado. Esta ação foi codificada com o valor 1
- Efetuar a compra. Esta ação foi codificada com o valor 2

Outra simplificação efetuada no desenvolvimento do sistema é o fato de que ele apenas opta por comprar ou não. Entretanto, um investidor real além de decidir o momento da compra, deve optar qual percentual de seu capital ele deve investir, fazendo assim o chamado *Money Management*. Aqui o Agente sempre opta por investir 100% de seu capital.

Tais simplificações se mostraram necessárias uma vez que se optou pela codificação da função Q na forma tabular. Com esta codificação adotada, a tabela possui $22 \times 20 \times 2 \times 2 = 1760$ entradas. Um aumento no número de estados ou ações possíveis poderia inviabilizar todo o sistema.

4.5 O Algoritmo

Uma vez codificados o espaço Estado-Ação, parte-se para a escolha do sistema inteligente a ser empregado. Neste trabalho se fará uso de uma política ϵ-gananciosa, com ϵ constante, pela necessidade contínua desbravamento nesse sistema altamente dinâmico que é a bolsa de valores. Optou-se pelo uso do SARSA tabular com Rastros de Elegibilidade. Este sistema, tal qual descrito no algoritmo 1, por aprender o $Q^{\pi}(s, a)$, os valores das ações-estados numa política π, parece se mostrar mais cauteloso que o *Q-Learning* em alguns exemplos do livro do Sutton[12] conseguindo auferir um retorno maior com o tempo.

Apenas uma modificação se mostrou necessária no algoritmo, e esta diz respeito aos Rastros de Elegibilidade. Considerou-se que as compras são mutuamente independentes, de modo que as decisões de compras passadas não interferem na compra presente. Deste modo, o algoritmo utilizado a cada compra partilha as recompensas com os pontos imediatamente anteriores onde se decidiu esperar, e depois zera a tabela de Rastros de Elegibilidade. Pode-se interpretar que cada compra consiste em um evento do sistema.

4.6 Codificação das Recompensas

Com este algoritmo em mente, por fim, resta apenas a codificação do sistema de recompensas. Para ações de compra, a recompensa dada consistiu no lucro percentual normalizado, uma resposta

bastante intuitiva para o problema. Para ações de espera, a recompensa foi sempre 1: foi utilizado uma recompensa para a manutenção baseando-se em um sistema a renda fixa.

Note que, embora a ação de manutenção tomada pelo sistema inteligente receba sempre recompensa constante igual a 1, isto não significa que seu valor será sempre unitário. Isto se deve ao fato de que, eventualmente, os Rastros de Elegibilidade atribuirão a esta ação uma recompensa recebida por uma ação de compra futura, sendo esta a principal maneira de obtenção de recompensas de um estado de manutenção. Afinal, em certos estados é muito mais valoroso esperar do que fazer uma compra ruim.

Outra decisão tomada foi reforçar a intensidade do lucro percentual. Na realimentação desse lucro na recompensa, multiplicou-se o lucro por seu valor absoluto, tendo isso o efeito de reforçar com intensidade quadrática compras bem efetuadas, e punir também com intensidade quadrática as compras inadequadas.

4.7 Sistema Benchmark

Para efeitos de comparação, e para justificar o uso de uma estratégia inteligente, optou-se por utilizar estratégias clássicas baseada no *HiLo Activator*, Bandas de Bollinger e *Stop Loss* e *Stop Gain*. Estas estratégias consagradas fazem o uso dos indicadores apresentados nas seções 2.5.5 e 2.5.7, respectivamente.

Para fins analíticos, a estratégia de compra do sistema inteligente foi sempre comparada usando a mesma estratégia de venda de um *trading system* clássico. Deste modo é possível tentar se provar uma suposta superioridade do algoritmo inteligente em relação a um indicador clássico simples.

A estratégia HiLo de Entrada consiste em efetuar a compra somente quando o fechamento do *candle* atual cruzar acima da média móvel aritmética de 7 períodos dos preços máximos dos *candles* anteriores. A de saída, por sua vez, era indicada pelo cruzamento abaixo da MMA de 7 períodos dos preços mínimos anteriores. A estratégia Bollinger consiste em comprar nos dois primeiros níveis e sair após os preços subirem além da média móvel exponencial de 20 períodos. Uma última estratégia testada, a dos *Stops*, foi empregada utilizando a estratégia HiLo como entrada e usando *Stop Loss* ou *Stop Gain* para sair do mercado.

Outro Benchmark natural para a comparação seria o análogo a um algoritmo ingênuo, o *Buy & Hold* no mesmo período. *Buy & Hold* consiste em comprar no primeiro período de comparação e vender apenas no último. Este simula a ação de um agente financeiro que não opera segundo as flutuações do mercado. Apesar de ser um Benchmark Clássico para a análise de *Trading Systems*, não se deve esperar muito dele. Isso se dá porque esse trabalho não se propõe a fazer um *trading system* completo, mas só um algoritmo de compras para mostrar que sistemas inteligentes podem ser aplicados com sucesso a Mercados Financeiros. O lucro final auferido pelo sistema resulta em grande medida da sua estratégia de venda, parte que não é o foco deste trabalho.

4.8 Critérios de Comparação

Os critérios de comparação entre estratégias, com a exceção do Sharpe Ratio, foram extraídos do livro "Trading Systems"[16] de Emilio Tomasini e Urban Jaekle. Eles visam simular o emprego de estratégias de maneira sistemática, e verificar seu comportamento quando aplicadas a determinados ativos. Para a análise da performance da estratégia, os autores deste livro empregaram medidas que avaliam não só o lucro obtido naquela simulação em particular, mas detalhes particulares a respeito de como este lucro foi obtido. A seguir, temos uma lista com as métricas empregadas neste trabalho, inspiradas por este livro.

Lucro Total Líquido Consiste no lucro obtido no término do investimento, ao se seguir a risca a estratégia empregada.

Lucro Bruto Consiste na somatória total dos lucros obtidos, desconsiderando todos os prejuízos.

Prejuízo Bruto Consiste na somatória total de todos os prejuízos sofridos, desconsiderando todos os lucros.

Fator de Lucro Consiste na razão entre o Lucro Bruto e o Prejuízo Bruto. Indica a proporção de Lucros para Prejuízos da estratégia analisada.

Número de Operações Consiste no total de operações efetuadas pela estratégia. No caso deste trabalho em particular, todas as operações foram de compra, não havendo operações de venda ou de reposicionamento.

Operações Vencedoras Consiste no número de operações que encerraram extraindo lucro do mercado.

Operações Perdedoras Consiste no número de operações que encerraram assumindo um prejuízo.

Percentual de Operações Lucrativas Consiste na razão percentual entre o número de operações vencedoras e o número total de operações realizado.

Operação Vencedora Média Consiste na média aritmética do lucro obtido considerando-se apenas as operações vencedoras.

Operação Perdedora Média Consiste na média aritmética do prejuízo assumido considerando-se apenas operações perdedoras.

Taxa Média Vencedora/Média Perdedora Consiste na razão entre operação média vencedora e a operação média perdedora.

Maior Operação Vencedora Consiste no valor do lucro obtido com a maior operação vencedora.

Maior Operação Perdedora Consiste no valor do prejuízo assumido com a pior operação perdedora.

Máximo de Operações Vencedoras Consecutivas Consiste no maior número de operações vencedoras que ocorreram sem que ocorresse entre elas uma operação perdedora

Máximo de Operações Perdedoras Consecutivas Consiste na maior sequência de operações perdedoras experimentada pela estratégia.

Média de Períodos em Operações Vencedoras Consiste no número de intervalos de tempo médio experimentado pela estratégia até que o lucro tenha sido obtido.

Média de Períodos em Operações Perdedoras Consiste no número de intervalos de tempo médio que a estratégia levou até assumir o prejuízo.

Sharpe Ratio Sharpe Ratio da estratégia.

Buy and Hold no mesmo Período Consiste em considerar uma compra efetuada no fechamento do primeiro *candle* do sinal e uma venda no fechamento do último *candle*, simulando um acionista que possua o ativo 100% do tempo.

Capítulo 5

Resultados Experimentais

5.1 Introdução

As simulações foram feitas utilizando-se o software MatLab R2013a©. Utilizou-se dados extraídos do programa ProfitChart© e MetaStock©, com granularidade de minuto a minuto, 5 em 5 minutos, 10 em 10 minutos, 15 em 15 minutos, 30 em 30 minutos, 60 em 60 minutos e diário. Utilizou-se a série histórica do Índice Bovespa (o chamado IBOV)[1] e dos instrumentos de maior liquidez da BM&FBovespa, PETR4, VALE5, CMIG4, DOLFUT, além do índice Dow Jones DJI.

Para atender aos critérios de treinamento e validação, dividiu-se parte do sinal histórico a ser utilizado exclusivamente para o treinamento do Sistema Inteligente, e a parte seguinte como comparativo entre o Sistema Inteligente e os outros dois sistemas de entrada. Em todos os casos, dado o conjunto de treinamento inicial, optou-se arbitrariamente em dividi-lo exatamente na metade, sendo que a metade inicial foi utilizada exclusivamente para treinamento, e ambos os algoritmos foram rodados e comparados na segunda metade.

Importante ressaltar que o algoritmo ainda aprende enquanto medimos seus resultados e se adapta a possíveis mudanças. Segue, na seção seguinte, a tabela de resultados obtidos. Todos os efeitos de *Slippage* e Taxas de Corretagem foram desconsiderados nestas simulações

5.2 Resultados

Nesta seção serão apresentados os diversos resultados obtidos nas simulações. Cada resultado será compilado em uma tabela mostrando os parâmetros explicados no capítulo anterior. Além disso, serão exibidos gráficos que melhor ilustram o comportamento do sistema sob aquelas condições

[1] o Cálculo do índice Bovespa é explicado no site da Bovespa: http://www.bmfbovespa.com.br/

5.2.1 Granularidade 1m - IBOV

Foi testado o Sistema Inteligente competindo contra um Sistema HiLo na entrada. Na saída, ambos consistiam em saídas por *Stop Loss* e *Stop Gain*. Haviam disponíveis 63752 pontos disponíveis. 31876 destes foram utilizados apenas para treinamento enquanto 31876 foram utilizados para comparação

Tabela 5.1: Relatório Comparativo de estratégias

Estratégia de entrada	Sistema Inteligente	Sistema clássico
Instrumento usado:	IBOV	IBOV
Granularidade:	1m	1m
Data de inicio:	07-Aug-2013	07-Aug-2013
Data de termino:	29-Nov-2013	29-Nov-2013
Lucro Total:	R$ 22749.03	R$ 11073.94
Lucro Bruto:	R$ 122970.65	R$ 68100.19
Prejuízo Bruto:	R$ -100221.62	R$ -57026.25
Fator de Lucro:	1.23	1.19
Número Total de Operações:	3333	1868
Percentual de Vencedoras:	76.06 %	76.82 %
Operações Vencedoras:	2535	1435
Operações Perdedoras:	798	433
Lucro Médio Total:	R$ 6.83	R$ 5.93
Lucro Médio das Vencedoras:	R$ 48.51	R$ 47.46
Prejuízo Médio das Perdedoras:	R$ -125.59	R$ -131.70
Razão Média Vencedoras/Média Perdedoras:	0.39	0.36
Maior Operação Vencedora:	R$ 932.54	R$ 383.64
Pior Operação Perdedora:	R$ -745.21	R$ -406.74
Maior Número de Vitórias Consecutivas:	33	24
Maior Número de Derrotas Consecutivas:	6	6
Média de Tempo das Operações:	7.56	8.29
Média de Tempo das Operações Vencedoras:	5.66	6.18
Média de Tempo das Operações Perdedoras:	13.58	15.28
Sharpe Ratio:	0.074439	0.067427
Estrategia Buy and Hold no mesmo período:	R$ 5156.36	R$ 5156.36

A diante, seguem os gráficos do percentual de lucro acumulado de ambas as estratégias comparadas:

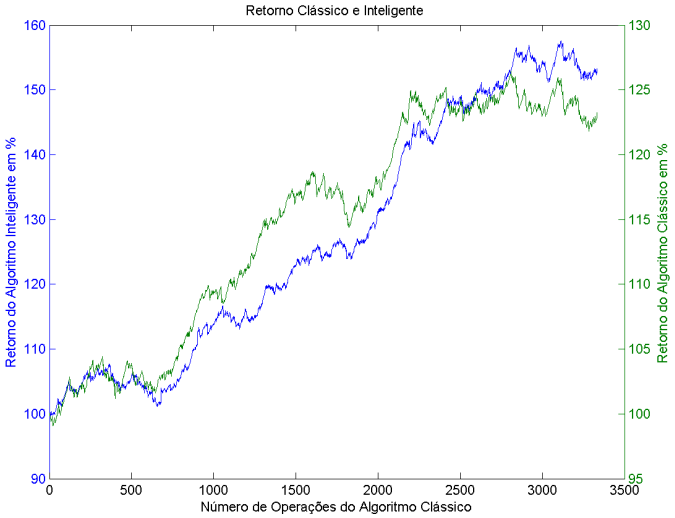

Figura 5.1: Comparação do Lucro Percentual Acumulado das Estratégias - IBOV 1m

O próximo gráfico ilustra a Tabela Q contida no interior do Sistema Inteligente. Este gráfico ilustra a valoração associada de cada estado considerado.

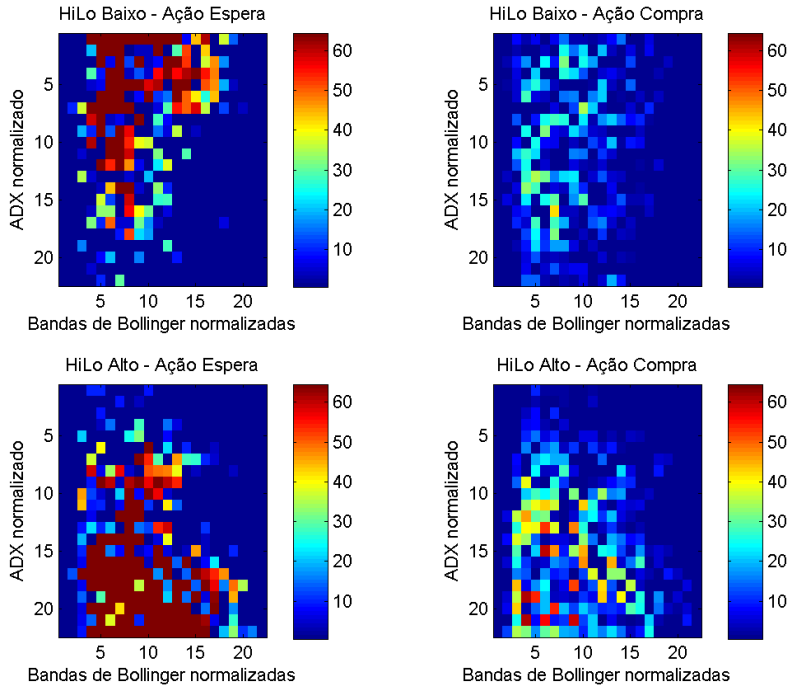

Figura 5.2: Tabela Q final do Sistema Inteligente - IBOV 1m

Em seguida, apresentam-se dois gráficos indicando a distribuição do lucro obtido com os trades.

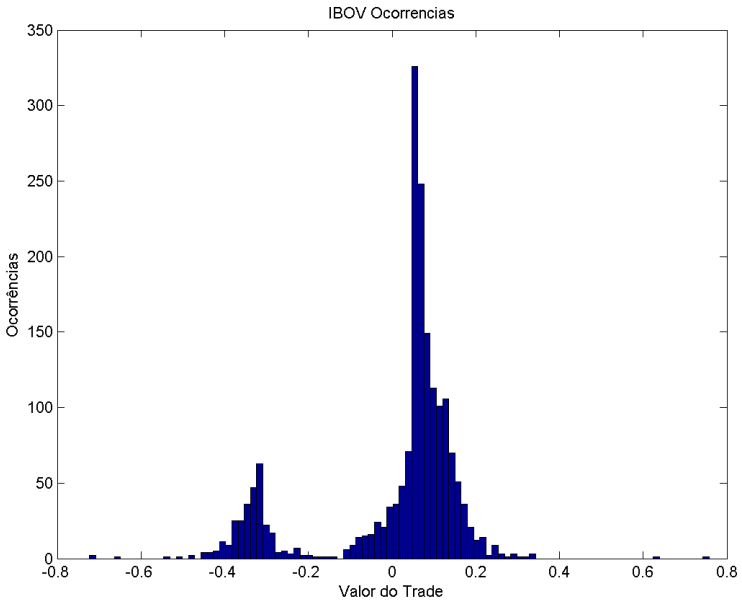

Figura 5.3: Distribuição do lucro na Estratégia Clássica - IBOV 1m

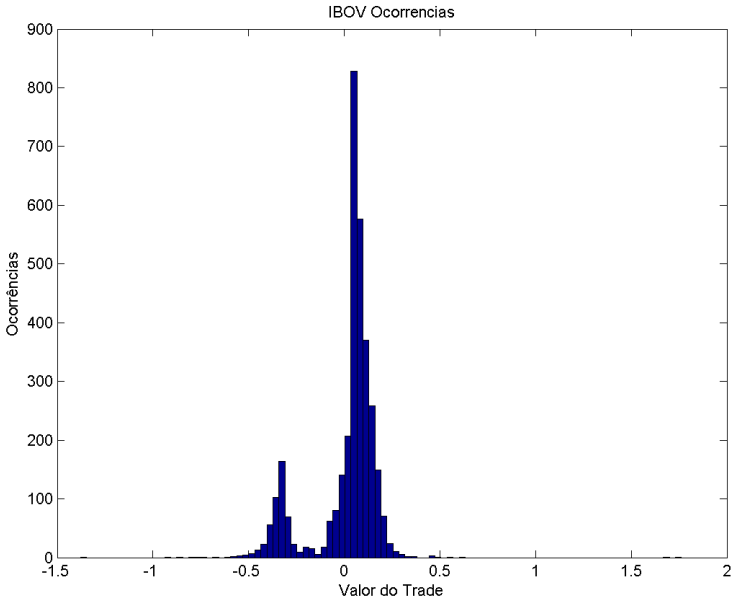

Figura 5.4: Distribuição do lucro no Sistema Inteligente - IBOV 1m

5.2.2 Granularidade 5 m - DJI

Foi testado o Sistema Inteligente competindo contra um Sistema HiLo na entrada. Na saída, ambos consistiam em saídas também controladas por um sistema HiLo. Haviam disponíveis 33302 pontos disponíveis. 16651 destes foram utilizados apenas para treinamento enquanto 16651 foram utilizados para comparação

Tabela 5.2: Relatório Comparativo de estratégias

Estratégia de entrada	Sistema Inteligente	Sistema clássico
Instrumento usado:	DJI	DJI
Granularidade:	5m	5m
Data de inicio:	06-Feb-2013	06-Feb-2013
Data de termino:	06-Dec-2013	06-Dec-2013
Lucro Total:	R$ 1714.49	R$ -110.00
Lucro Bruto:	R$ 12852.34	R$ 8951.86
Prejuízo Bruto:	R$ -11137.85	R$ -9061.86
Fator de Lucro:	1.15	0.99
Número Total de Operações:	972	972
Percentual de Vencedoras:	50.00 %	40.33 %
Operações Vencedoras:	486	392
Operações Perdedoras:	486	579
Lucro Médio Total:	R$ 1.76	R$ -0.11
Lucro Médio das Vencedoras:	R$ 26.45	R$ 22.84
Prejuízo Médio das Perdedoras:	R$ -22.92	R$ -15.65
Razão Média Vencedoras/Média Perdedoras:	1.15	1.46
Maior Operação Vencedora:	R$ 235.57	R$ 216.19
Pior Operação Perdedora:	R$ -215.22	R$ -138.20
Maior Número de Vitórias Consecutivas:	12	8
Maior Número de Derrotas Consecutivas:	9	10
Média de Tempo das Operações:	14.44	9.06
Média de Tempo das Operações Vencedoras:	17.09	13.77
Média de Tempo das Operações Perdedoras:	11.79	5.87
Sharpe Ratio:	0.045699	-0.003798
Estrategia Buy and Hold no mesmo período:	R$ 1851.32	R$ 1851.32

A diante, seguem os gráficos do percentual de lucro acumulado de ambas as estratégias comparadas:

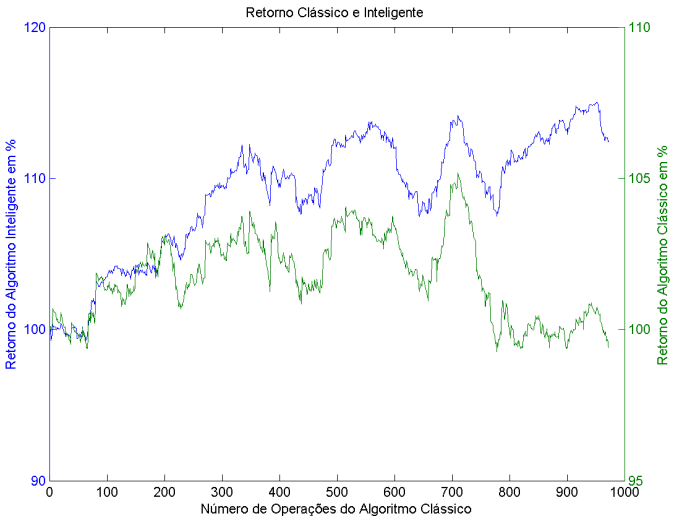

Figura 5.5: Comparação do Lucro Percentual Acumulado das Estratégias - DJI 5m

A seguir, a Tabela Q:

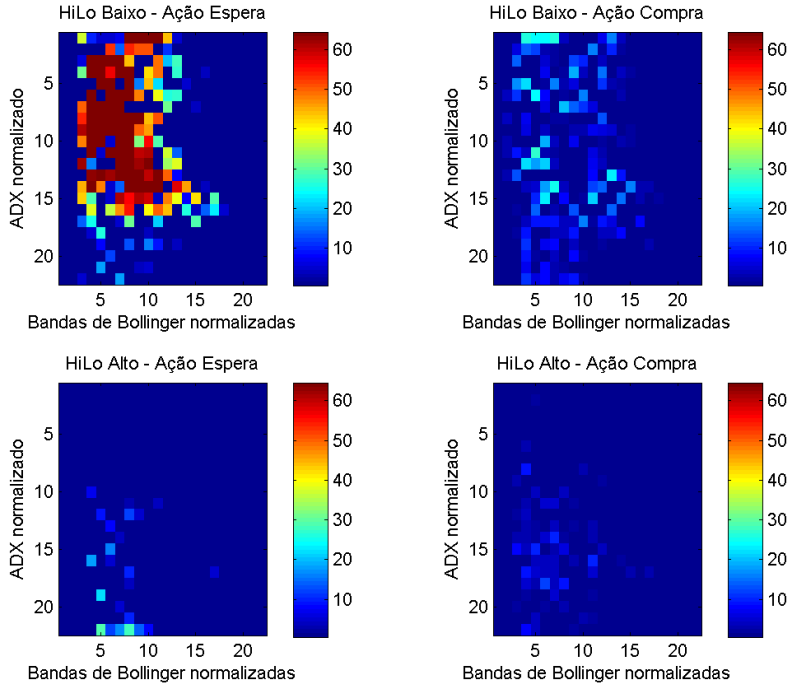

Figura 5.6: Tabela Q final do Sistema Inteligente - DJI 5m

Por fim, apresentam-se dois gráficos indicando a distribuição do lucro nas operações.

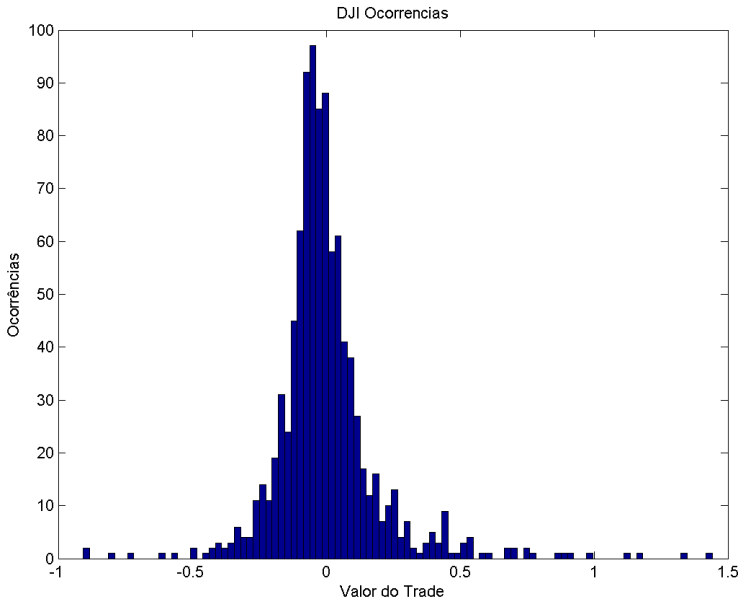

Figura 5.7: Distribuição do lucro na Estratégia Clássica - DJI 5m

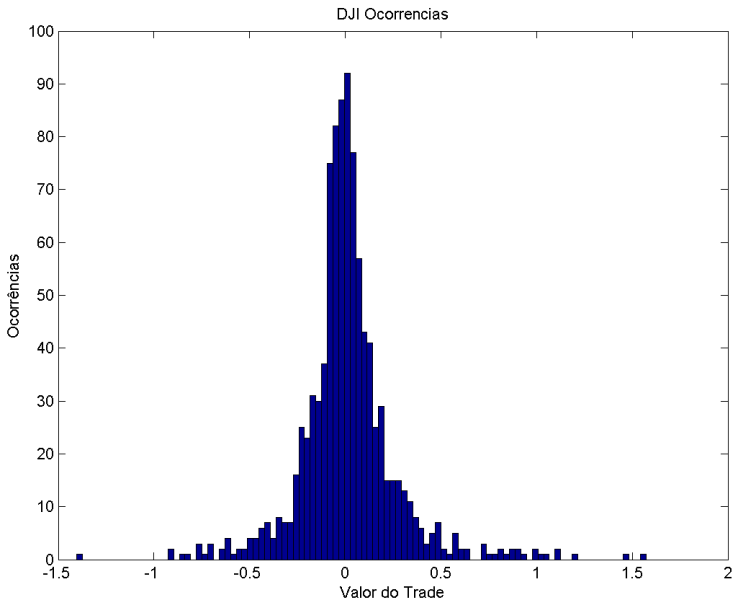

Figura 5.8: Distribuição do lucro no Sistema Inteligente - DJI 5m

5.2.3 Granularidade 10 m - PETR4

Foi testado o Sistema Inteligente competindo contra um Sistema Bollinger na entrada. Na saída, ambos consistiam em saídas baseadas em um sistema que utiliza as Bandas de Bollinger. Haviam disponíveis 18352 pontos disponíveis. 9176 destes foram utilizados apenas para treinamento enquanto 9176 foram utilizados para comparação.

Tabela 5.3: Relatório Comparativo de estratégias

Estratégia de entrada	Sistema Inteligente	Sistema clássico
Instrumento usado:	PETR4	PETR4
Granularidade:	10m	10m
Data de inicio:	13-Feb-2013	13-Feb-2013
Data de termino:	06-Dec-2013	06-Dec-2013
Lucro Total:	R$ 6.23	R$ 3.08
Lucro Bruto:	R$ 38.45	R$ 17.48
Prejuízo Bruto:	R$ -32.22	R$ -14.40
Fator de Lucro:	1.19	1.21
Número Total de Operações:	1228	213
Percentual de Vencedoras:	56.76 %	67.14 %
Operações Vencedoras:	697	143
Operações Perdedoras:	420	67
Lucro Médio Total:	R$ 0.01	R$ 0.01
Lucro Médio das Vencedoras:	R$ 0.06	R$ 0.12
Prejuízo Médio das Perdedoras:	R$ -0.08	R$ -0.21
Razão Média Vencedoras/Média Perdedoras:	0.72	0.57
Maior Operação Vencedora:	R$ 0.67	R$ 0.72
Pior Operação Perdedora:	R$ -1.67	R$ -1.67
Maior Número de Vitórias Consecutivas:	10	11
Maior Número de Derrotas Consecutivas:	7	4
Média de Tempo das Operações:	4.18	13.78
Média de Tempo das Operações Vencedoras:	2.55	7.68
Média de Tempo das Operações Perdedoras:	6.33	26.24
Sharpe Ratio:	0.042688	0.060673
Estrategia Buy and Hold no mesmo período:	R$ 0.54	R$ 0.45

A diante, seguem os gráficos do percentual de lucro acumulado de ambas as estratégias comparadas:

Figura 5.9: Comparação do Lucro Percentual Acumulado das Estratégias - PETR4 10m

A seguir, a Tabela Q:

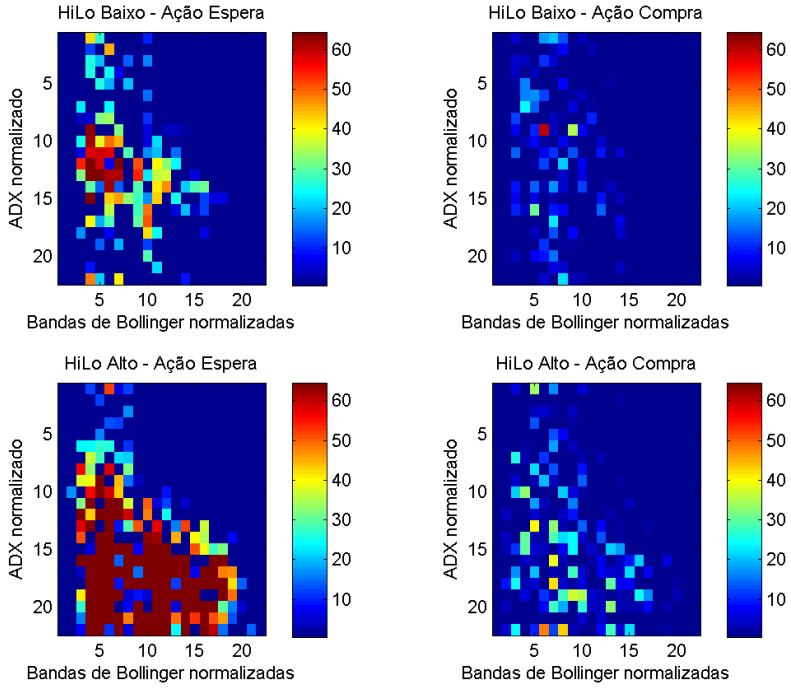

Figura 5.10: Tabela Q final do Sistema Inteligente - PETR4 10m

Por fim, apresentam-se dois gráficos indicando a distribuição do lucro obtido com os trades, um para cada estratégia.

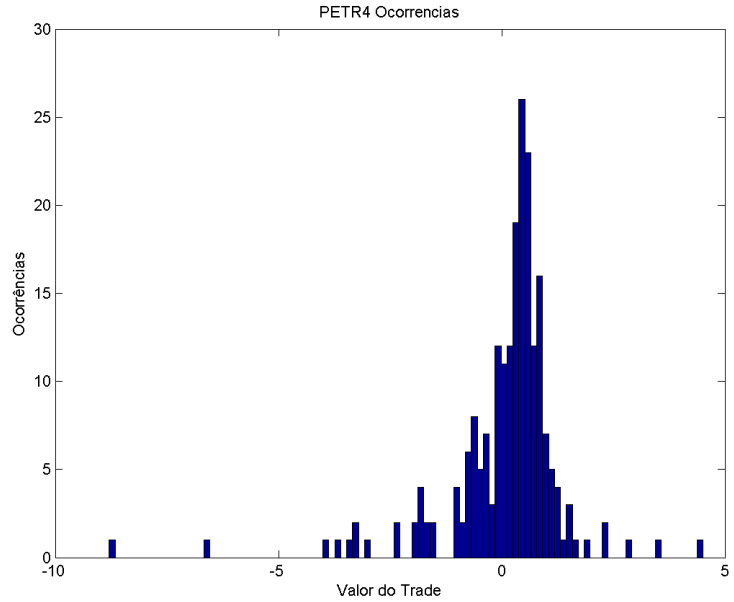

Figura 5.11: Distribuição do lucro na Estratégia Clássica - PETR4 10m

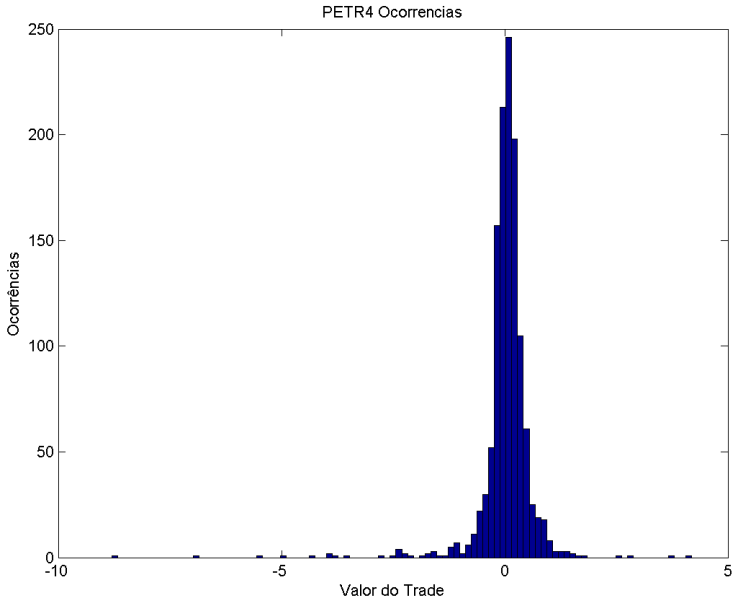

Figura 5.12: Distribuição do lucro no Sistema Inteligente - PETR4 10m

5.2.4 Granularidade 15 m - VALE5

Foi testado o Sistema Inteligente competindo contra um Sistema HiLo na entrada. Na saída, ambos consistiam em saídas baseadas em HiLo. Haviam disponíveis 12490 pontos disponíveis. 6245 destes foram utilizados apenas para treinamento enquanto 6245 foram utilizados para comparação.

Tabela 5.4: Relatório Comparativo de estratégias

Estratégia de entrada	Sistema Inteligente	Sistema clássico
Instrumento usado:	VALE5	VALE5
Granularidade:	15m	15m
Data de inicio:	15-Feb-2013	15-Feb-2013
Data de termino:	06-Dec-2013	06-Dec-2013
Lucro Total:	R$ -9.17	R$ -7.04
Lucro Bruto:	R$ 38.98	R$ 31.36
Prejuízo Bruto:	R$ -48.15	R$ -38.40
Fator de Lucro:	0.81	0.82
Número Total de Operações:	314	314
Percentual de Vencedoras:	37.26 %	31.21 %
Operações Vencedoras:	117	98
Operações Perdedoras:	191	209
Lucro Médio Total:	R$ -0.03	R$ -0.02
Lucro Médio das Vencedoras:	R$ 0.33	R$ 0.32
Prejuízo Médio das Perdedoras:	R$ -0.25	R$ -0.18
Razão Média Vencedoras/Média Perdedoras:	1.32	1.74
Maior Operação Vencedora:	R$ 1.55	R$ 1.76
Pior Operação Perdedora:	R$ -1.22	R$ -1.01
Maior Número de Vitórias Consecutivas:	5	5
Maior Número de Derrotas Consecutivas:	8	9
Média de Tempo das Operações:	16.47	9.55
Média de Tempo das Operações Vencedoras:	19.47	17.10
Média de Tempo das Operações Perdedoras:	14.70	6.12
Sharpe Ratio:	-0.072274	-0.067238
Estrategia Buy and Hold no mesmo período:	R$ -1.35	R$ -1.35

A diante, seguem os gráficos do percentual de lucro acumulado de ambas as estratégias comparadas:

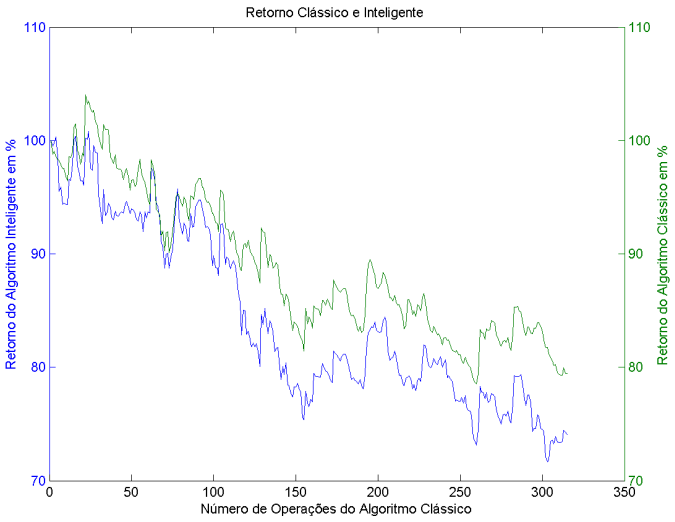

Figura 5.13: Comparação do Lucro Percentual Acumulado das Estratégias - VALE5 15m

A seguir, a Tabela Q:

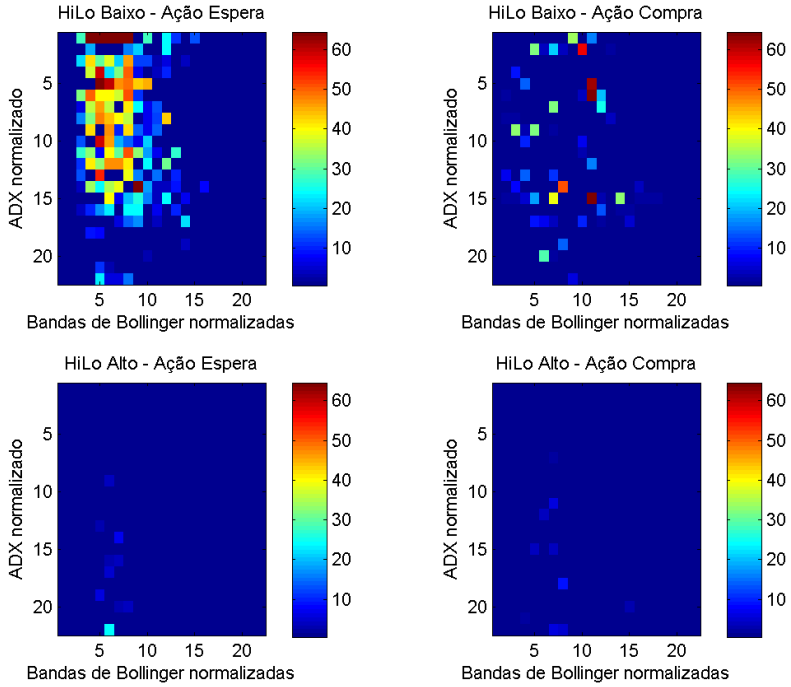

Figura 5.14: Tabela Q final do Sistema Inteligente - VALE5 15m

Por fim, apresentam-se dois gráficos indicando a distribuição do lucro obtido com os trades, um para cada estratégia.

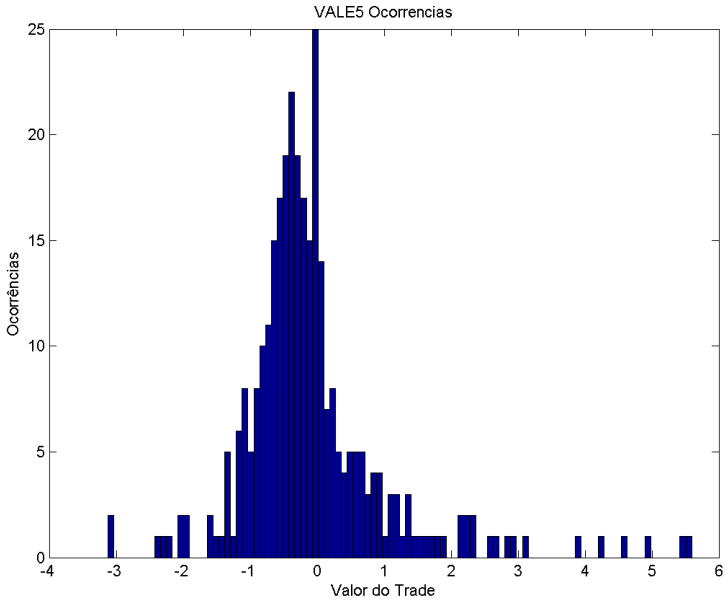

Figura 5.15: Distribuição do lucro na Estratégia Clássica - VALE5 15m

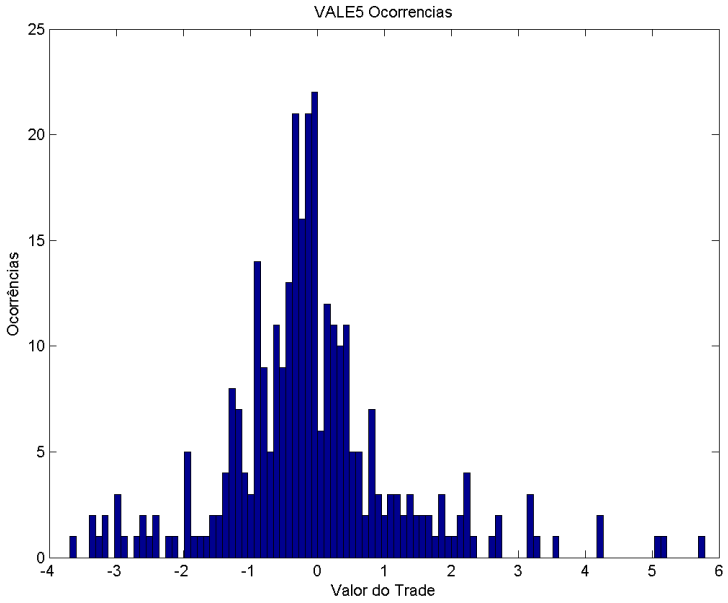

Figura 5.16: Distribuição do lucro no Sistema Inteligente - VALE5 15m

5.2.5 Granularidade 30 m - DOLFUT

Foi testado o Sistema Inteligente competindo contra um Sistema Bollinger na entrada. Na saída, ambos consistiam em saídas baseadas nas Bandas de Bollinger. Haviam disponíveis 7534 pontos disponíveis. 3767 destes foram utilizados apenas para treinamento enquanto 3767 foram utilizados para comparação

Tabela 5.5: Relatório Comparativo de estratégias

Estratégia de entrada	Sistema Inteligente	Sistema clássico
Instrumento usado:	DOLFUT	DOLFUT
Granularidade:	30m	30m
Data de inicio:	14-Feb-2013	14-Feb-2013
Data de termino:	06-Dec-2013	06-Dec-2013
Lucro Total:	R$ 43.10	R$ -44.36
Lucro Bruto:	R$ 1260.51	R$ 341.09
Prejuízo Bruto:	R$ -1217.41	R$ -385.45
Fator de Lucro:	1.04	0.88
Número Total de Operações:	665	57
Percentual de Vencedoras:	54.29 %	63.16 %
Operações Vencedoras:	361	36
Operações Perdedoras:	241	21
Lucro Médio Total:	R$ 0.06	R$ -0.78
Lucro Médio das Vencedoras:	R$ 3.49	R$ 9.47
Prejuízo Médio das Perdedoras:	R$ -5.05	R$ -18.35
Razão Média Vencedoras/Média Perdedoras:	0.69	0.52
Maior Operação Vencedora:	R$ 21.98	R$ 38.04
Pior Operação Perdedora:	R$ -64.92	R$ -54.85
Maior Número de Vitórias Consecutivas:	8	6
Maior Número de Derrotas Consecutivas:	8	4
Média de Tempo das Operações:	3.25	16.79
Média de Tempo das Operações Vencedoras:	2.09	9.53
Média de Tempo das Operações Perdedoras:	4.63	29.24
Sharpe Ratio:	0.008729	-0.044646
Estrategia Buy and Hold no mesmo período:	R$ 303.12	R$ 216.47

A diante, seguem os gráficos do percentual de lucro acumulado de ambas as estratégias comparadas:

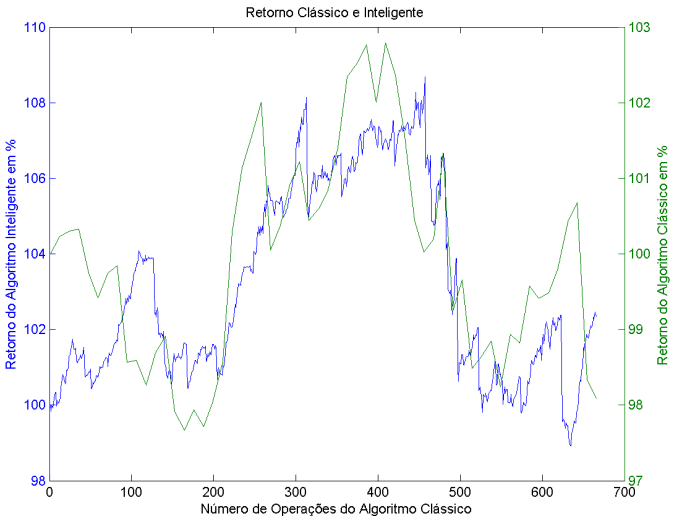

Figura 5.17: Comparação do Lucro Percentual Acumulado das Estratégias - DOLFUT 30m

A seguir, a Tabela Q:

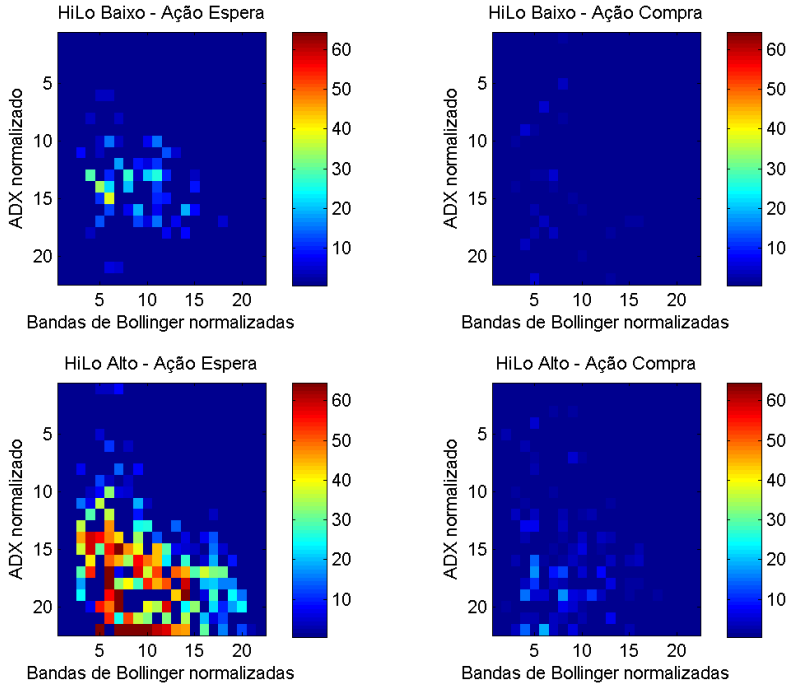

Figura 5.18: Tabela Q final do Sistema Inteligente - DOLFUT 30m

Por fim, apresentam-se dois gráficos indicando a distribuição do lucro obtido com os trades, um para cada estratégia.

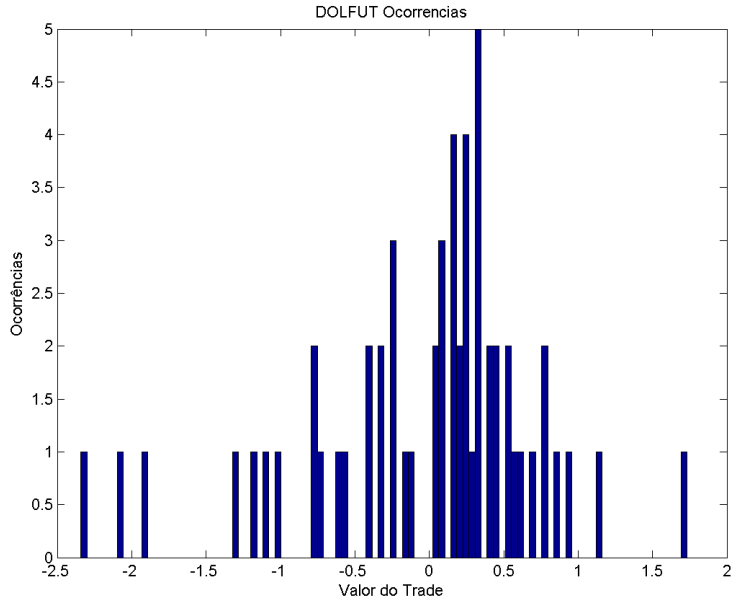

Figura 5.19: Distribuição do lucro na Estratégia Clássica - DOLFUT 30m

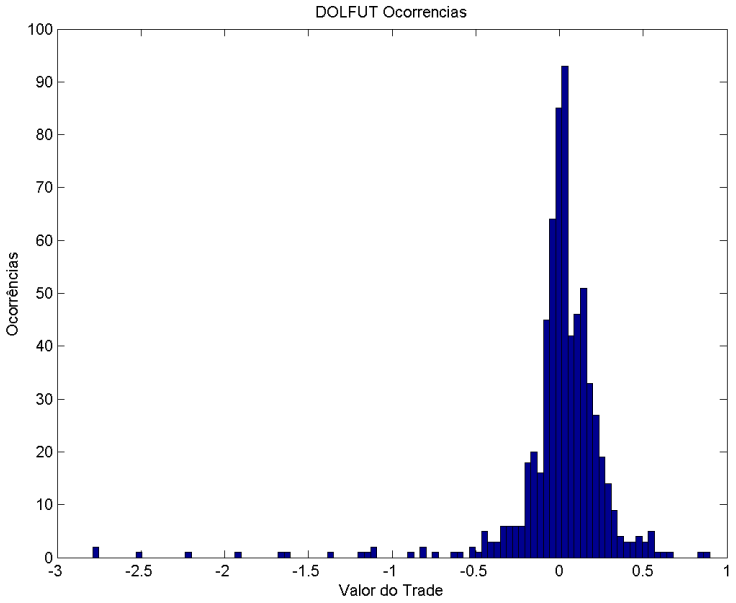

Figura 5.20: Distribuição do lucro no Sistema Inteligente - DOLFUT 30m

5.2.6 Granularidade 60 m - CMIG4

Foi testado o Sistema Inteligente competindo contra um Sistema Bollinger na entrada. Na saída, ambos consistiam em saídas baseadas no sistema Bollinger. Haviam disponíveis 1316 pontos disponíveis. 658 destes foram utilizados apenas para treinamento enquanto 658 foram utilizados para comparação

Tabela 5.6: Relatório Comparativo de estratégias

Estratégia de entrada	Sistema Inteligente	Sistema clássico
Instrumento usado:	CMIG4	CMIG4
Granularidade:	60m	60m
Data de inicio:	14-Aug-2013	14-Aug-2013
Data de termino:	29-Nov-2013	29-Nov-2013
Lucro Total:	R$ 1.93	R$ 0.91
Lucro Bruto:	R$ 6.41	R$ 2.29
Prejuízo Bruto:	R$ -4.48	R$ -1.38
Fator de Lucro:	1.43	1.66
Número Total de Operações:	134	13
Percentual de Vencedoras:	59.70 %	69.23 %
Operações Vencedoras:	80	9
Operações Perdedoras:	47	4
Lucro Médio Total:	R$ 0.01	R$ 0.07
Lucro Médio das Vencedoras:	R$ 0.08	R$ 0.25
Prejuízo Médio das Perdedoras:	R$ -0.10	R$ -0.34
Razão Média Vencedoras/Média Perdedoras:	0.84	0.74
Maior Operação Vencedora:	R$ 0.27	R$ 0.43
Pior Operação Perdedora:	R$ -0.62	R$ -0.46
Maior Número de Vitórias Consecutivas:	8	5
Maior Número de Derrotas Consecutivas:	4	1
Média de Tempo das Operações:	2.99	13.92
Média de Tempo das Operações Vencedoras:	1.84	8.44
Média de Tempo das Operações Perdedoras:	4.70	26.25
Sharpe Ratio:	0.118359	0.229231
Estrategia Buy and Hold no mesmo período:	R$ -0.49	R$ -1.10

A diante, seguem os gráficos do percentual de lucro acumulado de ambas as estratégias comparadas:

Figura 5.21: Comparação do Lucro Percentual Acumulado das Estratégias - CMIG4 60m

A seguir, a Tabela Q:

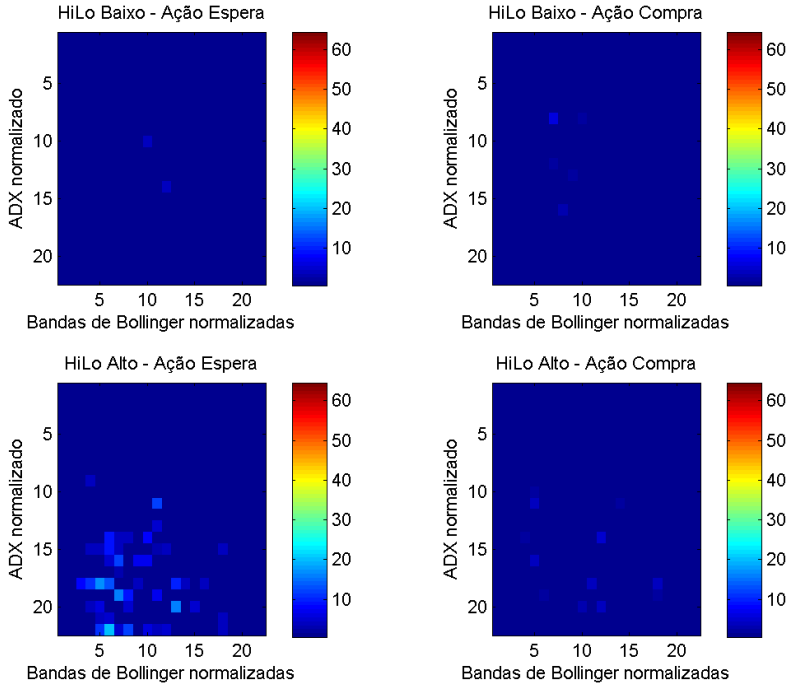

Figura 5.22: Tabela Q final do Sistema Inteligente - CMIG4 60m

Por fim, apresentam-se dois gráficos indicando a distribuição do lucro obtido com os trades, um para cada estratégia.

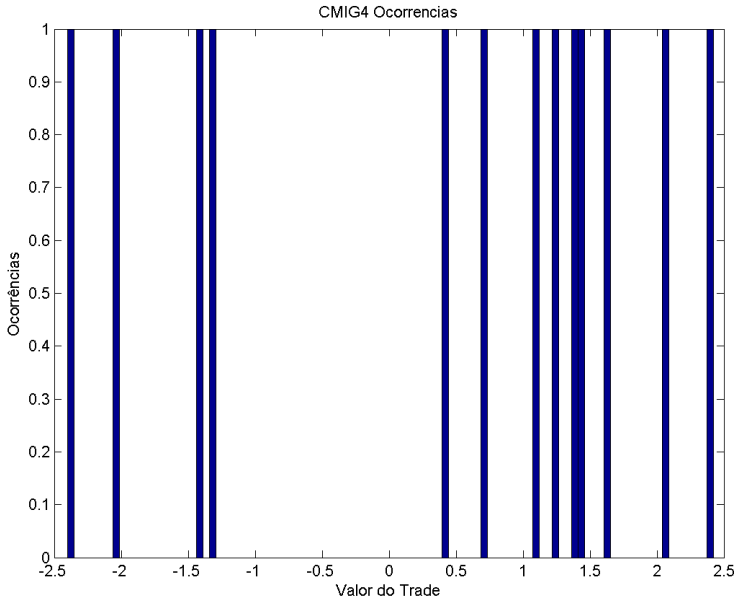

Figura 5.23: Distribuição do lucro na Estratégia Clássica - CMIG4 60m

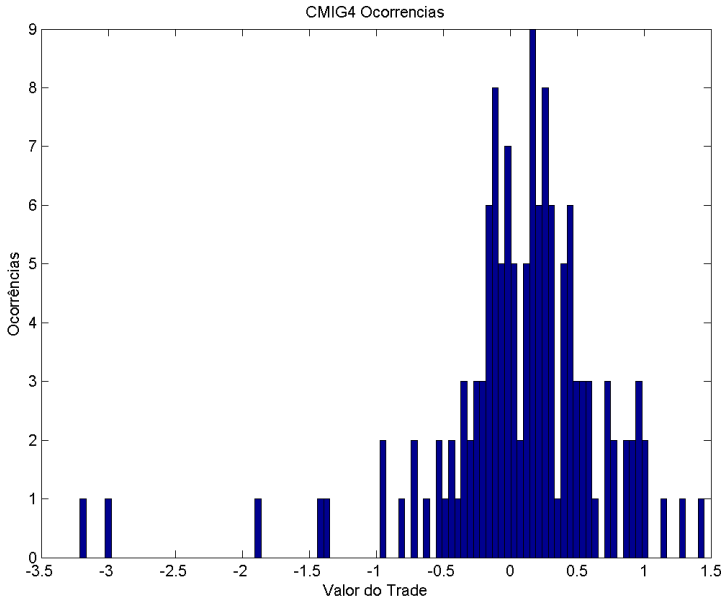

Figura 5.24: Distribuição do lucro no Sistema Inteligente - CMIG4 60m

5.3 Análise dos resultados

Primeiramente, é interessante ressaltar que, para dados de minuto a minuto, foi testado o algoritmo com 64 mil períodos amostrados. Considerando que a tabela tinha 1760 pontos e na metade dos períodos, aproximadamente, o algoritmo usado é o clássico para sair do mercado, temos em média $\frac{32000}{1760} \approx 18$ visitas a cada estado. Este número de visitas embora pareça pequeno, pelos resultados obtidos, parece fornecer uma boa aproximação. Além disso, após passado por todos os períodos, ao visualizar a tabela Q fica claro que alguns estados nunca chegam a ser realmente visitados[2], que faz com que outros sejam visitados bem mais que 18 vezes.

Em contrapartida, nas simulações efetuadas utilizando-se dados com granularidade de 30 minutos e 60 minutos, o número de pontos disponível era bastante inferior. Para o de 60 minutos, foi usado menos de 1400 períodos. Menos que o número de estados possíveis. Deste modo, a tabela Q possuiu um número bastante elevado de estados que nunca foram visitados. Além disso, mesmo os estados visitados não possuíram um número de visitas grande o bastante para que a lei dos grandes números pudesse ser aplicada, e o retorno seja adequadamente estimado. Isso também se mostra presente nas figuras 5.23, 5.24, 5.19 e 5.20. Vê-se que o número de compras efetuado não foi grande o suficiente nem para aproximar de maneira adequada a distribuição amostral do lucro.

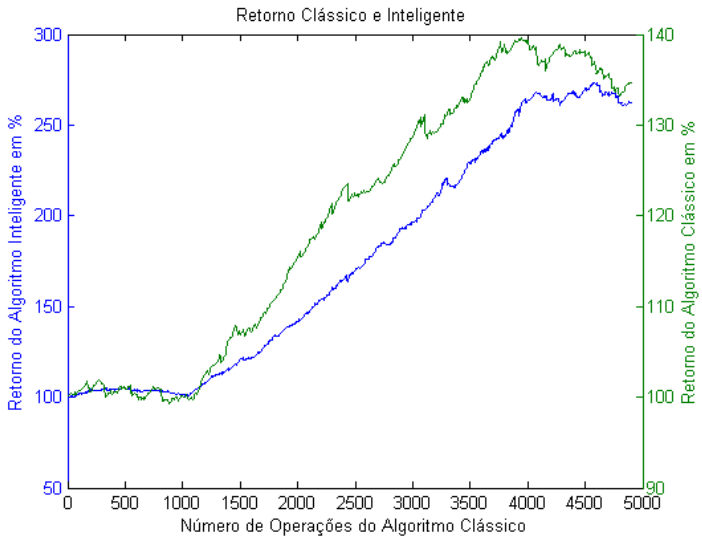

Figura 5.25: Comparação entre Retorno do algoritmo clássico e inteligente no IBOV de 1 minuto usando o Sistema Bollinger

Nas figuras 5.2, 5.6, 5.10, 5.14, 5.18, 5.22 pode-se notar o valor dos pares Estado-Ação de um experimento particular. É interessante notar que é consenso o fato de que esperar geralmente é melhor que comprar, com exceção de alguns pontos bem específicos. Ainda assim, fornecendo dados suficientes, o algoritmo inteligente consegue identificar muito mais pontos favoráveis de compra do que os algoritmos clássicos, como pode-se ver pelo número de operações maior e sem prejudicar o Fator de Lucro. Inclusive, com dados suficientes, os algoritmos inteligentes em praticamente todos

[2] A tabela Q é inicializada com valores inteiros e a cada atualização os valores adicionados são de ponto flutuante. Ao final do experimento, os pontos da tabela com valores ainda inteiros são estados nunca visitados.

os aspectos analisados sempre se mostraram superiores, ou pelo menos tão bons quanto, em relação ao algoritmo clássico.

Infelizmente, um ponto fraco deste sistema é a necessidade de um número muito grande de amostras para conseguir competir com um algoritmo clássico tradicional. Entretanto, em posse deste número mínimo de amostras (algo entre 14 e 20 mil períodos), o sistema se mostrou capaz de vencer todos os algoritmos clássicos testados.

Outro ponto extremamente interessante para se ressaltar é que o formato das curvas do dinheiro acumulado é extremamente semelhante no caso do algoritmo clássico e do inteligente. Na figura 5.25, assim como nas figuras 5.21, 5.17, 5.13, 5.9, 5.5 e 5.1, nota-se claramente que o formato é praticamente o mesmo, com exceção da ordem de grandeza e das amplitudes das variações do sinal, e isso foi visto em todos os gráficos. Isso se deve ao fato de ambos usarem a mesma saída. Isso traz análises incríveis. Aparentemente, independente da entrada escolhida, o formato do gráfico é dado pela estratégia de saída. Ou seja, se, no *Backtest*[3] de um *trading system*, o retorno está tendo um comportamento ruim, provavelmente será necessário trocar a estratégia de saída. Com a estratégia de entrada é possível apenas amenizar esse problema. A mudança da estratégia de entrada clássica por uma inteligente diminuiu a volatilidade proporcional do retorno e aumentou seu crescimento, mas ainda mostra as mesmas variações, como se isso fosse um timbre da estratégia de saída. Entretanto, isso também pode ser um reflexo de que o sinal de entrada do algoritmo inteligente é composto pelos mesmos indicadores das estratégias testadas.

Importante ressaltar que **ganhos passados não garantem ganhos futuros**. Logo, todos esses testes são quesitos mínimos, mas não suficientes para se ter uma estratégia vencedora.

[3]Teste de uma estratégia de compra e venda ou de um modelo preditivo usando dados históricos.

Capítulo 6

Conclusões

Neste trabalho, desejou-se demonstrar que o uso de uma estratégia baseada no aprendizado por reforço é capaz de superar um conjunto de Benchmarks envolvendo estratégias ativas e passívas na operação de instrumentos da Bolsa de Valores. Para isso, construiu-se um exemplar de um sistema SARSA e sob ele foi feita uma bateria de testes de modo a justificar seu uso em lugar de estratégias mais simples. De modo geral, dadas as restrições do Sistema Inteligente (ou seja, a necessidade de um número elevado de pontos para treinamento adequado) ele mostrou-se bastante apropriado em sua tarefa, superando todas as contrapartes clássicas contra a qual foi testado.

Alguns detalhes, no entanto, merecem ser discutidos. Embora sempre seja possível observar os valores da tabela Q estimados num dado momento, o sistema inteligente atua de maneira independente. Em outras palavras, ele funciona como uma caixa preta, simplesmente indicando momentos de compra sem uma justificativa mais embasada. Os algoritmos clássicos, embora não superem o Sistema Inteligente, possuem essa vantagem de se adequar a lógica do investidor.

Deste modo, caso se opte pelo uso do Sistema Inteligente para investimentos reais, deve-se estar pronto para operar os instrumentos de uma maneira que pode aparentar ser contra-intuitiva. E tais operações precisam ser feitas de maneira disciplinada, pois sem isso o sistema será contaminado pelo julgamento de seu usuário, e perderá sua validade estatística.

Um ponto interessante a ser discutido também é a codificação de estados. É interessante notar que a estratégia inteligente, apenas por experiência, atribuiu significado aos indicadores. Note que nas tabelas Q é possível ler e interpretar significados para os indicadores utilizados como estados. Tal significado se aproximou bastante do original, pelo qual o indicador é de fato utilizado.

Um possível trabalho futuro consistiria em estudar o efeito dos estados no sistema. Outras codificações de estado são possíveis, com outros indicadores, ou mesmo com as próprias informações de preço e volume do *Home Broker*. Entretanto, caso se amplie de maneira elevada o número de estados, se torna necessário modificar o algoritmo utilizado para abarcar estes estados. A tabela Q seria substituída por uma função Q, construída por um aproximador de funções qualquer. Deste modo, é possível construir sistemas com um maior número de estados e ainda assim empregá-los sem necessitar de um número exorbitante de pontos de amostra.

Algo que não foi discutido no trabalho, mas que cabe também a um trabalho futuro é o Gerenciamento Monetário (*Money Management*). Em todos os testes, supôs-se um investimento de capital fixo. Entretanto, em dados momentos no tempo existem perspectivas de lucro maiores associadas a riscos menores, indicando que um volume de capital maior poderia ser investido sob um risco menor de perda.

Outro ponto interessante que fica para um trabalho futuro é a elaboração de um Sistema Inteligente de saída. Conforme foi dito, embora o Sistema Inteligente de entrada, da forma como foi construído, garanta retornos mais estáveis, ou seja, menos sujeitos a volatilidade, uma boa estratégia de saída seria capaz de ampliar de maneira significativa o retorno esperado médio. Deste modo, cabe uma investigação a respeito do efeito real da estratégia de saída e se uma estratégia inteligente de saída é capaz de superar uma clássica.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] INVESTOPEDIA. Dec 2013. Disponível em: <http://www.investopedia.com/>.

[2] ELDER, A. *Trading for a Living.* [S.l.: s.n.], 1999.

[3] BACHELIER, L. *Théorie de la spéculation.* [S.l.]: Gauthier-Villars, 1900.

[4] FAMA, E. F. Efficient capital markets: A review of theory and empirical work*. *The Journal of Finance*, Blackwell Publishing Ltd, v. 25, n. 2, p. 383–417, 1970. ISSN 1540-6261. Disponível em: <http://dx.doi.org/10.1111/j.1540-6261.1970.tb00518.x>.

[5] MANDELBROT, B. *The (Mis)Behavior of Markets.* [S.l.: s.n.], 2004.

[6] ELDER, A. *Come into my Trading Room.* [S.l.: s.n.], 1996.

[7] GRANVILLE, J. *Granville's New Key to Stock Trading.* [S.l.: s.n.], 1963.

[8] WILLIAMS, L. R. *How I made one million dollars last year trading commodities.* [S.l.]: Windsor Books (Brightwaters, NY), 1979.

[9] WILDER, J. W. *New Concepts in Technical Trading Systems.* [S.l.: s.n.], 1978.

[10] KRAUSZ, R. *A WD Gann Treasure Discovered: Simple Trading Plans for Stocks & Commodities.* [S.l.]: Geometric Traders Institute, 1996.

[11] SHARPE, W. F. Mutual fund performance. *The Journal of Business*, The University of Chicago Press, v. 39, n. 1, p. pp. 119–138, 1966. ISSN 00219398. Disponível em: <http://www.jstor.org/stable/2351741>.

[12] SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction.* [S.l.: s.n.], 1998.

[13] RUMMERY, G. A.; NIRANJAN, M. *On-Line Q-Learning Using Connectionist Systems.* [S.l.], 1994.

[14] SINGH, S.; SUTTON, R. S. Reinforcement learning with replacing eligibility traces. In: *MACHINE LEARNING.* [S.l.: s.n.], 1996. p. 123–158.

[15] WATKINS, C. *Learning from Delayed Rewards.* Tese (Doutorado) — Cambridge, 1989.

[16] TOMASINI, E.; JAEKLE, U. *Trading Systems: A New Approach to System Development and Portfolio Optimisation.* [S.l.]: Harriman House Limited, 2009.

[17] NISON, S. *Japanese Candlestick Charting Techniques*. [S.l.: s.n.], 1991.

ANEXOS

I. GRÁFICOS

Esta seção explicará de maneira breve o sistema de gráficos utilizados por *traders* para exibir as informações do mercado disponíveis. Em particular, será explicado o sistema de gráficos denominado *Japanese Candlestick*.

No mercado de ações, muita coisa acontece com um papel no intervalo de um dia. Em dias mais agitados, mesmo no intervalo de um minuto pode haver muita informação, de modo que essa informação jamais ficaria corretamente representada por meio de apenas um ponto. Assim, costumeiramente existem 4 parâmetros principais de preço que caracterizam um intervalo:

- Valor de Abertura - Preço sob o qual foi realizado o primeiro (ou os primeiros) negócio do período.
- Valor Máximo - Maior preço no período pelo qual o papel foi negociado
- Valor Mínimo - Menor preço no período pelo qual o papel foi negociado
- Valor de Fechamento - Preço sob o qual foi realizado o último (ou os últimos) negócios do período

De posse destas informações cada ponto no gráfico deverá representar simultaneamente estas 4 informações. Para isso, é utilizado o sistema *Japanese Candlestick* (ou apenas *Candlestick*). Tal sistema foi inventado no século XVIII pelos negociantes de arroz do Japão. Ele foi trazido para o Ocidente por meio do livro de Steve Nison, "'Japanese Candlestick Charting Techniques"'[17].

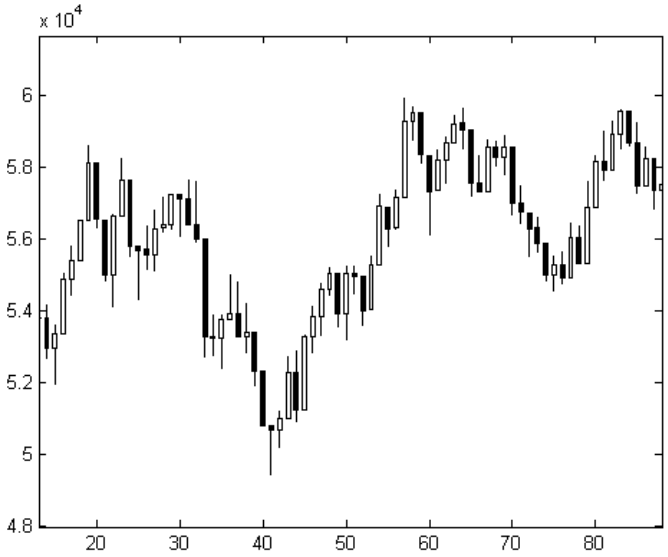

Figura I.1: Gráfico em *Candlesticks*

Abaixo vemos um exemplo de dois *candlesticks* em particular. Cada *candlestick* representa um período de negociação. Ele possui duas partes, o corpo e as sombras. O corpo representa os preços de abertura e fechamento daquele período particular. A sombra superior indica valor máximo do papel negociado naquele período, enquanto a sombra inferior, por sua vez, indica o valor mínimo.

A distinção entre a abertura e o fechamento se dá pela cor do *candle*. Um *candle* branco indica uma ascensão de preços. Deste modo a sua leitura se dá da seguinte maneira: o canto inferior esquerdo indica seu preço de abertura, enquanto seu canto superior direito indica o preço de fechamento. Já um *candle* preto indica uma queda de preços. Portanto a sua leitura é: canto superior direito representando o preço de abertura e canto inferior esquerdo como fechamento.

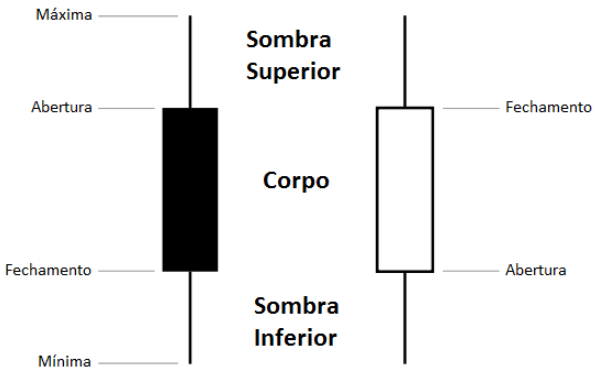

Figura I.2: *Candlesticks* possíveis

A principal vantagem do uso de *candlesticks* é o auxílio visual que ele fornece. Em um gráfico com tendencias de subida, a maior parte dos *candlesticks* será branco, enquanto um mercado em queda terá bastantes *candlesticks* pretos.

Além disso, outras informações cuja visualização se faz importante é o volume e a quantidade de negócios realizada em um dado período em particular. Tais informações são usualmente mostradas em gráficos com formato de barras, alinhadas com os *candlesticks*, conforme a figura abaixo:

Figura I.3: *Candlesticks* + Volume, extraído do programa InvestCharts

Além do sistema de *candlesticks*, outra maneira bastante comum de se mostrar tendências associadas ao mercado de ações é feito com o uso de um gráfico com linha contínua. Nessa linha contínua, cada período de um dia é representado por apenas um ponto, e pontos em dias consecutivos são unidos com uma linha reta. Em geral, é costume utilizar sempre o preço de fechamento associado àquele dia para plotar informações desta maneira, uma vez que o preço de fechamento é, dos quatro valores disponíveis, aquele que possui uma maior estabilidade. Isto se dá porque, segundo o Dr. Alexander Elder[2], durante o período de fechamento os agentes financeiros mais experientes fazem, utilizando a informação adquirida no decorrer do dia, as suas operações.

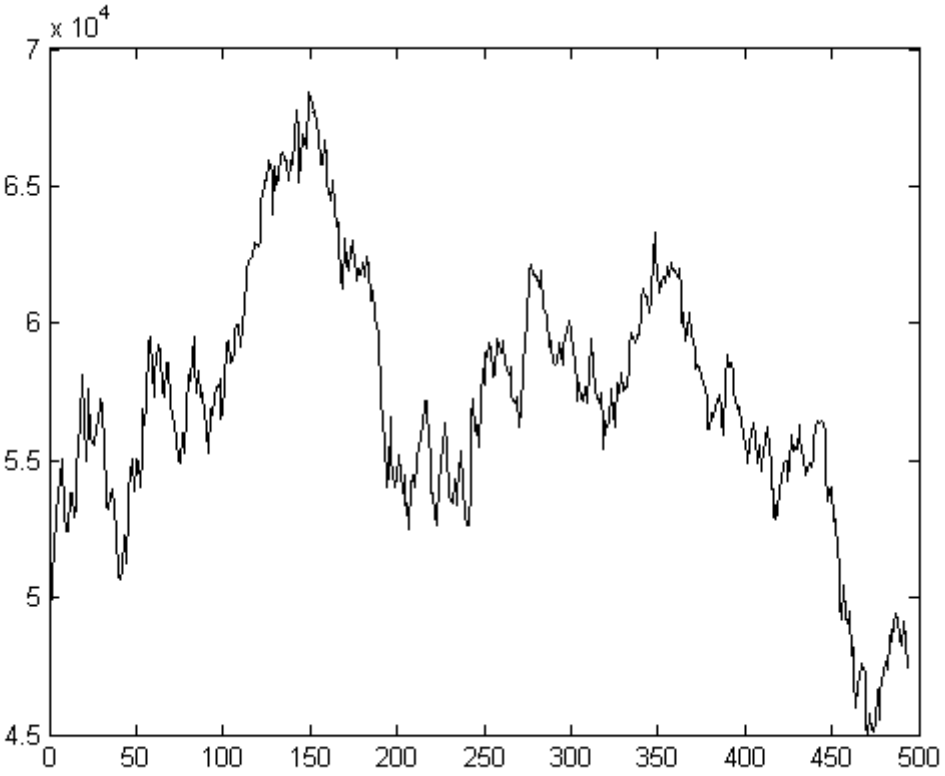

Figura I.4: Gráfico em linha do papel IBOV

II. DICIONÁRIO DE JARGÕES

Uma vez que este trabalho se encontra na ponte entre dois universos distintos (Engenharia Financeira e Aprendizagem por Reforço), pensou-se ser relevante um dicionário para familiarizar o leitor acostumado a apenas um destes universos, mas para o qual o outro é novidade. A primeira parte deste capítulo tratará dos jargões financeiros, enquanto que a segunda tratará dos jargões característicos do Aprendizado por Reforço.

II.1 Jargões Financeiros

Instrumento

Qualquer contrato que dê origem a um ativo financeiro para uma entidade e a um passivo financeiro ou instrumento de capital próprio para outra. Exemplos: Ações da Petrobras, Contratos Futuros de alguma *commodity*, etc.

Ativo financeiro

Um ativo do qual se deriva valor. Exemplos: ações, depósitos no banco, contratos futuros, títulos públicos e privados, etc.

Bid-Ask Spread

Distância existente entre a oferta de compra mais cara e a oferta de venda mais barata.

Stop Loss

Ordem de mercado que visa limitar as perdas do agente financeiro através de uma operação de compra ou venda, assim que o preço desce abaixo do preço do *Stop*.

Stop Gain

Ordem de mercado que visa realizar os ganhos do agente financeiro antes de uma reversão de tendência através de uma operação de compra ou venda, assim que o preço sobe acima do preço do *Stop*.

Trailing Stop

É idêntico ao *Stop Loss*, com a exceção de que ele se move com os preços em apenas uma direção a fim de minimizar as perdas, ao mesmo tempo que protege os ganhos.

Agente Financeiro

Agentes financeiros são entidades (pessoas físicas ou companhias) que fazem a gestão de seu patrimônio fazendo uso do Mercado Financeiro.

Taxa de Corretagem

Taxa cobrada por corretoras de valores ao se fazer uma operação no Mercado Financeiro, tendo seu valor usualmente tabelado pelas corretoras.

Touro

Animal que ataca de baixo para cima, representa a força que faz os preços aumentarem. Figura costumeiramente associada aos compradores.

Urso

Animal que ataca de cima para baixo, representa a força que faz os preços diminuírem. Figura costumeiramente associada aos vendedores.

Short

Entrar vendido; entrar como Urso; pegar emprestado, vender para comprar a um preço mais baixo.

Long

Entrar comprado; entrar como touro; comprar com intenção de vender mais caro.

Entrar comprado

Ver long.

Entrar vendido

Ver short.

Slippage

Diferença entre o preço esperado de uma operação e o preço que a operação foi realmente executada. *Slippage* geralmente acontece durante períodos de alta volatilidade e também quando grandes ordens são executadas e não existem suficientes agentes dispostos a pagar o mesmo preço por esta ordem.

Market Order

É uma ordem de compra ou venda executada imediatamente a preços atuais de mercado.

Limit Order

É uma ordem de compra limitada a um preço máximo determinado ou uma ordem de venda limitada a um preço mínimo determinado.

Operação

Uma operação, também chamado de *trade*, acontece quando a oferta de compra é maior ou igual à oferta de venda, vencendo assim o *bid/ask spread* e ocorrendo a troca de dinheiro por instrumentos financeiros.

Backtest

Teste de uma estratégia de compra e venda ou de um modelo preditivo usando dados históricos.

Sharpe Ratio

Indicador desenvolvido por William Sharpe, cujo objetivo é ponderar o lucro esperado por um determinado ativo contra o risco que se corre ao investir nesse ativo.

II.2 Jargões de Aprendizado por Reforço

Ação

Decisão do agente de aprendizado por reforço. Neste trabalho as únicas ações disponíveis são comprar ou esperar.

Agente

É a parte do algoritmo que tem memória e faz a tomada de decisão. Todo o resto é considerado ambiente.

Ambiente

É tudo que não é agente. A oscilação dos instrumentos, o cálculo da recompensa, os estados e os processos dele são todos parte do ambiente.

Episódio

Episódios são qualquer forma de interação repetida sob a qual o Agente está submetido.

Atualização *Online/Offline*

Em um treinamento *online*, as atualizações são feitas durante o episódio, assim que o incremento é computado. Por outro lado, em um treinamento *offline* os incrementos são acumulados à parte e não são usados para se atualizar as estimativas de valor até o final do episódio.

Retorno

A soma das recompensas recebidas a longo prazo. Pode se referir também a retorno descontado.

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{II.1}$$

Recompensa

Sinal do Ambiente passado ao Agente, que representa a proximidade do seu objetivo.