

Robustez como Critério de Decisão em Modelos Financeiros Sequenciais

Duarte Trigueiros

Informática

© Copyright 1991
by
Duarte Trigueiros

Esta copia é fornecida sob condição de que quem a consultar reconhece que os direitos de autor permanecem da posse do autor e que nenhuma citação deste trabalho, nem nenhuma informação derivada dele, poderá ser publicada sem a prévia autorização escrita do autor.

Introdução

Este trabalho tem por âmbito os projectos de investimento. Ele mostra como a capacidade redutora de complexidade dos algoritmos para indução de regras tem uma promissora aplicação na interpretação de estruturas sequenciais de decisão financeira onde se pretenda, não apenas a descoberta da decisão óptima, mas também a quantificação da sua robustez.

Este trabalho insere-se na linha de investigação que procura desenvolver sistemas capazes de descobrir relações subjacentes a experiência acumulada em bases de dados. A indução de regras a partir de observações é uma das técnicas usadas em sistemas capazes de aprender. Porém, em vez de se centrar na aprendizagem automática, este trabalho mostrará que tais ferramentas podem ter outros usos úteis.

O centro de referência deste trabalho é a decisão sequencial em projectos de investimento, i.e, o problema da decisão estruturada hierarquicamente ao longo do tempo. Far-se-á primeiro uma descrição de anteriores tentativas para abordar o problema da robustez. Em seguida criticar-se-ão os usos não-específicos que se têm feito das técnicas de indução de regras, com o fim de clarificar conceitos importantes para o problema em estudo. Por fim, explicar-se-á como o uso adequado do algoritmo ID3 [27], ou outro semelhante, é capaz de eliminar a redundância e assim reduzir a complexidade em modelos de decisão sequencial, originando medidas objectivas de robustez.

O saldo final será o de mostrar que algoritmos como o ID3 ou o C5 não são adequados para todos os tipos de modelação estatística em finanças e portanto não devem ser usados para esse fim. Porém, estes mesmos algoritmos mostram-se promissores como pós-processadores de árvores de decisão. Eles são capazes de hierarquizar o peso causal de cada uma das variáveis do problema. Este facto permite ao gestor ajuizar, não apenas da optimalidade de cada possível percurso, mas também do grau de domínio que tem sobre o processo.

Para compreender porquê podem estes resultados serem úteis em projectos de investimento, vale a pena recordar aquilo que um gestor obtém actualmente de uma árvore de decisão. Ele obtém um percurso óptimo e um valor correspondente. Não obtém informação acerca da robustez dessa particular solução óptima. Nada lhe é dito sobre a flexibilidade de que dispõe para introduzir modificações durante o processo sem comprometer demais o resultado.

Um óptimo é um ponto estático. Para os gestores, o problema é também dinâmico: Precisam de calcular as consequências de um ajustamento de planos a meio do processo ou de uma mudança nas condições exteriores.

Parece-nos, pelo estudo realizado, que os algoritmos para indução de regras são capazes de revelar algumas características dinâmicas dos problemas sequenciais de decisão financeira. E se mais não fizessem, a simples eliminação de redundância já seria interessante. É sabido com que facilidade as árvores de decisão se transformam em bosques intrincados.

Terminologia: A expressão “árvore de decisão” é usada em finanças para designar estruturas hierárquicas de decisão sequencial. Estas estruturas estão geralmente associadas a projectos de investimento. Em Inteligência Artificial, esta mesma expressão costuma significar qualquer hierarquia lógica, mesmo quando não tem nada a ver com decisões. No presente estudo, uma “árvore de decisão” conserva o seu sentido original. As estruturas lógicas são aqui chamadas “árvores de regras”.

Em modelação estatística não é frequente o uso do termo “classe” como sinónimo de variável explicada. Aqui, usar-se-á a seguinte terminologia: Uma variável explicadora é um “atributo”. Uma variável explicada é um “desenlace”. O valor previsto por um modelo é um desenlace “esperado” ou um “acontecimento”. Pensa-se que o uso destes termos facilitará a compreensão dos assuntos e não se afasta muito do sentido com que são usados em modelação estatística de variáveis nominais.

O termo “classe”, mantém o significado original, o de uma partição ou domínio para contagem de frequências.

Agradecimentos

Em primeiro lugar, o autor deseja manifestar o seu apreço e agradecimento ao Professor Augusto Afonso de Albuquerque, cujas qualidades de liderança e visão estão na raiz deste e outros estudos.

Este trabalho foi realizado durante a estadia do autor na Universidade de East Anglia, paralelamente aos estudos que também aí realizou tendo por objecto outros aspectos da aquisição automática de conhecimentos a partir de experiência passada existente em bases de dados contabilísticos e financeiros. O autor está agradecido ao Professor Robert H. Berry pelas interessantes discussões a que este tema deu azo.

O Professor Carlos Beltran Almeida reviu esta versão em português. O autor está muito agradecido pelas sugestões e correcções por ele sugeridas. A Professora Elisabete Reis também leu este estudo. O autor agradece as suas valiosas críticas e sugestões.

Por último, o autor deseja manifestar a sua gratidão ao ISCTE, Instituto de Ciências do Trabalho e da Empresa, pela cobertura financeira deste e outros estudos.

Índice

Introdução	iii
Agradecimentos	v
1 Robustez como Critério de Decisão	1
1.1 Introdução à Noção de Robustez	2
1.2 Discussão	5
1.3 Uma Nova Ideia de Robustez	8
1.4 Sumário	8
2 O Algoritmo ID3 para Indução de Regras	9
2.1 A Quantidade de Informação em Estatística	9
2.1.1 O Algoritmo ID3	11
2.1.2 Outros Algoritmos	12
2.2 A Diferença Entre Modelar e Ordenar	13
2.3 Crítica ao Uso da Indução de Regras em Estatística	15
2.3.1 Generalização	15
2.3.2 Indução de Regras e Estrutura	17
2.3.3 Como Podar Árvores de Regras?	21
2.4 Indução de Regras e Modelos Usuais em Finanças	21
2.4.1 Problemas de Medição	22
2.4.2 Probabilidades e Correlações A-Priori	23
2.5 Sumário	24
3 Indução de Regras e Robustez	25
3.1 O ID3 em Modelos Sequenciais	25

3.1.1	Estruturas Concordantes	25
3.1.2	Quantificação da Causalidade	26
3.1.3	Como Aplicar o ID3 a Árvores de Decisão	30
3.2	Quantificando a Robustez	32
3.2.1	A Árvore de Decisão para “Prism Paints Inc.”	32
3.2.2	Discussão dos Resultados	35
3.3	Sumário	41
4	Conclusões	42

Capítulo 1

Robustez como Critério de Decisão

Os sistemas de suporte à decisão têm procurado até agora apresentar a solução *ótima* para problemas de Gestão. Por definição, essa solução deveria ser a melhor de entre as possíveis. Porém, tem provado, com demasiada frequência, ser inaceitável por parte dos gestores.

As razões para esta falha são múltiplas. Uma das mais discutidas na literatura é a pobreza de especificação de objectivos e critérios, devida em grande parte à impossibilidade de quantificar metas sentidas como importantes pelos gestores. Em outras palavras, os gestores sentem que o modelo não contempla aspectos importantes.

Os modelos usuais limitam-se a maximizar os lucros esperados ou a minimizar os custos, ou ainda a descobrir valores extremos de uma qualquer função-objectivo que possa ser tratada como sinónimo de lucros ou custos. Este modo de trabalhar reduz situações genuinamente multi-objectivo, mas onde importantes aspectos da decisão não são quantificáveis, a simples problemas de optimização.

Acresce que a decisão de investir tem geralmente profundas implicações estratégicas. Aqui, uma simples maximização de lucros, mesmo quando o modelo é pretensamente muito elaborado e abrangente, não irá obrigatoriamente conduzir à melhor solução: Há objectivos relevantes, como o ganho de certas posições e opções a longo prazo, difíceis de encaixar numa problemática do “ótimo”, e no entanto decisivos.

Existem porém outros tipos de razões para o desencanto com que certos sectores acolhem os modelos baseados no ótimo. Uma delas é a consciência de que tais modelos só são efectivos nos casos onde uma decisão conduz a acções repetitivas, uma vez que os valores ótimos só são ótimos “em média”, isto é, são valores esperados. Ora as mais importantes decisões que os gestores enfrentam costumam ser únicas, não repetitivas.

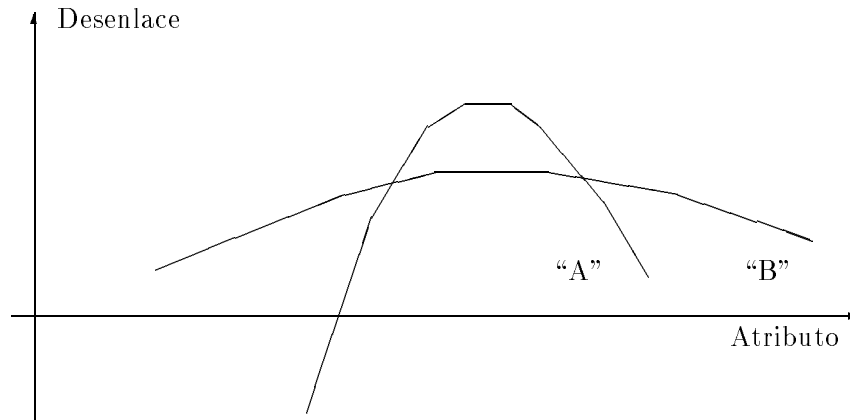


Figura 1: Visão simplificada de dois tipos de ótimos exibindo diferente robustez. A relação “B” mostra um valor ótimo robusto. “A” é um ótimo frágil.

Abordaremos de novo o problema das decisões únicas no capítulo 2. Não iremos tratar de todos os interessantes problemas levantados pelo “ótimo” em Planeamento estratégico. Aqui, interessa-nos explorar apenas uma destas questões. Trata-se do problema da robustez como critério. Esse será, juntamente com o algoritmo ID3 e os modelos financeiros sequenciais, o objecto do presente estudo.

Este breve capítulo tem por tema a *robustez* como critério de decisão: Uma vez que se pretende ter obtido uma quantificação da robustez em decisões sequenciais, convirá apresentar o problema antes de resolve-lo.

1.1 Introdução à Noção de Robustez

A ideia intuitiva de robustez de uma decisão está ligada à *largura* da solução ótima. Se um atributo pode originar desenlaces diversos, o gestor estará interessado em conhecer o valor do atributo ao qual corresponde o melhor desenlace, i.e., um máximo ou um mínimo. Porém, o conhecimento deste ótimo não chega. Na vida real é raro que os projectos se consigam cingir a um valor exacto. O acaso, por exemplo, irá impôr um certo grau de incerteza. Sendo assim, podem dar-se desvios em relação ao atributo capaz de gerar um desenlace ótimo.

Um desenlace será robusto quando, perante tais desvios, se mantém a um nível aceitável de optimalidade. E será frágil no caso contrário: Quando um pequeno desvio pode deteriorar seriamente ou mesmo inverter o sentido do desenlace.

A figura 1 procura transmitir a ideia intuitiva de robustez. Nela, encontram-se representadas duas relações entre um atributo (em abcissas) e um desenlace (em ordenadas). A relação “B” seria o exemplo de um óptimo robusto. “A”, o de um óptimo frágil. De facto, a relação “B” admite uma certa largura de desvios em redor do valor óptimo sem que este se deteriore. Pelo contrário, mesmo um pequeno desvio em relação ao óptimo pode comprometer seriamente o desenlace no caso da relação “A”.

Claro que, em situações reais, a relação entre atributos e desenlace não é contínua. Isto modifica a forma como o problema da robustez se apresenta. O exemplo acima é artificial e destina-se apenas a transmitir a ideia geral de robustez.

A robustez segundo Rosenhead *et al.*: Rosenhead *et al.* (1972) [32] elaboram a ideia intuitiva de robustez apresentada acima. Depois de frisarem que a noção de robustez só faz sentido uma vez estabelecida a diferença entre plano e decisão — pois a característica básica de um plano é, para eles, a possibilidade de rever decisões — propõem uma definição de robustez próxima da de flexibilidade da estrutura ou operativa:

“Tanto na indústria como nos serviços públicos, o plano é aquilo que dá coesão às decisões táticas. O plano permite que as decisões, uma vez tomadas, sejam compatíveis com, ou se dirijam a, resultados considerados desejáveis. Planejar foi definido como o desenhar de um futuro desejável e dos procedimentos para torná-lo realidade. Porém, ao ser implementado, novas informações surgirão, as atitudes acerca do que é desejável podem mudar e assim o horizonte daquilo que está a ser planeado ficará ainda mais nas mãos do futuro.

Pouco do que é conhecido sobre o objecto de um planeamento pode sê-lo com certeza. E, do que não é conhecido, pouco pode ser expresso em termos de probabilidades — o comportamento futuro de outras companhias, agências e governos, bem como mudanças nas atitudes do público e nas prioridades. A situação é de incerteza. À medida que estes acontecimentos imprevistos e incontroláveis se desenrolam, mais informação fica disponível acerca do estado futuro desejável e sobre como o atingir. À luz desta informação é natural que se reconsiderem e talvez que se modifiquem os passos do plano ainda não dados. Porém, se a possibilidade de fazer revisões e modificações não foi considerada ao especificar decisões anteriores — as já implementadas —, pode não existir flexibilidade residual que permita fazê-lo mais tarde. Qualquer decisão limita o

futuro ao assumir compromissos no presente. Um plano cujas decisões iniciais comprometem o menos possível o futuro terá vantagens num mundo incerto.

Considere-se um problema de planeamento onde uma decisão deve ser escolhida de entre um conjunto $D \equiv \{d_i\}$ de decisões a curto prazo; destas, uma de entre um conjunto S de planos alternativos (ou soluções) será realizado a longo prazo. Qualquer decisão inicial d_i restringirá o conjunto dos planos atingíveis a um subconjunto S_i de S .

Suponha-se que um certo subconjunto, \hat{S} de S é presentemente considerado bom ou aceitável de acordo com alguma combinação de critérios. Um subconjunto \hat{S}_i de \hat{S} será atingível depois de uma decisão inicial d_i . Então, a *robustez* de d_i é definida (ver também [13] e [11]) como

$$r_i = \frac{n(\hat{S}_i)}{n(\hat{S})}$$

onde $n(S)$ é o número de elementos no conjunto S .

A robustez, uma medida da flexibilidade operativa contida numa decisão, tem características que a tornam um critério apropriado para a tomada de decisão sequencial sob condições de incerteza. Ela maneja a incerteza do ambiente, não através da imposição de uma estrutura probabilística, mas antes sublinhando a importância da flexibilidade. Ela torna explícita a distinção entre decisões já tomadas e soluções planeadas. Ela reflecte a natureza sequencial da tomada de decisões ao colocar menos ênfase no Plano e mais no processo contínuo de planear.

Depois, os autores definem outro critério, a *estabilidade*, complemento da robustez:

(...) por razões tanto externas como internas à organização que planeia, a sequência de decisões pode ser truncada depois da implementação das decisões iniciais, de tal modo que as etapas seguintes do plano não são atingidas. Dizemos que a decisão inicial é estável se o sistema, depois de modificado por essa decisão, tem um desempenho a longo prazo que é satisfatório com respeito às alternativas, quando as etapas seguintes da sequência de decisões não são implementadas.

Finalmente, os autores sublinham o facto de que tanto a robustez como a estabilidade devem ser vistas como um critério para a escolha de decisões iniciais de entre as que o conjunto D oferece e não para a escolha da solução final de entre as existentes no conjunto S .

A linha de raciocínio expressa acima foi posteriormente divulgada por um dos autores, Rosenhead, em diversos escritos (ver, por exemplo, [31]). Devido a essa divulgação, o tema da robustez é hoje geralmente associado com Rosenhead. Mas não é assunto que tenha merecido muita atenção: Em alguns meios académicos, a robustez das decisões sequenciais gerou um interesse que se mantém. Na prática dos sistemas de suporte à decisão o impacto desta proposta foi pequeno ou nulo.

1.2 Discussão

O presente trabalho tem propostas a fazer quanto à ideia de robustez de uma decisão. Essas propostas serão delineadas na secção 1.3 e depois concretizadas no capítulo 3. Aqui, interessa discutir a definição de Rosenhead *et al.*, expondo algumas objecções e limitações ao seu uso. O objectivo é preparar e justificar a nova definição de robustez que este estudo irá sugerir.

O primeiro aspecto característico da definição de robustez dada acima é a explícita negação do uso de probabilidades. Vão-se primeiro discutir dois problemas que decorrem desta negação.

Incerteza e Risco: Usa-se em certos meios o termo “incerteza” como contraponto do termo “risco” (ver por exemplo Wipperfurth (1966) [34]). A “incerteza” designa uma situação de total ausência de informação acerca de acontecimentos futuros. O termo “risco” é reservado para quando se conhecem as probabilidades associadas a cada desenlace. Rosenhead *et al.*, no trabalho citado acima, põem ênfase nesta definição de incerteza. Note-se que no capítulo 3 o termo “incerteza” é também usado, mas com o sentido tradicional de simples não-certeza ou falta de informação completa.

Perante acontecimentos incertos — no sentido de Rosenhead *et al.* — a informação disponível não é nula. É, ela própria, inexistente. Pode ser que um dado desenlace seja altamente provável mas o gestor não tem possibilidade de aceder a esse facto. Portanto Rosenhead *et al.* colocam-se num mundo muito especial, onde é preciso planear às cegas: Prescindindo de qualquer conhecimento sobre a verosimilhança de acontecimentos futuros.

A primeira objecção ou limitação ao uso deste critério deriva do facto de ele apenas contemplar situações de incerteza. Acontece que, ao contrário daquilo que os autores deixam supôr, as situações de total incerteza são raras em projectos de investimento. Elas coexistem

com outras, a maioria, de simples risco. Ao ignorar toda a informação a-priori existente, o critério proposto deita fora dados que geralmente são muito importantes. Com isto, torna-se pouco realista e propenso a medir pela mesma bitola situações completamente dispares.

O peso causal das decisões: Outra consequência da renúncia a qualquer informação a-priori é a necessidade de tratar todas as decisões como se tivessem o mesmo peso com vistas à obtenção dos resultados. Para o critério descrito, interessa apenas o número de decisões de cada conjunto. O peso causal de cada uma delas é ignorado. Na prática isto é equivalente a aceitar que todas têm igual peso.

Ver-se-á no capítulo 3 que o peso causal das decisões pode ser muito variado. As decisões não contribuem no mesmo grau para o desenlace. Há decisões que contribuem muito para o desenlace e outras pouco. Projectos há em que certas decisões não contribuem em nada para o desenlace, apesar de aparentarem uma certa importância.

O facto de que existe umnexo causal, forte ou fraco, entre decisões e desenlace, deveria estar na base da ideia de robustez. Se deveras se pretende aceder à robustez das decisões, torna-se necessário começar por quantificar o seu peso causal.

Peso causal e comprometimento do futuro: Expõe-se agora o reparo mais sério ao critério proposto por Rosenhead *et al.* Este reparo tem a ver com as objecções já enunciadas mas vai mais longe.

Para que a definição de robustez dada acima fosse aceitável, seria preciso que o grau em que cada decisão compromete o futuro fosse inversamente proporcional — ou pelo menos independente — do grau em que cada decisão causa o desenlace desejado. Ora não há razão nenhuma para supôr que assim tenha de ser.

Foi citada acima a afirmação

Uma decisão limita o futuro ao assumir compromissos no presente. O plano cujas decisões iniciais comprometem *o menos possível* o futuro terá vantagens num mundo incerto.

que é a síntese do método de Rosenhead *et al.* A segunda sentença deste parágrafo não parece muito realista. Vai contra a experiência comum dos gestores. Um plano, pensado para que as suas primeiras decisões comprometam o menos possível o futuro, não terá por isso especiais vantagens, mesmo num mundo incerto.

A ideia de caminhar para o desenlace desejado através de decisões que comprometam o futuro o menos possível não parece aceitável por parte de um gestor experiente. É talvez mais uma ideia ligada a problemas negociais — em que os objectivos são, eles próprios, flexíveis — e pouco tem a ver com a gestão de projectos de investimento.

Os bons gestores, se sabem alguma coisa é isto: Não existem — ou são raras — as decisões ao mesmo tempo eficazes e não-comprometedoras.

O grau em que uma decisão compromete o futuro não é geralmente independente do seu peso causal, i. e., da sua capacidade para causar o desenlace pretendido. Uma decisão que compromete pouco o futuro também será pouco eficaz com vistas à meta a atingir e vice versa. São precisamente as decisões que mais comprometem o futuro aquelas que permitem geralmente alcançar o desenlace desejado e não outro.

Caminhar através de decisões cuidadosamente escolhidas para não comprometerem o futuro é habitualmente a melhor forma de conseguir não chegar a parte nenhuma.

Sendo assim, a definição de robustez proposta por Rosenhead *et al.*, baseada como está na independência entre o grau de comprometimento do futuro e o poder de alcançar objectivos, terá que considerar-se com reservas.

Robustez e estrutura: Por último, a flexibilidade a que Rosenhead *et al.* aludem depende da estrutura do modelo sequencial em questão. Ela é conseguida à base de uma crescente sofisticação. Só depois de prever cuidadosamente todas as eventualidades se podem pôr de parte as decisões que não se mostram flexíveis.

Assim, o problema da robustez fica ligado ao problema referido no início como o da sofisticação do modelo. Quanto mais completo o modelo, melhor se pode medir a robustez das decisões. Modelos que sejam uma simplificação não permitem tal medida.

Acontece que a simplicidade nem sempre é um defeito. Em muitos casos ela é procurada como uma meta importante: Os analistas introduzem deliberadamente simplificações com o fim de manter a complexidade em níveis aceitáveis ou para conseguir modelos mais robustos.

De facto, a robustez pode mesmo opôr-se à complexidade. Quanto maior o número de atributos, mais possível é um erro ou uma avaliação enganosa. Por tudo isto, seria desejável que a noção de robustez não estivesse ligada à de complexidade e grau de sofisticação da estrutura.

1.3 Uma Nova Ideia de Robustez

Este estudo irá basear a sua proposta de definição de robustez na ideia de peso causal: As decisões robustas serão aquelas que, num dado projecto, tiverem mais peso causal do que os atributos que o gestor não controla.

Peso causal significa capacidade de intervenção efectiva. Durante o desenrolar de um plano, as alternativas são o ser-se um joguete nas mãos de acontecimentos impossíveis de controlar ou, pelo contrário, permanecer no comando. Uma decisão robusta é aquela em que o gestor pode introduzir modificações ao plano, com a garantia de que tais modificações influem no curso dos acontecimentos.

Assim definida, a robustez surge como independente da estrutura ou grau de sofisticação do modelo. A mesma estrutura pode originar diagnósticos muito dispares quanto à robustez de decisões. É também uma medida que contabiliza toda a informação existente acerca dos acontecimentos futuros. Como referido acima, isto não parece ser um ponto fraco, tanto mais que já existem formas de introduzir nos modelos a ignorância do futuro a par da falta de informação completa.

No capítulo 3 ver-se-á, no caso concreto das decisões sequenciais em projectos de investimento, como o algoritmo ID3 para a extracção de regras é capaz de medir o peso causal relativo das decisões quando comparadas com o peso causal dos atributos que o gestor não controla. O próximo capítulo introduz este algoritmo.

1.4 Sumário

A robustez de uma decisão está intuitivamente ligada à ideia de largura do óptimo. Rosenhead *et al.* propuseram uma medida de robustez baseada na flexibilidade estrutural. Este estudo contrapõe a tal sugestão a ideia de que a robustez deveria basear-se em critérios não ligados à estrutura. Por um lado, a flexibilidade estrutural não aproveita muita da informação existente acerca do futuro. Por outro, o peso causal das decisões não pode ser visto como independente do comprometimento do futuro. Além disso, uma ideia de robustez baseada na estrutura não é aplicável com generalidade. Sendo assim, sugere-se uma nova noção de robustez baseada no peso causal relativo de cada decisão.

Capítulo 2

O Algoritmo ID3 para Indução de Regras

Este capítulo introduz o algoritmo ID3 para indução de regras e discute o seu uso em dois tipos diferentes de problemas: Modelação estatística e redução de complexidade.

Originalmente, os algoritmo para indução de regras tinham como finalidade a extracção de conhecimentos a partir de dados determinísticos. Porém, pouco depois de terem sido divulgados, algoritmos como o ID3 começaram a ser usados para modelação estatística. Este estudo mostra que tal uso — quando indiscriminado — levanta problemas. O objectivo, sempre presente, é o de clarificar qual o campo de aplicação destes algoritmos e assim preparar a aplicação do ID3 como pós-processador de árvores de decisão.

Conteúdo: Em primeiro lugar introduzir-se-ão as noções básicas de que este estudo se vai servir: Quantidade de informação, o algoritmo ID3, indução. Em seguida apontar-se-á a diferença entre modelar e ordenar, central neste estudo. Por fim discutir-se-á, em duas secções separadas, o uso do ID3 em modelação. Na primeira delas reunir-se-ão as objecções gerais. Na seguinte, apenas as que mais directamente limitam o seu uso em modelação financeira.

2.1 A Quantidade de Informação em Estatística

A ideia de medir a quantidade de informação vem da engenharia de telecomunicações. Para efeitos de comunicação, a informação é aditiva. O número de dígitos necessários para

distinguir um acontecimento de entre todos os possíveis, será proporcional ao logaritmo do número N de acontecimentos possíveis. Por exemplo, 99 acontecimentos possíveis requerem dois dígitos decimais para a sua catalogação e 999 requerem três.

Diz-se portanto que $\log N$ é a *variedade* ou incerteza *a priori* de uma colecção de acontecimentos possíveis. A variedade mede o número de dígitos ou quantidade de informação que é precisa para que um acontecimento, entre vários possíveis, passe a ser conhecido. O que se pretende é saber a quantidade de informação necessária para comunicar qual o desenlace, isto é, qual dos N acontecimentos possíveis acabou por dar-se.

Quando existe alguma regularidade na colecção de acontecimentos possíveis a variedade ou quantidade de informação que é precisa para distingui-los deixa de ser $\log N$. Dá-se um *ganho* em informação ao saber-se que, por exemplo, k_i dos N acontecimentos estão agrupados numa classe ou *atributo*, isto é, todos eles partilham uma característica comum. O conhecimento de que existe tal atributo traz consigo uma certa quantidade de informação sobre o acontecimento que se espera. Este ganho por classificação deve pois ser subtraído à variedade total sempre que se pretenda saber a quantidade de informação necessária para descrever completamente um acontecimento entre vários.

Perante uma classificação múltipla, i.e., quando existem vários atributos comuns a grupos de acontecimentos, a quantidade de informação que ainda falta conhecer para identificar correctamente cada acontecimento será a diferença, H , entre a informação que faltava antes da classificação e a quantidade de informação média que tal classificação trouxe consigo. Isto é,

$$H = \log N - \sum_i \frac{k_i}{N} \log k_i \quad (1)$$

Esta diferença H é conhecida pelo nome de “entropia” porque tem sido usada como medida de irregularidade. De facto, quando $H = 0$, não existe nenhuma falta de informação sobre o acontecimento esperado. Cada um dos acontecimentos pode ser completamente descrito pelos seus atributos, i.e., pelas classes a que pertence, de tal modo que quem conhecer os atributos fica também a saber qual é o desenlace.

No polo oposto, quando $H = \log N$, não existe nenhuma informação a-priori sobre o desenlace. A incerteza é máxima porque a irregularidade da colecção de acontecimentos possíveis é também máxima. Pode pois considerar-se a irregularidade ou entropia como a diferença entre a variedade e o ganho devido a conhecimentos prévios sobre regularidades na colecção de acontecimentos.

2.1.1 O Algoritmo ID3

O ganho referido acima,

$$G = \sum_i \frac{k_i}{N} \log k_i \quad (2)$$

é a estatística que o algoritmo ID3 maximiza. O ID3 começa por determinar qual o atributo que mais explica o desenlace quando G é usado como critério. Depois, a amostra é dividida em tantas sub-amostras quantas as classes do atributo encontrado e o processo repete-se com cada uma destas sub-amostras. O resultado é uma estrutura com atributos embricados uns nos outros começando pelos mais informativos e acabando nos menos informativos. Esta estrutura costuma ser interpretada em termos de regras. O resultado é uma “árvore de regras”. A secção 2.3.2 contém um exemplo de como aplicar o ID3.

O algoritmo ID3 (“Iterative Dichotomizer 3”) foi proposto em 1979 por Quinlan [27] com base em trabalhos de Hunt, Martin e Stone [16]. A finalidade desta ferramenta era a obtenção de regras a partir de experiência acumulada. Para testá-la, Quinlan usou sequências de decisões no tempo. Mais especificamente, colecções de finais de Xadrez. Pouco depois, Breiman *et al.* [6] propuseram um algoritmo semelhante mas orientado para problemas de classificação em modelos estatísticos.

Recentemente, diversos autores publicaram estudos onde se apresenta o ID3 como um instrumento de modelação estatística de uso geral. Além dos já citados Breiman *et al.*, que no seu livro “Árvores de Decisão e de Regressão” tentam enaltecer as virtualidades da estrutura hierárquica nestes domínios, também Quinlan (1986 e 1987) [28] [29], Niblett [26], Cestnik *et al.* [8], Kodratoff e Manago [19], se dedicaram a este tipo de experiências. Race e Thomas [30] sugeriram a introdução desta técnica como modelador de “cash-flows” simulados por meio de “análise do risco” [14]. Mingers [25] comparou o ID3 com regressões lineares e pronunciou-se pela superioridade do primeiro. Braun [5] usou-o na previsão do preço de acções cotadas na Bolsa.

Toda esta literatura põe a obtenção de regras acima de qualquer outra consideração. Casos há em que tanto a variável que se pretende explicar como os atributos usados para explicá-la são claramente de evolução contínua — e é preciso proceder a uma discretização mais do que forçada antes de usar o algoritmo.

Claramente, aquilo que notamos ao ler esta literatura é uma ausência de reflexão sobre o que é a modelação de relações estatísticas e quais os requisitos necessários à obtenção de uma boa generalização. Nas próximas secções ver-se-á porquê.

2.1.2 Outros Algoritmos

A possibilidade de modelar variáveis nominais e obter regras baseadas na maximização do ganho não é nova.

O emprego em estatística da entropia e outras medidas usadas em comunicações começou em 1954 quando McGill [23] generalizou estes conceitos para mais de duas variáveis. Kullback, poucos anos depois, desenvolveu testes de significância para estas estatísticas e estendeu a sua base teórica [21]. Ele mostrou, entre outras coisas, que o popular *Qui-Quadrado* pode ser visto como uma medida aproximada do ganho em informação. Expandindo o logaritmo em potências da frequência esperada, o termo linear desaparece e o seguinte, o quadrático, é o *Qui-Quadrado*. Os testes de significância para o ganho desenvolvidos por Kullback são importantes para o ID3. O continuador de Kullback foi Good (1963) [12], que estendeu os testes de hipóteses em outras direcções.

Uma contribuição importante para o uso prático de estatísticas baseadas em quantidade de informação ficou a dever-se a Ashby (1964) [1] [2] que desenvolveu uma base para modelação estrutural de dados nominais. Muitos outros estudos refinaram este processo, aplicando-o especialmente a problemas em Ciências Sociais e Biologia (ver, por exemplo, Conant (1979) [9], [10], Cavallo (1979) [7], Klir (1976) [18] e Kripendorff (1981) [20]). É deste último autor um estudo de que nos serviremos mais adiante.

A entropia ou o ganho têm muitos pontos em comum com a log-verosimilhança. Como medidas da bondade de um ajuste, são ambas aditivas em todos os modelos onde não existam “loops” ou zeros estruturais. Este facto torna as estimativas do grau de ajuste aos dados sob sucessivas simplificações uma tarefa simples. Os resultados da aplicação destas ferramentas costumam ser similares aos obtidos ao usar modelos Log-Lineares [4].

Hoje, existe um corpo de conhecimentos bem estabelecido e reconhecido como útil, que aplica medidas baseadas na teoria das comunicações à modelação estatística. Isto, especialmente quando os dados são de tipo nominal. Ao estudar as potencialidades do algoritmo ID3 em finanças, tal corpo de conhecimentos é importante para comparação pois usa o mesmo critério que o ID3.

Além disso, a existência e a qualidade destas ferramentas não pode ser ignorada. Tarefas que elas possam desempenhar melhor do que o algoritmo de Quinlan ou semelhantes, não deveriam ser abordadas como se todo este conjunto de técnicas não existissem. Por outras palavras, não é objectivo dizer-se que o algoritmo ID3 vem preencher uma lacuna no campo dos instrumentos de modelação estatística.

2.2 A Diferença Entre Modelar e Ordenar

O problema da extracção das regras subjacentes a observações faz parte do problema mais geral da aquisição de conhecimentos a partir de casos particulares por indução. Pretende-se, de um conjunto de exemplos ou amostra, generalizar para a população. É isto o que se chama aprendizagem por indução.

O que é específico de algoritmos como o ID3 é a possibilidade de se obterem estruturas lógicas como o resultado final da aprendizagem. Mas a ideia de aprender a partir de exemplos é mais abrangente do que a extracção de regras.

Durante mais de vinte anos, entre a década dos sessenta e a dos oitenta, as ferramentas fornecidas pela computação simbólica foram consideradas como as únicas interessantes para abordar problemas que requeressem inteligência artificial. Daí a importância que adquiriu a extracção de estruturas hierárquicas de regras.

Note-se porém que qualquer instrumento de modelação capaz de quantificar ou classificar é também apto para este fim embora exija um pouco mais de sofisticação.

Generalização e ordenação: Um aumento em conhecimento por indução pode obter-se por muitas vias. Aqui estamos interessados em comparar duas dessas vias: A generalização, que se aplica a desenlaces não completamente determinados pelos seus atributos, e a ordenação ou o re-arranjo, que se aplica a desenlaces que são completamente determinados pelos seus atributos, mas complexos. Um desenlace é determinado ou determinístico quando pode ser previsto por observação das suas qualidades ou atributos.

Dados não-determinados: No caso de desenlaces não completamente determinados, desde que se descubra alguma regularidade na relação atributos-desenlace, pode ser legítimo extrapolar essa regularidade para toda a população da qual os dados são uma amostra. Tais *modelos* são valiosos para efeitos de aprendizagem por indução na medida em que sejam representações gerais. Por isso a sua construção deve ignorar os detalhes e sublinhar as características comuns. Neste processo deixam de interessar os valores concretos que cada caso assume. Passam a interessar os valores esperados. Supõe-se que procedendo assim se conseguirá uma descrição aplicável, não apenas a um dado conjunto de casos, mas a todos os pertencentes à mesma população.

Re-arranjo ou ordenação: A segunda via referida para extracção de conhecimentos é a que se aplica a observações determinísticas mas complexas: Ordenação, classificação, re-arranjo, como meio para eliminar redundância em conteúdo informativo e descobrir alguma qualidade simples escondida pela aparente complexidade. Trata-se portanto de um processo redutor de complexidade.

Um re-arranjo, uma ordenação, são apenas transformações que se aplicam aos dados. Não se dá com elas nenhuma perda de informação; só se “perde” redundância. Pelo contrário, na modelação estatística, dá-se um processo redutor de informação: O que é particular é desprezado com o fim de obter uma descrição geral.

Comparação das duas tarefas: Este estudo insiste na distinção entre complexidade determinística e indeterminação, no que respeita à tarefa de aprendizagem por indução.

As técnicas de indução de regras foram inventadas para dar resposta a um problema de complexidade determinística e não de indeterminação. Pretendia-se converter experiência — expressa de uma forma complexa — em estruturas lógicas. Esta tarefa requer uma capacidade redutora de complexidade mas não uma capacidade para generalizar, própria da modelação estatística. O final de um jogo de Xadrez, com todas as possíveis hipóteses, é sem dúvida um problema complexo; mas os desenlaces são completamente determinados pelos atributos — que neste caso são as jogadas e posições iniciais dos jogadores.

Pelo contrário, nos problemas em que o desconhecimento do futuro, os erros de medição ou outra fonte de indeterminação estão presentes, a indução deixa de ser possível através da simples ordenação e classificação. Surge a necessidade de se dispôr de uma nova capacidade, a de generalização. A natureza indeterminada dos dados exige, não um processo redutor de complexidade apenas, mas um processo redutor de informação considerada particular.

O algoritmo ID3 é originalmente uma ferramenta para a extracção de regras a partir de dados determinísticos complexos. Nisto, tem-se mostrado superior já que não lhe é pedida a capacidade de generalizar mas a de ordenar.

As medidas baseadas na quantidade de informação são adequadas a ambas as tarefas referidas. A pesquisa em árvore, própria do ID3, parece eficaz em tarefas que exijam redução da complexidade determinística já que, ao ordenar os atributos por poder explicativo do desenlace, vai eliminando uma boa parte ou mesmo toda a redundância associada com cada um deles. Porém, ver-se-á a seguir que esta estrutura não é a mais adequada à tarefa de generalizar excepto em casos particulares.

2.3 Crítica ao Uso da Indução de Regras em Estatística

Esta secção mostra as objecções que podem fazer-se ao uso do ID3 em problemas de modelação estatística. Começar-se-á por introduzir um tipo de considerações que, não sendo elas próprias muito ponderosas, servem de introdução aos problemas apresentados a seguir.

2.3.1 Generalização

Frequentemente é difícil descobrir, a partir de uma colecção de observações, um modelo com boas qualidades de generalização. Boas descrições de uma amostra podem ser apenas válidas para essa amostra, mas não para outras extraídas da mesma população. Não seria legítimo considerar as regularidades encontradas como características comuns a todos os casos possíveis.

Generalização e suposição a-priori: A capacidade de um modelo para generalizar adequadamente está muitas vezes condicionada ao prévio conhecimento e uso de uma propriedade inerente às observações. Só depois de descoberta e assumida essa propriedade pode o modelo tentar ser uma representação geral. É esta talvez a razão pela qual as técnicas de modelação estatística são, em tantos casos, uma suposição a-priori sobre as mais importantes características da população. Assim, a Regressão Linear é usada sempre que a hipótese de co-variâncias lineares parece plausível. E a Análise de Variância supõe a existência de efeitos centrais.

A experiência mostrou que estas opiniões prévias acerca de qualidades gerais são de grande importância prática: Para que um modelo consiga generalizar aceitavelmente é preciso que tenha havido um acerto entre este pressuposto inicial e uma característica da população. Por isso, os utensílios de modelação estatística que se têm mostrado capazes de uma boa generalização procuram captar, não uma descrição razoável das observações, mas uma qualidade geral: Uma característica, um mecanismo, uma estrutura.

Explicar um máximo de variabilidade não é modelar: Existem porém outros instrumentos que não partem de nenhum pressuposto acerca da população. Procuram apenas descrever um máximo de variabilidade e usam para isso técnicas muito gerais.

Uma Regressão Linear é naturalmente limitada na sua capacidade para explicar a variabilidade de uma amostra pela hipótese de linearidade. Tal hipótese será também a fonte da

sua capacidade para generalizar, caso a linearidade seja uma característica da população. Pelo contrário, quando se usam polinómios de grau arbitrário, um simples atributo pode descrever toda a variabilidade de uma amostra — mas sem utilidade. Enquanto que a primeira se encontra limitada na sua capacidade para descrever a variabilidade observada, os segundos, por não partirem de nenhuma hipótese prévia, são ilimitados na capacidade de modelar, podendo descrever a amostra com cada vez mais detalhe, até ao ponto de produzirem descrições exactas dos dados a que foram expostos.

Quando um instrumento de modelação é ilimitado na sua capacidade de aproximar observações, existe o problema de descobrir quando se deve parar de acrescentar detalhes ao modelo. A Indução de Regras sofre de esta falta de moderação ao aproximar dados. Ao usá-la, é fácil obter modelos detalhados e inúteis. E é difícil ver as autênticas características da população, escondidas debaixo de descrições exaustivas.

Autores tem havido que comparam o ID3 com outras técnicas e demonstram que o primeiro explica mais variabilidade. Se as técnicas de indução de regras têm que ser comparadas com instrumentos de modelação estatística como a Regressão Linear, deveriam sê-lo com base na sua capacidade para generalizar, não na variabilidade explicada.

Não faz sentido falar de rendimento de modelos estatísticos em termos da variabilidade que eles conseguem explicar. Seria mais interessante mostrar a possível capacidade das técnicas de indução de regras para encontrarem representações gerais. É o que se fará agora, usando a linearidade como paradigma.

O ID3 não pode modelar relações lineares: A indução de regras, quando usada como ferramenta de modelação, produz uma estrutura lógica hierárquica. Será que uma relação linear pode ser modelada por tal tipo de estrutura? Este parece ser um problema interessante posto que muitos processos estatísticos são lineares.

No caso das observações puramente nominais, linearidade significa pura aditividade ou ausência de interações entre atributos. O desenlace é explicado apenas através de efeitos de primeira ordem. Quer isto dizer que a capacidade de um dado atributo para explicar o desenlace não é afectada pela presença dos outros atributos. E portanto, a variabilidade explicada será a mesma, quer se construa um único modelo com todos os atributos presentes, quer se construam tantos modelos quantos os atributos e tendo cada um o seu.

Se o algoritmo ID3 tiver que descrever uma relação linear, começará por descobrir o atributo com maior poder para explicar o desenlace. A seguir, suposta já explicada essa

porção da variabilidade, procurará de novo, de entre os atributos restantes, aquele com maior poder explicativo; e por aí fora. Porém, a maneira como o modelo foi crescendo corresponde a uma estrutura muito particular: Os atributos aparecem sempre embricados dentro de atributos, como se se pretendesse explicar atributos com atributos. Os efeitos estão presentes neste modelo sempre condicionados ao atributo previamente entrado: Se o atributo A foi o primeiro a ser aceite, então B , o seguinte a entrar, há-de modelar um efeito “ B embricado em A ”. Nunca se dará o cruzamento livre ou a adição de efeitos. Daqui que o algoritmo ID3 não seja capaz de modelar relações lineares enquanto tais.

Estruturas hierárquicas não conseguem descrever aditividade. Aquilo que há de mais comum escapa a estes algoritmos.

Graus de liberdade e indução: No caso em estudo, a variabilidade explicada por todos os atributos menos o primeiro será apenas devida ao acrescentar de mais e mais graus de liberdade ao modelo da mesma forma que um polinómio de grau arbitrário explica uma relação, qualquer que ela seja. A potência do ID3 é obtida à custa do uso de muitos graus de liberdade e não porque tenha acertado com uma característica da população. Como resultado, a sua capacidade para generalizar será pobre. É uma vez que a generalização é, para dados estatísticos, a base da indução, não havendo a primeira, não se pode falar da existência da segunda. O ID3 produz regras mas induz pobremente.

2.3.2 Indução de Regras e Estrutura

Esta secção contém o mais forte e mais básico argumento contra o uso do ID3 como ferramenta de modelação estatística: A indução de regras não deve ser um instrumento de modelação estatística de uso geral pois é capaz de induzir regras erróneas. Para mostrá-lo descrever-se-á um caso em que isto acontece.

Um caso: Krippendorff [20] cita uma tabela aparecida no New York Times Magazine (11 de Março de 1979) contendo a frequência com que a pena de morte foi aplicada no estado da Florida (Estados Unidos) entre 1973 e 1979, por raça do assassino e da vítima. Reproduzimos tais dados na tabela 1. O problema de construir um modelo estatístico para explicar o desenlace, morte ou outra pena, em termos das raças do assassino e da vítima tornou-se uma espécie de padrão em modelação estatística com variáveis nominais. Parece existir um modelo estrutural muito simples mas capaz de explicar quase a totalidade da

RAÇA:		PENA APLICADA:	
DO ASSASSINO	DA VITIMA	MORTE	OUTRAS PENAS
Negro	Branco	48	239
Negro	Negro	11	2209
Branco	Branco	72	2074
Branco	Negro	0	111

Tabela 1: A frequência de penas de morte na Florida, entre 1973 e 1979, por raça do assassino e da vítima.

variabilidade encontrada nestes dados [20].

Que nos poderia dizer um algoritmo para indução de regras quando tentando modelar as observações descritas pela tabela 1? Em primeiro lugar, o atributo que melhor descreve o desenlace é Vítima, como se vê pelas tabelas de contingência abaixo.

ASSASSINO	DESENLACE		VÍTIMA	DESENLACE	
	MORTE	OUTRO		MORTE	OUTRO
Branco	72	2185	Branco	120	2313
Negro	59	2448	Negro	11	2320
Qui-Quadrado: 3.2			Qui-Quadrado: 88.6		

O atributo Vítima produz um Qui-Quadrado de 88.6 ao cruzar-se com o desenlace enquanto que o atributo Assassino produz um Qui-Quadrado de 3.2. Lembramos que o Qui-Quadrado é uma aproximação do ganho. Portanto, as tabelas de contingência acima mostram que aquilo que se ganha em informação acerca do desenlace quando se conhece a raça da vítima é muito superior ao que se ganha ao conhecer a raça do assassino.

O passo seguinte consiste em construir duas novas tabelas cruzadas, uma para as vítimas brancas (à esquerda) e outra para as vítimas negras (à direita):

ASSASSINO	DESENLACE		ASSASSINO	DESENLACE	
	MORTE	OUTRO		MORTE	OUTRO
Branco	72	2074	Branmco	0	111
Negro	48	239	Negro	11	2209
Qui-Quadrado: 96.5			Qui-Quadrado: Não Sig.		

As regras induzidas pelo ID3 são pois as seguintes:

Se Vítima é branca então:

se Assassino é branco, sentença é
outra que não a morte;

se Assassino é negro, sentença é
de morte.

Se Vítima é negro então:

sentença é
outra que não morte.

As observações não foram todas classificadas correctamente, mas já não há mais atributos disponíveis.

Discussão do resultado: O algoritmo, devido à sua própria forma de trabalhar, não pode construir uma solução que contenha duas relações binárias entre o desenlace e atributos diferentes. Um conjunto de regras não-exclusivas como estas:

Se Vítima é branca então a sentença é
de morte.

Se Assassino é negro então a sentença é
de morte.

nunca podem ser atingidas por um algoritmo que impõe aos dados uma estrutura em árvore, onde tudo passa pela raiz, e portanto onde as relações só podem ser cada vez de maior ordem. No nosso exemplo o algoritmo criou uma relação ternária, muito sedutora, mas que, por não existir nos dados, é errónea.

E é errónea, não apenas porque produziu regras erradas. É-o, muito basicamente, por duas razões:

- Quaisquer que fossem os dados, uma relação ternária (Pena — Assassino — Vítima ou Pena — Vítima — Assassino) teria sempre que surgir. O estado da Florida estava à partida condenado pelo ID3 a um veredicto em que as penas eram explicadas em termos das raças dos assassinos e das vítimas.
- Seria impossível, também à partida, obter uma estrutura onde a raça da vítima e a do assassino explicassem a pena de morte de uma forma aditiva.

A título de informação complementar, convirá saber-se que uma modelação cuidadosa mostrou que os únicos efeitos significativos são dois, e são binários: A pena de morte é mais provável quando a vítima é branca, qualquer que seja a raça do assassino. Os assassinos são mais provavelmente sentenciados à morte quando negros, qualquer que seja a raça da vítima. A ideia de que assassinos negros de vítimas brancas são mais prováveis candidatos à pena capital não encontra confirmação nos dados. Tal relação ternária está muito longe de acrescentar uma quantidade significativa de variabilidade explicada ao modelo.

O exemplo apresentado mostra pois que a mensagem transmitida pelos dados pode ficar completamente distorcida pelo facto do algoritmo ID3 impôr uma estrutura lógica.

Isto dar-se-á especialmente em problemas onde a verbosidade se opõe à precisão. A indução de regras não se mostra adequada à modelação de problemas simples. Havendo grande variedade de atributos, o algoritmo parece conseguir bons progressos no sentido da simplificação, especialmente ao princípio. Mas diante de problemas com poucos atributos e onde a estrutura interna é importante, a indução de regras é de evitar.

Conclusão: Num final de Xadrez o desenlace é determinado pelos atributos. Mas não é simples. Aí, a tarefa de um sistema que aprende seria a de reduzir a complexidade eliminando redundâncias, ordenando e classificando. Procura-se como meta a simplicidade. Neste caso, a indução de regras parece convincente.

A distribuição de penas de morte na Florida é simples. Mas o desenlace não é determinado pelos atributos. Aqui, a tarefa de um sistema capaz de aprender é a de modelar, isto é, reduzir a informação não-significativa e revelar apenas o que é geral. Neste caso, a indução de regras mostra-se pouco convincente. além disso, a estrutura imposta pelo algoritmo pode conduzir a inferências erróneas.

Sistemas periciais: Os sistemas periciais comuns fogem de soluções em termos de efeitos porque estas, ao contrário das árvores de regras, não conduzem a um veredicto lógico. Porém, esta razão de conveniência não justifica o uso do algoritmo ID3 em problemas onde ele falsearia as conclusões.

Acresce que a descrição em termos de efeitos estatísticos que os modelos proporcionam, também se presta à inferência. O facto da computação simbólica não se dar bem com esta forma de inferir mostra apenas que todos os utensílios têm a sua especificidade.

2.3.3 Como Podar Árvores de Regras?

A objeção seguinte decorre da anterior. Não existindo correspondência entre a estrutura desenhada pelo modelo e qualquer característica interna da população, o problema de determinar até que ponto se devem considerar como significativos os desenvolvimentos de regras descobertas pelo modelo (isto é, como “podar” a árvore) torna-se um exercício cego.

Faz sentido medir a relativa importância de, por exemplo, um efeito isolado em Análise de Variância. Faz sentido discutir a significância de uma pendente em Regressão Linear. Relações causais podem ser testadas para avaliar o seu peso relativo, e isso faz sentido. Mas não parece tão claro como se há-de discutir a significância de pedaços isolados de representações embricadas obtidas a partir de observações não-embricadas. Não havendo correspondência entre o efeito e qualquer característica da população, um teste de significância passa a centrar-se no próprio modelo; deixa de ser um teste da importância de um efeito quando acareado com a variabilidade total.

Mesmo assim, sugere-se uma modificação no procedimento sugerido por alguns autores (ver por exemplo [24]). A aproximação correcta, caso um algoritmo de indução de regras faça sentido, é a de calcular G ou outra estatística para a árvore toda, i.e., sem podar. Depois de cada poda “para trás”, calcula-se G sobre o que resta da árvore original.

Pode obter-se a significância da diferença entre ambas as medidas, a qual se distribui segundo um Qui-Quadrado com tantos graus de liberdade quantos os que se perderam ao podar. Assim, ter-se-ia uma ideia de quando parar. É difícil de dizer se este método iria dar resultados muito diferentes dos obtidos pelos métodos habituais; mas parece mais adequado do que medir a significância de folhas soltas que por sua vez não representam nada de real.

2.4 Indução de Regras e Modelos Usuais em Finanças

Têm-se vindo a estudar as principais objecções ao uso dos algoritmos para indução de regras como ferramentas estatísticas. Vamos agora ver outros problemas não tão gerais, mais próprios dos modelos que se usam em finanças. A este respeito salientar-se-ão duas limitações: Os problemas decorrentes da discretização de observações e a avaliação de probabilidades a-priori.

2.4.1 Problemas de Medição

Em modelação estatística a informação disponível pode apresentar-se em diferentes escalas de medição. É costume seguir Stevens [33] e considerar quatro escalas, segundo o poder ou riqueza de informação. Lembra-se brevemente as suas características.

Medidas Nominais, simples conjuntos de categorias ou etiquetas. Classificam conjuntos disjuntos como o mesmo ou diferente, mas não qualificam as diferenças.

Medidas Ordinais, capazes de comparar ordenações relativas sem contudo quantificarem o tamanho dos intervalos que separam um caso do seguinte.

Escalas Intervalares, que introduzem uma métrica de intervalos iguais mas com origem arbitrária.

Escalas Racionais, a forma mais poderosa de medir, que possui um zero e iguais intervalos. As suas propriedades correspondem às dos números reais.

É sempre possível extrair ordens ou etiquetas de medidas mais ricas. Em tal caso, perde-se informação. Mas não é possível extrair medidas de maior conteúdo informativo a partir de escalas pobres: As etiquetas contêm menos informação do que as ordens; e estas, menos do que as escalas.

Deve ainda notar-se, com Stevens, que as operações permitidas nos dados são específicas do tipo de medição a que pertencem, ou de maior ordem com perda de informação. Um instrumento de modelação estatística apropriado para um dado tipo de medição será geralmente inválido com tipos menos ricos pois requer operações não suportadas por esse nível. Inversamente, uma técnica estatística apropriada para níveis pobres será pouco eficaz no uso que faz da informação disponível quando em níveis mais ricos.

A indução de regras lida com o mais pobre de entre os tipos de medição: Só pode manejar simples categorias não-ordenadas. Quando aplicamos estes algoritmos a observações racionais, como costumam ser as encontradas em modelos financeiros, é preciso perceber-se qual a informação que se está a perder e como esta perda irá influir nos resultados.

Assim, o uso de dados racionais como simples etiquetas compreende três sucessivas reduções que danificam o conteúdo informativo das observações da seguinte forma:

1. Ignorar a existência de um zero real é o mesmo que considerar o zero como uma posição arbitrária. Em modelação financeira, a diferença entre ganhos e perdas esvai-se.

2. Ignorar a métrica intervalar é o mesmo que limitar a informação sobre um conjunto de empresas a uma simples ordenação (primeiro, segundo, terceiro, ...) mesmo quando se tratar de empresas tão diferentes em tamanho como a General Motors, a IBM e a mercearia local.
3. Finalmente, ignorar mesmo a ordenação, como os algoritmos de indução de regras fazem quando aplicados a escalas racionais, é o mesmo que limitar a informação acerca das empresas referidas acima à afirmação de que são diferentes, como de facto são, e mais nada.

Claramente, a perda de informação é severa. Isto pode reflectir-se no comportamento final do modelo. Só uma razão muito especial pode conduzir a aceitar tal simplificação.

O dano causado no conteúdo informativo das observações pelo uso de técnicas de modelação redutoras é às vezes difícil de reconhecer intuitivamente. Uma possível razão está no facto de que as etiquetas possuem geralmente informação elas próprias, embora não acessível directamente. Uma etiqueta a que se deu o nome de “segundo” ou “negativo”, quando aposta sobre dados nominais, pode dar a impressão de que esses dados contêm em si mesmos informação sobre ordem ou escala. E de facto, pode recuperar-se alguma da informação perdida guardando-a nas etiquetas e criando uma tabela de correspondências ou mapa para interpretar resultados. Porém, em modelos onde se pretenda uma aprendizagem automática, este domínio sobre a informação perdida pode tornar-se difícil.

As técnicas de indução de regras, quando usadas sobre dados financeiros discretizados da maneira descrita acima, não produzirá regras erradas. Mas os resultados podem ser enganadores devido a esta falta de controlo sobre o conteúdo informativo que a discretização fez perder. Duas situações muito semelhantes podem gerar regras que estabelecem alternativas diferentes; e duas posições realmente diferentes podem facilmente ser alisadas numa simples, grosseira, regra que as descreve a ambas.

2.4.2 Probabilidades e Correlações A-Priori

Os modeladores baseados na indução de regras, pelo facto de usarem apenas variáveis nominais, sofrem ainda de outro contratempo: Passa a ser preciso ter em conta, explicitamente, pressupostos Bayesianos quanto a probabilidades a-priori. De outra forma não é possível interpretar correctamente os resultados (ver por exemplo [3] [17]).

Os mecanismos usados para indução de regras têm como base a comparação de diferentes

frequências. Assim, torna-se obrigatório um ajustamento bayesiano segundo as probabilidades a-priori existentes, como acontece com outros classificadores (Análise Discriminante ou Regressão Logística por exemplo).

Quando Mingers [25] aplica o algoritmo ID3 ou outro dos algoritmos para indução de regras à previsão de resultados de futebol, usa 164 jogos. Na sua amostra, a proporção de cada possível resultado é de 79/164 (vitória), 40/164 (empate) e 45/164 (derrota). Será que estas proporções concordam com qualquer pressuposto acerca das proporções observadas na população? Podem ser usadas, como o são, como probabilidades a-priori implícitas?

A inferência bayesiana caracteriza-se pelo subjectivismo. As discussões em torno dos valores adequados para as probabilidades a-priori raramente são conclusivas. Assim, quando é possível usar instrumentos mais ricos em capacidade de modelar a informação existente, o facto de com eles se evitar subjectivismos já parece razão suficiente para os preferir.

2.5 Sumário

A pesquisa em árvore, própria do ID3 parece eficaz em tarefas que exijam redução da complexidade determinística já que, ao ordenarem os atributos por poder explicativo do desenlace, vão eliminando uma boa parte ou mesmo toda a redundância associada com cada um deles.

Perante dados não determinados a tarefa de um sistema capaz de aprender é a de reduzir a informação não-significativa e revelar apenas o que é geral. Neste caso a indução de regras mostra-se pouco convincente. além disso a estrutura embricada imposta pelo algoritmo pode conduzir a inferências erróneas.

Capítulo 3

Indução de Regras e Robustez

O último capítulo deste estudo mostra que o ID3 tem uma promissora aplicação em modelos sequenciais de decisão financeira. Este algoritmo é capaz de dar ao gestor, entre outras interessantes medidas, uma quantificação da robustez, tal como foi introduzida na secção 1.3.

3.1 O ID3 em Modelos Sequenciais

Em face do exposto no capítulo anterior, qual é o futuro das técnicas de indução de regras em modelação financeira? Em primeiro lugar, mostrar-se-á que existem modelos que criam estruturas semelhantes à que o ID3 supõe nos dados. Na secção 3.1.2 discutir-se-á a adequação das medidas baseadas em quantidade de informação a problemas de decisão.

3.1.1 Estruturas Concordantes

No capítulo 2 foram apontados alguns inconvenientes do uso de algoritmos para indução de regras como utensílios de modelação estatística. Catalogando as técnicas usuais em finanças, estes reparos aplicam-se especialmente aos modelos baseados em sistemas de equações, à simulação e aos classificadores estatísticos. Mas existe um âmbito de utilização ao qual eles não se aplicam. Trata-se dos modelos sequenciais de decisão, conhecidos como árvores de decisão. Com efeito, as observações (atributos e desenlace) são, em projectos de investimento, de natureza nominal ou ordinal; e a estrutura interna é embricada, hierárquica, coincidindo portanto com a estrutura que o ID3 supõe existir nos dados.

Árvores de decisão: Para um dado projecto de investimento, procura-se com a respectiva árvore de decisão representar todos os desenlaces possíveis em função das decisões e das circunstâncias exteriores que a eles conduzem. As árvores de decisão contêm dois tipos diferentes de atributos: Aqueles que dependem dos gestores — as decisões — e aqueles que não dependem — as “jogadas da natureza” ou respostas do azar —. Estes últimos são o elemento probabilístico. As árvores de decisão usam-se para descobrir as decisões que conduzem ao desenlace óptimo esperado, para depois analisar a sensibilidade deste desenlace a diferentes cenários, ou para simulação [15].

Uma vez que as referidas decisões e respostas do azar formam uma sequência no tempo, as árvores de decisão são modelos hierárquicos onde os efeitos se encontram embricados dentro dos referentes ao período anterior. São estruturalmente semelhantes a finais de Xadrez onde um elemento indeterminado estivesse presente.

Decisões sequenciais originam sempre este tipo de desenhos embricados. Como vimos, muitos dos primeiros ensaios de algoritmos para extracção de regras foram conduzidos sobre problemas de decisão sequenciais. O trabalho de Quinlan — extrair regras de finais de Xadrez — é um exemplo típico de estrutura sequencial embricada: O primeiro movimento condiciona todos os restantes, o segundo os seguintes e por aí fora. Como acontece com as decisões num investimento.

Sendo assim, a mais ponderosa objecção ao uso do ID3 — a possibilidade de induzir regras erróneas quando a estrutura embricada se impõe a dados que não o são — deixa de estar de pé. Por outro lado, a incapacidade do ID3 para generalizar também não é importante no caso de árvores de decisão. Não se espera de uma árvore de decisão — construída para retratar um dado projecto de investimento — que possa ser usada em todos os restantes projectos do mesmo género.

Ver-se-á agora que a estatística usada pelo ID3 como critério é especialmente adequada à quantificação da causalidade dos atributos e, no caso concreto de desenlaces discretos, é também capaz de reduzir a redundância porventura existente.

3.1.2 Quantificação da Causalidade

No capítulo 2 foram introduzidas duas medidas baseadas na quantidade de informação: O ganho e a entropia. Aplicar-se-ão estas medidas à comparação de jogos de azar. O objectivo é determinar o real peso, em projectos de investimento, do conteúdo probabilístico das diversas hipóteses que o gestor enfrenta. Mais especificamente, os próximos parágrafos

procuram determinar qual a quantidade de informação que falta aos gestores para tomar uma decisão à qual se seguem jogadas da natureza.

Informação, incerteza e probabilidades: Nas árvores de decisão é costume usarem-se probabilidades, p_i , para quantificar a verosimilhança de cada desenlace, i , associado a atributos indeterminados.

As probabilidades são apenas uma entre várias possíveis maneiras de expressar expectativas ou tendências acerca de acontecimentos incertos. Existem expectativas quando há alguma informação disponível sobre o desenlace mas esta é incompleta. A probabilidade de 9/10 associada à extracção de uma bola branca de uma urna dá ideia de uma forte expectativa ou a existência de informação quase completa sobre tal desenlace.

Porém, a forma como as probabilidades medem expectativas não é a mais sugestiva no caso de árvores de decisão. Para que um gestor fique de posse de todos os dados necessários a uma decisão, precisará de saber, além das probabilidades associadas a acontecimentos incertos, o número desses acontecimentos e a sua estrutura.

Por exemplo, uma probabilidade de 1/2 associada ao desenlace i mede algo muito diferente consoante se trate de um jogo com dois ou seis possíveis desenlaces. No primeiro caso, expressa expectativas nulas ou ausência de qualquer informação. No segundo, esta mesma probabilidade expressa uma tendência ou expectativa forte a favor da ocorrência de i ou uma apreciável quantidade de informação acerca do desenlace. Em termos mais gerais, se N é o número de possíveis desenlaces, uma probabilidade de $1/N$ de que um deles se dê significa ausência de expectativas ou um máximo de incerteza.

Seria desejável medir, com uma só observação, a expectativa associada a jogos sequenciais. Isto tornaria a interpretação de árvores de decisão muito mais fácil. Ver-se-á nos parágrafos seguintes que o algoritmo ID3 é capaz de fornecer uma tal medida.

A incerteza de um jogo de azar: É costume usar-se a entropia para medir a incerteza associada a um atributo não determinado. Seria fácil de ver que H , tal como foi definida em (1) (página 10) pode ser escrita como a média ou valor esperado da informação que falta para conhecer completamente um desenlace:

$$H = - \sum_{i=1}^N p_i \log p_i \quad (3)$$

onde p_i é a probabilidade de ocorrência de cada um dos N possíveis desenlaces que esse

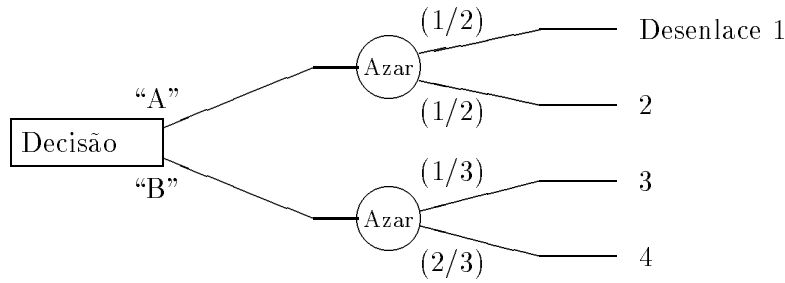


Figura 2: Um elemento básico de qualquer árvore de decisão. A uma decisão de um gestor (“A” ou “B”) segue-se a “resposta da natureza”. As probabilidades associadas a cada desenlace estão entre parêntesis.

atributo — um jogo de azar — contempla. De novo, $H = 0$ denotaria completa informação sobre o desenlace. Nesse caso o atributo seria determinístico. No outro extremo, $H = \log N$ significaria uma total ignorância sobre o desenlace ou uma completa ausência de expectativas. O atributo seria maximamente indeterminado.

Entre os dois casos extremos, uma expectativa ou uma tendência moderada fica completamente descrita por H . Considerar a entropia em vez de uma colecção de probabilidades e sua estrutura pode simplificar e fazer mais realista a tomada de decisões.

Para entender porquê, considerar-se-á uma decisão à qual se seguem, para cada possível movimento do gestor, uma jogada da natureza. Esta simples árvore de decisão, esquematicamente representada na figura 2, pode considerar-se como o elemento básico de qualquer decisão. Neste caso, o uso da entropia dá ao gestor a possibilidade de comparar, com mais realismo do que se usasse colecções de probabilidades, cada um dos possíveis jogos da natureza associados às opções que enfrenta.

A figura 2 mostra uma decisão à qual se segue um entre dois jogos envolvendo dois possíveis desenlaces com probabilidades de ocorrência de p_1 e $p_2 = 1 - p_1$. Ao contrário do que seria intuitivo, a diferença entre uma incerteza dada por

$$p_1 = \frac{1}{2}, p_2 = \frac{1}{2} \quad \text{e a incerteza gerada por} \quad p_1 = \frac{1}{3}, p_2 = \frac{2}{3}$$

é negligível e não merece ser tida em consideração excepto quando o jogo associado a cada decisão tem que repetir-se muitas vezes. Tal facto é claramente visível quando se usa a entropia de cada um deles como uma estimação da incerteza. A tabela 2 mostra o valor de H para pares de probabilidades associados a um jogo com dois possíveis desenlaces.

Não há diferença significativa na entropia associada a um jogo onde as probabilidades são $\{1/2, 1/2\}$ e outro onde elas são $\{1/3, 2/3\}$. As diferenças só começam a ser importantes

PROBABILIDADES	ENTROPIA	PROBABILIDADES	ENTROPIA
1/2 e 1/2	0.30	1/5 e 4/5	0.22
1/3 e 2/3	0.28	1/10 e 9/10	0.14
1/4 e 3/4	0.24	1/20 e 19/20	0.09

Tabela 2: Relação entre Entropia e Probabilidade num jogo com dois desenlaces.

a partir de 1/3. Na figura 2, apesar das probabilidades parecerem indicar uma quebra na incerteza quando a decisão é “B”, esta é na realidade mínima.

A relação entre probabilidades e incerteza — a informação que falta conhecer sobre um desenlace — não é linear. Em redor de valores de $p = 1/N$, grandes variações de p conduzem a pequenas variações da incerteza. E nos extremos, quando p se aproxima de zero, qualquer pequena variação de p produz uma grande variação em incerteza.

Quantificação da causalidade: Quando, em vez do exemplo simples descrito acima, se está perante uma árvore de decisão mais complexa, o efeito enganador das probabilidades ou dos valores esperados — que são combinações lineares de probabilidades — pode fazer com que percursos semelhantes pareçam diferentes ou ao invés.

O algoritmo ID3, ao juntar probabilidades, número de possíveis acontecimentos e estrutura numa única medida de informação, vem ao encontro deste problema. Em primeiro lugar, estruturas sequenciais de probabilidades são transformadas num desenho ou representação da importância relativa de cada atributo para o desenlace. E em segundo lugar, este algoritmo reduz a redundância ou complexidade aparente de tal modo que certas características dinâmicas da decisão, antes escondidas, ficam visíveis.

O algoritmo ID3 clarifica a relativa importância das causas de cada possível desenlace em problemas de decisão com atributos probabilísticos baseando-se em medidas da quantidade de informação. Citando Ashby [2],

A teoria da comunicação é basicamente uma contagem. (...) Tal contagem foi estendida para dar respostas a problemas que teriam desafiado os métodos do contabilista tradicional: Casos onde as causas são contínuas, ou onde se encontram causas relevantes misturadas com outras que o não são. Quando estudamos a forma como a quantidade de informação se distribui e evolui dentro de um sistema, o que estamos realmente a medir são as *quantidades de causa e efeito* que trabalham dentro desse sistema.

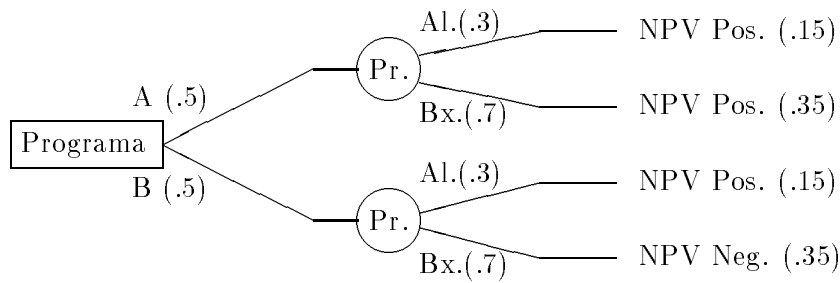


Figura 3: Uma árvore de decisão muito simples. *Pr.* é o atributo Procura. Entre parêntesis, a verosimilhança de cada classe.

O facto de algoritmos como o ID3 serem capazes de quantificar causas é central para este estudo e está directamente ligada à noção de robustez que se pretende explorar.

3.1.3 Como Aplicar o ID3 a Árvores de Decisão

Vai-se agora mostrar como o algoritmo ID3 pode extrair regras a partir de árvores de decisão representando projectos de investimento. Usar-se-á um exemplo elementar, com o fim de introduzir a metodologia. Mais tarde descrever-se-á um caso próximo da vida real. Nessa altura explicar-se-á como o ID3 quantifica a robustez de cada decisão.

Numa árvore de decisão, a verosimilhança de cada possível desenlace calcula-se multiplicando as probabilidades associadas a todos os atributos ao longo do percurso que conduz a esse desenlace. Esta verosimilhança tem uma interpretação intuitiva: Ela é a frequência esperada (relativa) de esse desenlace quando, hipoteticamente, o projecto se repete muitas vezes nas mesmas condições.

O conjunto de desenlaces e respectivas frequências esperadas e atributos é aquilo que o algoritmo ID3 vai transformar. Como resultado, obtém-se uma árvore de regras. Para este estudo é importante que se continue a observar o resultado sob a forma de árvore (estrutura hierárquica) e não apenas como uma colecção de regras.

O exemplo simples de árvore de decisão com o qual se ilustrará esta metodologia está representado esquematicamente na figura 3. O atributo “Programa” é uma decisão. O gestor pode escolher entre os programas “A” ou “B”. O atributo “Procura” é um acontecimento incerto que não depende do gestor (uma “jogada da Natureza”). No alto da árvore podem ver-se os desenlaces possíveis com a forma de um Valor Actual Líquido (NPV). Como é prática corrente, só interessa saber se os NPV são positivos ou negativos.

Vamos descrever o uso do ID3 por etapas. Elas são:

VALOR ACTUAL LIQUIDO:	PROCURA	PROGRAMA	FREQUÊNCIA
Positivo (+)	Alta	A	15
Positivo (+)	Baixa	A	35
Positivo (+)	Alta	B	15
Negativo (-)	Baixa	B	35

Tabela 3: O conjunto de desenlaces, seus atributos e frequências esperadas.

Transformar os desenlaces para que sejam nominais: O Valor Actual Líquido (NPV) de cada desenlace é considerado como podendo ser apenas “positivo” ou “negativo”. É esta a informação que importa quantificar em projectos de investimento. Recorda-se que só é recomendável o uso do ID3 quando desenlaces nominais fazem sentido.

Considerar a informação a-priori associada com os atributos: Caso não exista nenhuma informação deste tipo, os atributos que representam decisões são igualmente verosímeis a-priori. Portanto, iguais probabilidades devem ser atribuídas a cada um deles. Neste caso, estas probabilidades são $1/2$ pois há duas decisões possíveis.

Construir o conjunto de observações: A tabela 3 mostra este conjunto. Cada desenlace aparece associado com os seus atributos e com a respectiva verosimilhança, escrita como se fosse uma frequência. A verosimilhança foi calculada multiplicando as probabilidades associadas ao atributo “programa” pelas associadas ao atributo “Procura”. Multiplicando depois esta verosimilhança por um factor arbitrário, o mesmo para todos os desenlaces, obteve-se uma “frequência”. Portanto, tudo se passa como se existisse um conjunto de observações com tantos casos quantos os desenlaces, depois de multiplicados por um factor proporcional à sua verosimilhança. Na tabela 3, a coluna “frequência” foi obtida multiplicando por 100 a verosimilhança de cada desenlace.

Aplicar o algoritmo ID3 a este conjunto de observações: Neste caso muito simples, a primeira etapa de execução do ID3 consistiria em gerar duas tabelas de contingência: Desenlace por Programa e Desenlace por Procura. Depois, ir-se-ia calcular o ganho em informação cruzada de cada uma destas tabelas. E por fim usar-se-ia como raiz da árvore de regras em criação o atributo que aportasse um maior ganho em informação — uma maior causalidade — à previsão do desenlace. Tanto as referidas tabelas como uma medida aproximada do ganho, o Qui-Quadrado, são mostradas a seguir.

PROCURA	DESENLACE		PROGRAMA	DESENLACE	
	$NPV > 0$	$NPV < 0$		$NPV > 0$	$NPV < 0$
Alta	30	0	A	50	0
Baixa	35	35	B	15	35
Qui-Quadrado: 23.1			Qui-Quadrado: 53.8		

Portanto o atributo Programa explica o desenlace melhor do que o atributo Procura. Sendo assim, o programa figurará como raiz da árvore de regras. Depois, considerar-se-iam separadamente os dois casos do programa ser A ou B e repetir-se-ia o processo para cada um deles. O resultado é esta estrutura lógica:

Se Programa é A então o desenlace é NPV Positivo.

Se Programa é B então:

Se Procura é Alta então o desenlace é NPV Positivo.

Se Procura é Baixa então o desenlace é NPV Negativo.

O ID3 transformou uma árvore de decisão numa árvore de regras. O atributo mais próximo da raiz é o de maior peso causal. Ver-se-á na próxima secção qual o interesse desta transformação.

3.2 Quantificando a Robustez

Nesta secção explorar-se-ão as potencialidades do ID3 como pós-processador de árvores de decisão. Dar-se-á relevo à quantificação da robustez. Para este fim citar-se-á um caso concreto descrito na literatura.

3.2.1 A Árvore de Decisão para “Prism Paints Inc.”

O interesse da indução de regras como pós-processador ou interpretador de árvores de decisão em projectos de investimento será testado com uma versão do problema descrito por Magee [22], que é típico. Citam-se alguns períodos da sua introdução.

“Prism Paints vai precisar de decidir o destino a dar a uma das suas fábricas, muito pequena e incapaz de suprir a qualidade requerida pelo mercado. Existe

uma considerável controvérsia entre os gestores sobre a melhor acção — modernizar a operação construindo boas instalações no local, fechar a fábrica e começar tudo de novo noutra local. (...)”

“Existem 3 cenários operativos que parecem promissores:

Programa A: Modernizar a fábrica e também expandir noutras locais. Este é o menos dispendioso dos três programas supondo uma procura anual baixa (inferior a uma margem cujo valor é conhecido).

Programa B: Fechar a fábrica em questão e expandir noutra sítio. Este programa é o menos dispendioso quando a procura anual é média e se situa entre dois valores conhecidos.

Programa C: Modernizar e também aumentar a fábrica em questão: Este programa é o menos dispendioso quando a procura é alta a ponto de subir acima de uma cota cujo valor é conhecido.”

“As alternativas enunciadas envolvem investimentos significativamente diferentes e parecem conduzir a posições marcadamente distintas em economia operativa também. Subjacente à controvérsia sobre o que fazer, está a incerteza quanto à procura futura do produto. (...)”

“A evidência anterior e as previsões permitem estimar quão verosímil é cada nível de procura: Sabe-se qual a probabilidade de que a procura venha a situar-se no nível baixo, médio ou alto, para cada uma das três etapas do projecto de expansão. Estas previsões aparecem condicionadas à procura na etapa precedente, isto é, a procura em cada etapa depende da procura em etapas anteriores.” (ver a tabela 4 na página 34 onde se mostram estas probabilidades).

“Os Meios Libertos são calculados a partir dos tipos de marketing, operações, engenharia e análise financeira mencionados em outros documentos. No caso de Prism Paints, um estudo da distribuição mostra que o lucro operativo relativo de cada um dos três possíveis projectos pode ser expresso em Meios Libertos Líquidos anuais em termos dos três níveis da procura. (...)”

“A administração de Prism Paints estabeleceu uma taxa de rendibilidade desejável, ou um custo de capital que ronda os 14% anuais. Esta é a taxa a ser usada para obter os Valores Actuais dos meios libertos previsionais, sempre que se pretender compará-los.”

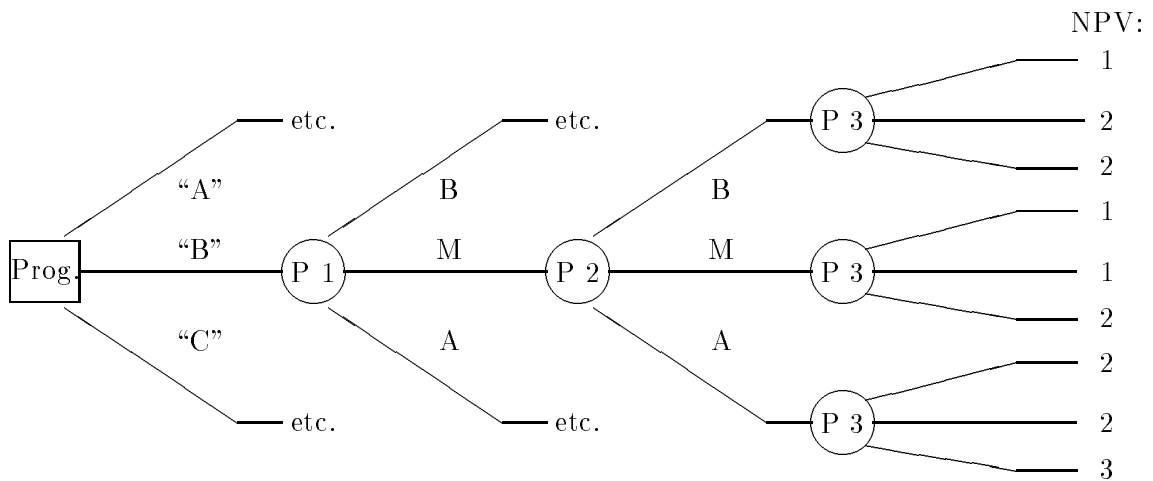


Figura 4: A parte central da árvore de decisão para Prism Paints Inc.

PRISM PAINTS INC.		Ano 2			Ano 3		
		Se a procura no ano 1 for:			Se a procura no ano 2 for:		
Nível da Procura	Ano 1	Baixa	Média	Alta	Baixa	Média	Alta
Baixa	.50	.35	.15	0	.20	.05	0
Média	.43	.50	.45	.40	.60	.35	.20
Alta	.07	.15	.40	.60	.20	.60	.80

Tabela 4: Probabilidades associadas às procuras Baixa, Média, e Alta no projecto para Prism Paints Inc.

Conhecido o investimento que cada alternativa supõe, e estabelecida a taxa de rendibilidade, podem expressar-se todos os desenlaces sob a forma de um Valor Actual Líquido (NPV).

Discretização do NPV: Discussão. Sob condições normais, é raro que um projecto ofereça apenas perspectivas de NPV positivos. Tal situação conduziria ao aparecimento de uma dura concorrência a muito curto prazo. Os lucros excessivamente fáceis são raros e nunca duram muito. Por isso, em modelos de decisão sequenciais que contemplem vários anos, as alternativas reais são o escolher entre cenários que conduzem a NPV ligeiramente positivos e aqueles que levam a NPV negativos. Esta característica é específica da análise de projectos de investimento. Não se aplica a outros tipos de modelação sequencial.

Assim, o NPV parece ser uma variável muito apropriada para algoritmos como os de indução de regras, requerendo desenlaces de tipo nominal. Em projectos de investimento apenas dois desenlaces são importantes, e podem definir-se claramente.

Nas presentes simulações, o NPV foi autorizado a ter três valores: “Negativo”, “Positivo”, e “Lucros Extra”. Isto torna o modelo mais complexo, o que parece desejável ao rastrear a robustez de cada cenário.

Diferenças em relação ao projecto original: No nosso caso, os NPV não foram copiados do texto original: construiu-se um modelo contendo todos os parâmetros deste problema e introduziu-se nele um grau regulável de aleatoriedade. Obtiveram-se assim várias colecções de NPV simulados, e respectivos atributos, cada uma constituindo um diferente cenário. O modelo tem provisões para evitar dominância absoluta e outras situações em que a decisão seria por demais óbvia.

Outro ponto em que o modelo é diferente do original é o facto de os três programas em questão, A, B, e C, permanecerem sem modificação ao longo do tempo: Não se considera a possibilidade de saltos de um programa para outro a meio do projecto. A razão para esta simplificação é o desejo de evitar, neste estudo, situações de difícil interpretação no que respeita à robustez de cada possível cenário.

A procura é modelada como no original (Baixa, Media e Alta) e a cada caso corresponde uma verosimilhança expressa em probabilidades. Assim, a árvore de decisão resultante tem $3^4 = 81$ possíveis caminhos ou “folhas terminais”, com desenlaces (NPV) de três possíveis tipos. A figura 4 (página 34) documenta o aspecto desta árvore de decisão.

Não foram consideradas tendências a-priori: Os três programas objecto de escolha apresentam-se, à partida, como igualmente verosimilhantes — na ausência de outra informação. A verosimilhança de cada desenlace determina a frequência de cada caso, como se viu atrás.

3.2.2 Discussão dos Resultados

Limitaremos esta discussão à análise das árvores de regras obtidas a partir de dois dos cenários simulados. Estas árvores encontram-se na figura 5, na página 36 e na figura 6, na página 39. A tabela 5 mostra os 81 NPV e seus atributos correspondentes ao primeiro dos cenários gerados pelo simulador (os que originaram a árvore de regras da figura 5). Estes valores são semelhantes aos do artigo original.

O segundo dos cenários distingue-se do primeiro pela maior aleatoriedade que foi autorizada a entrar no modelo e pelo realce do peso do atributo Programa para o desenlace. Trata-se também de um cenário menos pessimista.

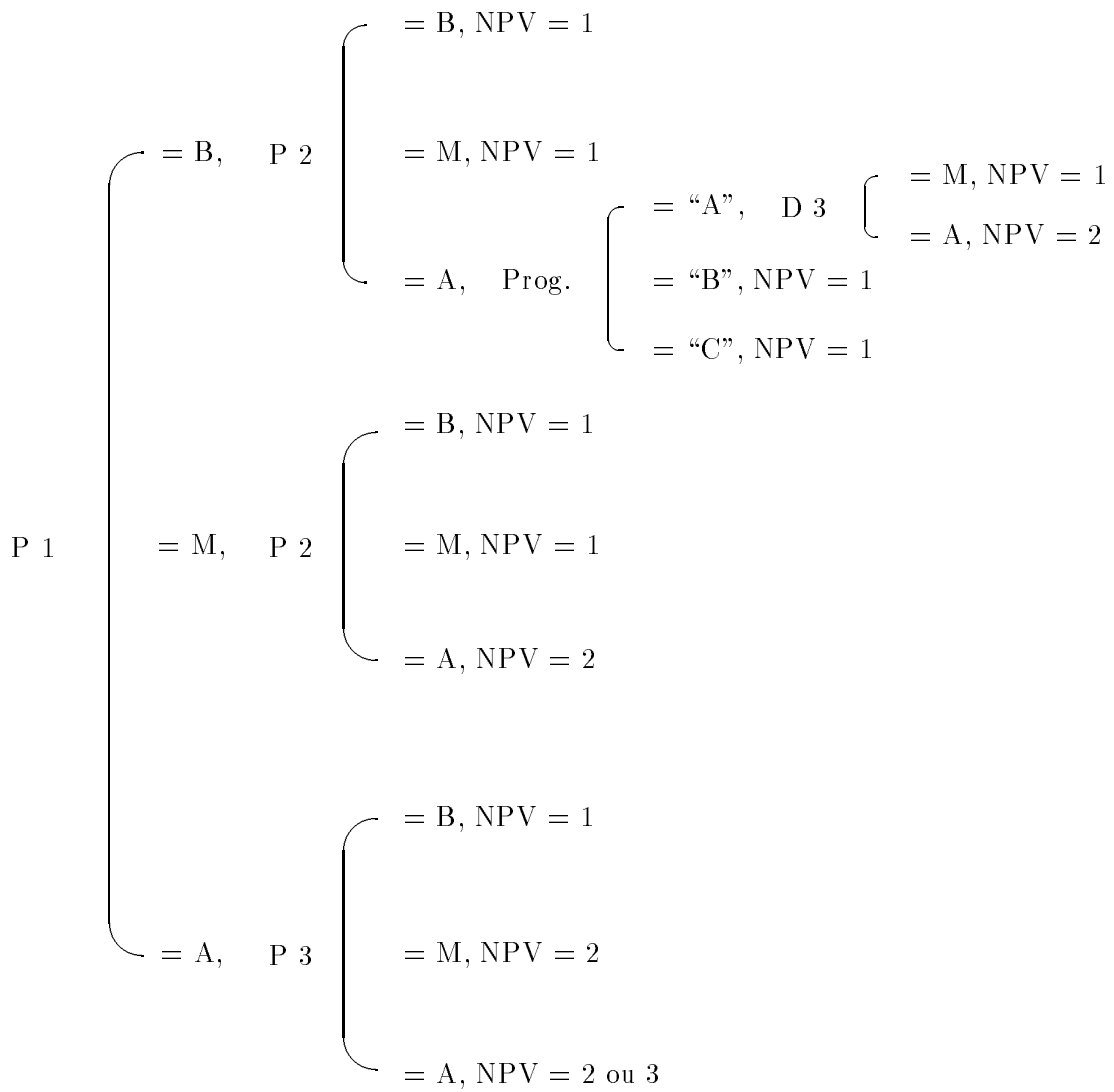


Figura 5: Prism Paints Inc.: A árvore de regras obtida pelo algoritmo ID3 a partir do primeiro cenário simulado. NPV= 1 é o negativo; NPV= 2 é o positivo; NPV= 3 é o de “lucros extra”. P 1, P 2 e P 3 são a procura nos anos 1, 2 e 3.

PROCURA			VALOR ACTUAL LÍQUIDO		
Ano 1	Ano 2	Ano 3	Programa 1	Programa 2	Programa 3
Alta	Alta	Alta	Positivo	Lucros extra	Lucros extra
Alta	Alta	Média	Positivo	Positivo	Positivo
Alta	Alta	Baixa	Positivo	Negativo	Negativo
Alta	Média	Alta	Positivo	Positivo	Positivo
Alta	Média	Média	Positivo	Positivo	Positivo
Alta	Média	Baixa	Negativo	Negativo	Negativo
Alta	Baixa	Alta	Positivo	Negativo	Negativo
Alta	Baixa	Média	Negativo	Negativo	Negativo
Alta	Baixa	Baixa	Negativo	Negativo	Negativo
Média	Alta	Alta	Positivo	Positivo	Positivo
Média	Alta	Média	Positivo	Positivo	Positivo
Média	Alta	Baixa	Negativo	Negativo	Negativo
Média	Média	Alta	Positivo	Positivo	Positivo
Média	Média	Média	Positivo	Negativo	Negativo
Média	Média	Baixa	Negativo	Negativo	Negativo
Média	Baixa	Alta	Negativo	Negativo	Negativo
Média	Baixa	Média	Negativo	Negativo	Negativo
Média	Baixa	Baixa	Negativo	Negativo	Negativo
Baixa	Alta	Alta	Positivo	Negativo	Negativo
Baixa	Alta	Média	Negativo	Negativo	Negativo
Baixa	Alta	Baixa	Negativo	Negativo	Negativo
Baixa	Média	Alta	Negativo	Negativo	Negativo
Baixa	Média	Média	Negativo	Negativo	Negativo
Baixa	Média	Baixa	Negativo	Negativo	Negativo
Baixa	Baixa	Alta	Negativo	Negativo	Negativo
Baixa	Baixa	Média	Negativo	Negativo	Negativo
Baixa	Baixa	Baixa	Negativo	Negativo	Negativo

Tabela 5: Prism Paints Inc: Valores originais do NPV para cada programa em função da procura nos três anos que dura o projecto.

Como poderão os gestores beneficiar com este pós-processamento de árvores de decisão? Claramente, de duas formas: A indução de regras, ao eliminar redundâncias, produz redução da complexidade. Como consequência, dá-se também uma quantificação da robustez de cada alternativa. A redução da complexidade é capaz de revelar algumas características dinâmicas dos problemas sequenciais de decisão financeira.

Vejamos estes pontos com maior detalhe.

1. Redução da complexidade: Os 81 possíveis percursos, tal como se encontram nas árvores de decisão, originaram um número menor de regras depois de aplicado o algoritmo ID3. A árvore de regras é mais simples do que a árvore de decisão. E é de mais fácil interpretação também. Muita da redundância foi removida. A aplicação do ID3 à primeira das simulações (ver página 36) apenas gerou 12 percursos (uma redução de 85%) e a segunda (ver página 39) gerou 40 percursos (50% menos). A redução da complexidade obtida pelo ID3 é inversamente proporcional à aleatoriedade presente na relação atributos-desenlaces.

É esperável que uma interpretação mais fácil de árvores de decisão por parte dos gestores seja agora possível. Ela ficar-se-á a dever apenas a esta redução de complexidade. Como é sabido, o aspecto intrincado das árvores de decisão é uma das barreiras à sua utilização.

2. Quantificação da Robustez: Como foi visto, numa árvore de decisão existem *atributos de decisão*, que correspondem a escolhas que os gestores podem fazer; e existem também *atributos da natureza* correspondendo a acontecimentos incertos que os gestores não dominam. O algoritmo ID3 mostra-se capaz de hierarquizar os atributos de acordo com a sua importância para o desenlace. A posição relativa que ocupa cada um destes tipos de atributos na nova árvore de regras irá mostrar o grau de controlo que o gestor tem sobre o processo sequencial que se desenrola.

Na secção 3.1.2 viu-se como as expectativas, quantificadas sob a forma de probabilidades, podem ser traduzidas de modo a reflectir importância causal. O ID3, ao escalar atributos, está a indicar qual a relevância relativa de cada um deles para o desenlace. Os atributos mais relevantes serão os mais próximos da raiz na árvore de regras.

Quando, entre os atributos relevantes, se encontra um atributo de decisão, o gestor tem o poder de determinar o desenlace. Pelo contrário, quando os atributos relevantes são jogadas da natureza, o gestor tem pouco poder sobre o projecto.

Uma decisão que se mostra pouco relevante para o desenlace final indica também um

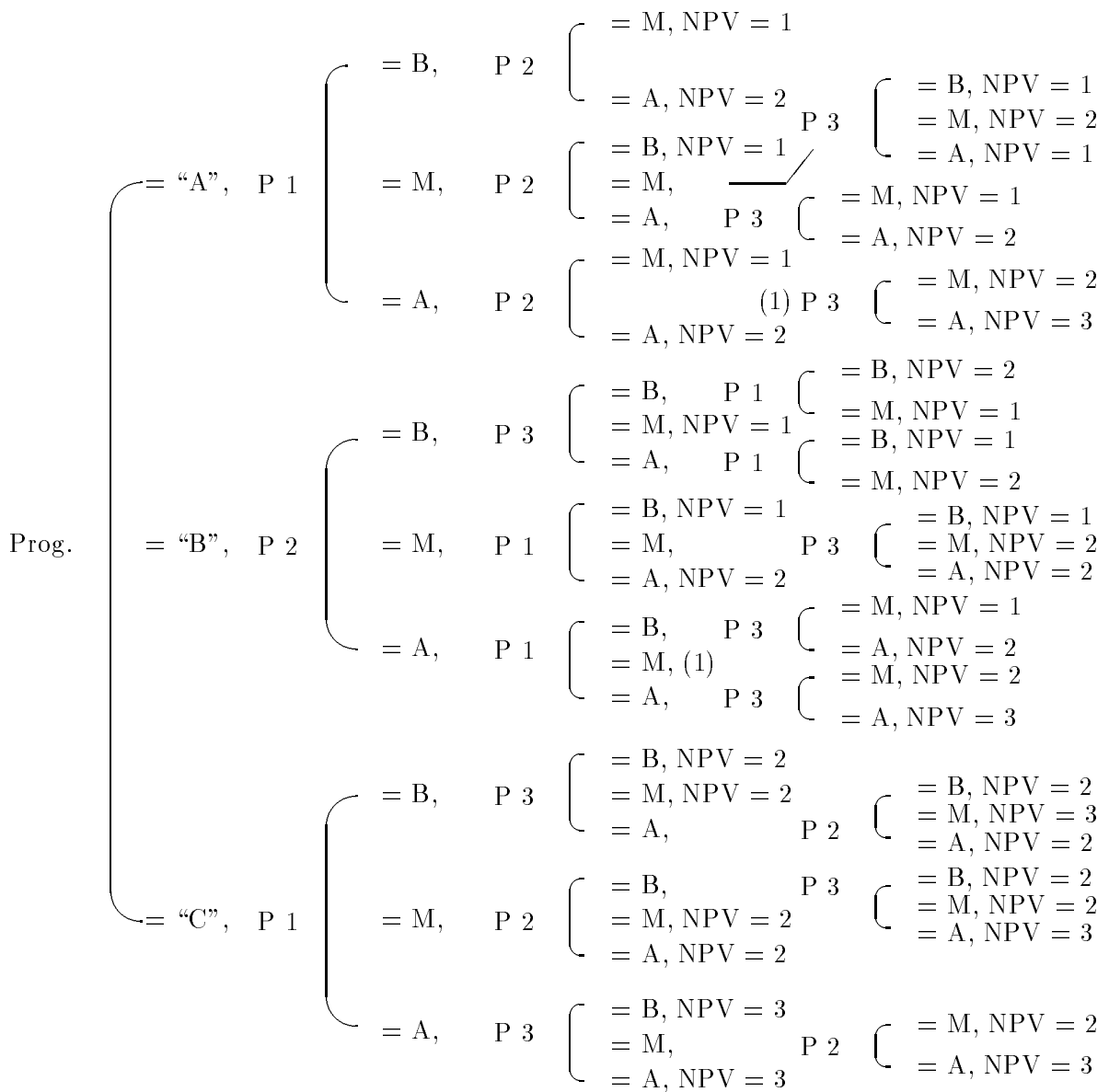


Figura 6: Prism Paints Inc.: A árvore de regras obtida pelo algoritmo ID3 a partir do segundo cenário simulado. NPV= 1 é o negativo; NPV= 2 é o positivo; NPV= 3 é o de "lucros extra". P 1, P 2 e P 3 são a procura nos anos 1, 2 e 3.

percurso pouco robusto: O gestor, faça o que fizer, enfrenta um desenlace pré-determinado. O seu poder de intervenção é pequeno. Se fôr preciso, no decorrer do projecto, introduzir alterações, é de esperar que tais alterações tenham pouca eficácia na determinação do desenlace. No extremo oposto, quando um atributo de decisão ocupa um lugar perto da raiz da árvore de regras produzida pelo ID3, tal decisão mostra-se como robusta: O desenlace vem determinado por factores que estão na mão do gestor durante o processo, e que ele pode alterar — produzindo modificações eficazes.

Na primeira simulação, a escolha do programa — o atributo de decisão — encontra-se geralmente longe da raiz (ver página 36). Uma vez que a questão é decidir qual dos programas, A, B ou C, deve ser escolhido, pode dizer-se que o ID3, através da árvore de regras, mostra aos gestores que a decisão é irrelevante para o resultado. O gestor tem muito pouco controlo sobre o projecto.

Eventualmente, depois de analisar a árvore de regras da página 36, o gestor poderia preferir um percurso menos óptimo mas mais robusto. Um percurso onde visse que o seu poder de intervenção seria maior. E também estaria iuninteressado no facto do atributo mais causal ser a procura no ano um. Perceberia que, para Prism Paints Inc., o ano crítico seria o primeiro. Também neste caso, uma baixa procura no primeiro ano torna o programa importante. Caso contrário ele não importa nada. Ao reparar que a escolha do programa só parece pesar quando a procura é baixa, um gestor concluiria que as alternativas postas ao investimento não estão proporcionadas à procura. O ID3 conseguiu pôr à mostra uma possível incoerência formal do projecto.

Quanto ao cenário que originou a segunda árvore de regras (página 39), a escolha do programa é o atributo com mais peso para o desenlace. Está-se perante uma decisão robusta. O gestor sabe que as suas decisões têm peso no processo.

3. Quantificação da Complexidade da Incerteza: Nas árvores de regras obtidas, alguns ramos conduzem a zonas “boscosas”, com muitos galhos, enquanto que outras conduzem rapidamente a um conjunto simples de desenlaces. Os primeiros, desdobram-se em atributos até atingirem um NPV final. Os últimos, pelo contrário, levam a um NPV dependente de poucos atributos.

Até agora, pouca atenção tem sido dada na literatura ao problema da *complexidade da incerteza* em problemas de decisão. O ID3, devido à sua capacidade eliminadora de redundância, parece capaz de assinalar esta qualidade.

Os percursos simples são atractivos para os gestores. A incerteza em percursos simples é causada por poucos atributos. É uma incerteza de pequena *dimensão*. Existem menos graus de liberdade a considerar. Na primeira das simulações apresentadas, uma procura elevada ou média no primeiro ano conduz o projecto para uma zona complexa. Os NPV positivos só se alcançam através de percursos complicados enquanto que os NPV negativos são simples de alcançar. Isto é um aviso útil para os gestores.

3.3 Sumário

A experiência apresentada neste capítulo mostra que a indução de regras pode ser útil como um pós-processador para modelos sequenciais de decisão. A capacidade de redução de complexidade inerente a algoritmos como o ID3 é bem-vinda e é correcta quando usada em tais modelos. A forma como o ID3 consegue ordenar os atributos segundo o peso da sua capacidade para determinar o desenlace dá aos gestores uma ferramenta interessante na avaliação das características dinâmicas dos projectos. Nomeadamente, eles podem avaliar a robustez das decisões pela posição relativa dos atributos na árvore de regras. É também possível medir a complexidade da incerteza de cada percurso.

Capítulo 4

Conclusões

Este estudo descreveu uma nova ferramenta para ser usada em projectos de investimento conjuntamente com árvores de decisão. Trata-se de um pós-processador capaz de reduzir a redundância existente nesse tipo de modelos, de modo a conseguir uma quantificação da robustez, entre outras medidas úteis para os gestores.

Na primeira parte deste estudo explicou-se o que se entendia por robustez e as razões pelas quais o algoritmo ID3 para indução de regras não era adequado como ferramenta de modelação estatística de uso geral. A robustez tem sido apresentada como um critério de decisão, candidato a figurar, junto com o óptimo, em modelos financeiros sequenciais. Porém, a simples flexibilidade operativa descrita por Rosenhead *et al.* [32] não satisfaz as exigências dos projectos de investimento.

Uma medida do poder ou peso das decisões de desencadear o desenlace esperado, independente da estrutura do modelo, parece preferível já que incorpora toda a informação disponível sobre o futuro e não depende do grau de sofisticação pretendido.

Por último, este trabalho explorou o uso do algoritmo ID3 como pós-processador de árvores de decisão em projectos de investimento. Mostrou-se como o ID3 encontra nos dados a estrutura que supõe existir. Sublinharam-se as vantagens decorrentes da existência, neste tipo de modelo, de desenlaces nominais ou ordinais. E finalmente, ilustrou-se com um exemplo conhecido o uso do ID3 em árvores de decisão e a forma como este algoritmo mostra aos gestores a robustez e a dimensão da complexidade de cada percurso.

Bibliografia

- [1] W. Ashby. Constraint analysis of many-dimensional relations. *General Systems Yearbook*, 9:99–105, 1964.
- [2] W. Ashby. Measuring the internal information exchange in a system. *Cybernetica*, 8(1):5–22, 1965.
- [3] G. Barnard and D. Cox. *The Foundations of Statistical Inference*. Methuen, London, 1962.
- [4] Y. Bishop, S. Fienberg, and P. Holland. *Discrete Multivariate Analysis*. MIT Press, Cambridge, MA., 1978.
- [5] H. Braun and J. Chandler. Predicting stock market behaviour through rule induction. In R. Trippi and E. Turban, editors, *Investment Management - Decision Support and Expert Systems*, pages 78–93. Boyd and Fraser, Cambridge, MA, USA, 1990.
- [6] L. Breiman, J. Freidman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth International, California, 1984.
- [7] R. Cavallo and G. Klir. Reconstructability analysis of multi-dimensional relations: A theoretical basis for computer-aided determination of acceptable systems models. *International Journal of General Systems*, 5:143–171, 1979.
- [8] G. Cestnik, I. Kononenko, and I. Bratko. Assistant 86: A knowledge elicitation tool for sophisticated users. In I. Bratko and N. Lavrac, editors, *Progress In Machine Learning*. Sigma Press, London, England, 1987.
- [9] R. Conant. Structural modelling using a simple information measure. *International Journal of Systems Sciences*, 11(6):721–730, June 1979.

- [10] R. Conant. Detection and analysis of dependency structures. *International Journal of General Systems*, 7:81–91, 1981.
- [11] J. Friend and W. Jessop. *Local Government and Strategic Choice*. Tavistock, London, 1969.
- [12] I. Good. Maximum entropy for hypothesis formulation especially for multidimensional contingency tables. *Annals of Mathemat. Statistics*, 34:911–934, 1963.
- [13] S. Gupta and J. Rosenhead. Robustness in sequential investment decisions. *Management Sciences*, 15:B–18, 1968.
- [14] B. Hertz and H. Thomas. *Risk Analysis and its Applications*. Wiley, New York, 1983.
- [15] F. Hespos and P. Strassman. Stochastic decision trees for the analysis of investment decisions. *Management Science*, 11:224–259, 1965.
- [16] E. Hunt, J. Marin, and P. Stone. *Experiments in Induction*. Academic Press, New York, 1966.
- [17] E. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4(3):227–341, 1968.
- [18] G. Klir. Identification of generative structures of empirical data. *International Journal of General Systems*, 3:89–104, 1976.
- [19] Y. Kodratoff and M. Manago. Generalisation and noise. *Internatrional Journal of Man-Machine Studies*, 27:181–204, 1987.
- [20] K. Krippendorff. An algorithm for identifying structural models of multi-variate data. *International Journal of General Systems*, 7:63–79, 1981.
- [21] S. Kullback. *Information Theory and Statistics*. Wiley, 1959.
- [22] J. Magee. How to use decision trees in capital investment. *Harvard Business Review*, pages 79–96, September 1964.
- [23] W. McGill. Multivariate information transmission. *Psychometrika*, 19(2):97–116, June 1954.

- [24] J. Mingers. Expert systems — rule induction with statistical data. *Journal of the Operational Research Society*, 38(1):39–47, 1986.
- [25] J. Mingers. Rule induction with statistical data — a comparison with multiple regression. *Journal of the Operational Research Society*, 38(4):347–351, 1986.
- [26] T. Niblett. Constructing decision trees in noisy domains. In I. Bratko and N. Lavrac, editors, *Progress In Machine Learning*. Sigma Press, London, England, 1987.
- [27] J. Quinlan. Discovering rules from large collections of samples — a case study. In D. Michie, editor, *Expert Systems in the Micro Electronic Age*. Edimburgh University Press, 1979.
- [28] J. Quinlan. The effect of noise on concept learning. In R. Michalsky, T. Carbonell, and T. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*. Morgan Kaufman, Los Altos, 1986.
- [29] R. Quinlan. Simplifying decision trees. *Internatrional Journal of Man-Machine Studies*, 27:221–234, 1987.
- [30] P. Race and R. Thomas. Rule induction in investment appraisal. *Journal of the Operational Research Society*, 39(12):1113–1123, 1988.
- [31] J. Rosenhead. Planning under uncertainty. *Journal of the Operational Research Society*, 31:209–216, 1980.
- [32] J. Rosenhead, M. Elton, and S. Gupta. Robustness and optimality as criteria for strategic decisions. *Operational Research Quarterly*, 23(4):413–431, 1972.
- [33] S. Stevens. On the theory of scales of measurement. *Science*, 103:677–680, 1946.
- [34] R. Wipperfurth. A note on the equivalent risk class assumption. *Engineering Economist*, Spring:13–22, 1966.