

**Seleção Probabilística – Melhorando a Eficiência
da Conferência Aduaneira**

2.º Lugar

MARCOS ANTONIO CARDOSO FERREIRA

Engenharia Mecânica - ITA - São José dos Campos - SP
AFRF - Aduaneira
DRF São José dos Campos - SP

SELEÇÃO PROBABILÍSTICA – MELHORANDO A EFICIÊNCIA DA CONFERÊNCIA ADUANEIRA

RESUMO

O Siscomex (Sistema Integrado de Comércio Exterior) utiliza uma técnica chamada Seleção Parametrizada para selecionar ou liberar operações de importação para verificação e fiscalização. A Seleção Parametrizada apresenta um desempenho considerado insatisfatório porque, com certa frequência, seleciona Declarações de Importação sem infração, enquanto libera outras declarações contendo infrações.

Objetivos básicos

Perante essa situação, os objetivos desse trabalho são: propor uma outra técnica que possa ser integrada ao sistema atual, com o propósito de melhorar a seleção de Declarações de Importação para conferência, e fornecer à Administração Aduaneira uma ferramenta de avaliação probabilística quanto a ocorrência de infrações aduaneiras em despachos aduaneiros de importação.

Metodologia utilizada

Esse trabalho foi precedido de uma análise de várias técnicas

na área de Inteligência Artificial, visando escolher a mais adequada ao problema. Entre as técnicas consideradas, destacam-se: Sistemas de Aprendizado por Reforço, Redes Neurais, Fuzzy Logic e Redes de Crença. Esta última é muito utilizada em sistemas inteligentes de diagnóstico médico, os quais se assemelham muito ao problema de identificar infrações, utilizando as informações prestadas na Declaração de Importação (DI).

A técnica aqui proposta é denominada Seleção Probabilística e utiliza a teoria de Redes de Crença e o potencial de lançamento tributário de cada infração como elementos para o auxílio na tomada de decisão.

A metodologia utilizada para avaliar o desempenho da técnica proposta pode ser descrita pelas seguintes etapas: dominar a teoria que fundamenta o uso de Redes de Crença; delimitar as condições para sua aplicação prática e obter um modelo representativo do problema; implementar o modelo anterior em uma linguagem de programação e simulá-lo com dados de operações reais; e analisar os resultados obtidos e compará-los com a técnica atual.

Adequação do trabalho aos critérios de julgamento

I - Relação custo versus benefício

É difícil mensurar os ganhos exatos que sua adoção permitiria obter, embora a técnica tenha sido simulada com dados reais e exibido bons resultados.

A Seleção Probabilística fornece ao sistema a vantagem de efetuar a escolha das Declarações de Importação (DIs) com um espectro mais amplo e adaptável a novos casos, pois ele passa a contar com capacidade de aproximação e aprendizado. Com o tempo, a fiscalização aduaneira perceberá uma melhor qualidade na seleção, já que mais DIs contendo irregularidade serão fiscalizadas e menos DIs com irregularidades serão liberadas. Acrescente-se o fato de que a administração aduaneira disporá de uma ferramenta de avaliação probabilística que aperfeiçoará o controle submetido às importações.

II - Aumento de produtividade

A simulação da técnica mostrou que, para os dados coletados,

ela foi capaz de descobrir 72,5% das DIs com infração, com uma perda hipotética de crédito tributário em torno de R\$ 1.143.438,99. A simulação da Seleção Parametrizada teria uma perda hipotética de R\$ 3.628.757,77 e um acerto de 40%. Ainda que o desempenho real da Seleção Parametrizada possa ser melhor ou pior que aquele apresentado nos testes, já que não foi possível replicar fielmente as condições existentes à época de registro de cada DI, ela dificilmente alcançaria o desempenho da Seleção Probabilística, pois esta possui capacidade de avaliação qualitativamente melhor.

A Seleção Probabilística, ao ponderar os casos com maior potencial de irregularidade, seleciona uma quantidade menor de casos inócuos e assegura maiores lançamentos de crédito tributário.

III - Viabilidade de implementação

Para avaliar a viabilidade técnica da solução aqui proposta, o sistema de seleção de DIs foi reproduzido da forma mais fiel possível. Dois conjuntos contendo dados de operações reais foram coletados e extraídos: o maior dos conjuntos foi usado para a modelagem, e o outro foi utilizado para avaliar a Seleção Probabilística. A implementação não provoca impacto nas unidades aduaneiras, e o tempo de processamento das probabilidades do modelo foi baixo, em torno de 1,3 segundo, em um microcomputador Pentium III.

Tecnicamente, o custo da implantação dessa proposta pode ser considerado baixo, já que seriam acrescentados ao Siscomex dois novos bancos de dados de estrutura relacional, um com 10 tabelas e outro com 4 tabelas, e seriam adicionados programas com as seguintes finalidades: registrar as infrações cometidas; efetuar a inferência probabilística; estimar o potencial de lançamento de crédito tributário; e proceder à escolha com base no critério de ordenação das DIs.

Avalizada pelo gestor do sistema, a decisão de implementação deve ser precedida de uma análise dos requisitos exigidos e da consecução de um projeto de *software* de acordo com as normas atuais.

IV - Valorização do servidor

O elemento humano é fundamental nessa proposta, pois ele é o único capaz de alimentar o sistema com informações confiáveis sobre

as infrações aduaneiras. Cada servidor aduaneiro deve estar ciente de que o relato preciso das infrações aduaneiras cometidas será usado por outras unidades, além da sua própria. Agindo comunitariamente, ele receberá em troca a experiência de outras pessoas e aprimorará suas habilidades para identificar infrações. O que a Seleção Probabilística faz é avaliar esse conhecimento sobre as infrações aduaneiras, tornando-o valioso para melhorar a qualidade da seleção e, conseqüentemente, do trabalho fiscal. Ela consegue transformar informações em conhecimento coletivo da organização e dele extrair decisões úteis.

V - Melhoria da qualidade do serviço prestado

Em relação ao cumprimento dos objetivos gerais do programa de trabalho da SRF para os anos de 2003 e 2004, o presente trabalho está vinculado aos seguintes:

Objetivo 1 – Subsidiar a formulação da política tributária e de comércio exterior.

Isso é proporcionado pela apresentação de uma ferramenta capaz de melhorar o controle aduaneiro, avaliando DIs com uma abordagem probabilística.

Objetivo 3 – Intensificar a atuação da SRF no combate ao crime organizado.

A técnica proposta permite ao sistema discernir as tentativas de burlar os controles aduaneiros, por exemplo, identificando as infrações aduaneiras de contrafação. Esse assunto tem causado muito atrito com outros países.

Objetivo 7 – Aprimorar a qualidade e a produtividade do trabalho fiscal;

Objetivo 8 – Aumentar a eficácia da vigilância e da repressão aos ilícitos aduaneiros; e

Objetivo 9 – Simplificar, padronizar e agilizar o controle aduaneiro.

Conferindo o mesmo número de DIs, a Seleção Probabilística provê a possibilidade de encontrar mais infrações, aumentando a perspectiva de recuperação de crédito, o que aumenta a produtividade, agiliza o controle aduaneiro e intensifica a repressão aos ilícitos.

Objetivo 14 – Aprimorar a política de gestão da informação e de infra-estrutura de tecnologia.

Abordando de forma diferente a informação hoje existente, pode-se obter conhecimento até então desprezado. O foco gerencial para seleção de DIs adquire uma componente dinâmica pela atualização automática das probabilidades.

VI - Promoção da justiça fiscal e social dos tributos

Indiretamente o sistema cumpre o seu papel social, pois aumenta a chance de encontrar irregularidades que poderiam, por exemplo, significar concorrência desleal com os produtos nacionais e acarretar desemprego ou a introdução de produtos contrafeitos ou prejudiciais à saúde.

O trabalho aqui desenvolvido vem ao encontro de educar o contribuinte para o cumprimento voluntário das obrigações tributárias. Entre as variáveis utilizadas para avaliar a probabilidade de ocorrência de alguma infração aduaneira em uma DI está aquela que identifica o importador. Assim, o próprio importador perceberá que o cumprimento freqüente e diligente das normas que regem a importação diminuirá o risco potencial de sua seleção.

A monografia conta com 6 capítulos. O primeiro faz uma introdução dos pontos principais do problema de seleção de DIs. O segundo apresenta a teoria de Redes de Crença. O terceiro relata as etapas necessárias para obter um modelo de acordo com a teoria que representasse o problema de seleção. O quarto apresenta as linhas gerais para implementação da técnica proposta. O quinto apresenta os resultados obtidos da simulação de um protótipo com dados reais. O sexto efetua comentários e faz conclusões sobre o presente trabalho.

SELEÇÃO PROBABILÍSTICA – MELHORANDO A EFICIÊNCIA DA CONFERÊNCIA ADUANEIRA

CAPÍTULO 1

INTRODUÇÃO

Há muito tempo a Aduana Brasileira vem adotando uma política de informatização de seus procedimentos com o objetivo de enfrentar a crescente demanda pelos serviços aduaneiros. O Sistema Integrado de Comércio Exterior (Siscomex) é um dos exemplos mais conhecidos desse esforço governamental, e por meio dele são automatizadas as operações de importação e exportação, além de atuar como uma plataforma que concentra e sincroniza as ações de todos os órgãos governamentais envolvidos.

O Siscomex utiliza uma técnica chamada Seleção Parametrizada para selecionar ou liberar operações de importação para verificação e fiscalização. Esse instrumento é fundamental para direcionar o trabalho da fiscalização para as Declarações de Importação com suspeitas de irregularidades. Para efetuar a sua escolha, a Seleção Parametrizada compara informações declaradas nas Declarações de Importação com parâmetros fornecidos pela administração aduaneira. Em caso de coincidência, atribui um canal de conferência, o qual delimita

ta os procedimentos a serem executados pela fiscalização aduaneira.

Com o tempo, percebeu-se que a Seleção Parametrizada apresentava um desempenho considerado insatisfatório porque, com certa frequência, ela selecionava Declarações de Importação sem infração enquanto liberava outras declarações contendo infrações. Esse comportamento é explicado parcialmente pela forma usada para fixar os valores dos parâmetros; pela liberdade limitada para diferenciar ou ordenar as declarações selecionadas; por depender do julgamento humano; e por não aproveitar o histórico de infrações cometidas.

Este trabalho aplica-se no contexto do despacho aduaneiro de importação e tem como objetivos: propor uma solução que melhore a eficiência da seleção de Declarações de Importação (DIs) para a conferência aduaneira e aparelhar a Administração Aduaneira de uma ferramenta probabilística para avaliação do potencial de cometimento das infrações aduaneiras.

A avaliação da probabilidade de ocorrência de infração aduaneira utilizará os dados informados na Declaração de Importação (DI) registrada pelo importador.

A Declaração de Importação é composta de duas partes: a primeira, chamada informações ou dados gerais da DI; e outra, de informações específicas, denominada adição. As informações gerais contêm dados que compõem e identificam genericamente uma declaração de importação, como o nome do importador, número do conhecimento de carga e valores de frete e seguro. As informações específicas contêm dados como nome do fabricante ou produtor da mercadoria, a classificação e denominação desta, seu valor unitário e alíquotas de imposto aplicável.

O Siscomex, para descongestionar as unidades aduaneiras, aplica ao despacho aduaneiro de importação o conceito de canais de conferência aduaneira. Cada canal de conferência estipula o nível de fiscalização a ser efetuado. São quatro os canais de conferência: Verde, Amarelo, Vermelho e Cinza. O Canal Verde libera a mercadoria sem verificação. O Canal Amarelo destina-se à conferência apenas dos documentos que instruem o despacho aduaneiro. O Canal Vermelho destina-se, além da conferência documental, à verificação física da mercadoria. O Canal Cinza exige o mesmo grau de fiscalização do Canal Ver-

melho, acrescido do exame de valor aduaneiro da mercadoria e da investigação de indícios de interposição fraudulenta.

A Seleção Parametrizada pode atribuir um canal de conferência de duas maneiras: a primeira utiliza conjuntos de tipos de parâmetros armazenados com seus respectivos valores e canal de conferência. Quando os valores da DI coincidem com um conjunto, atribui-se o canal associado à DI; a segunda utiliza regras fixas, do tipo *Se Então*. Toda vez que informações do Despacho Aduaneiro tornam a condição *Se verdadeira*, o *Então* é acionado, sendo-lhe atribuído o respectivo canal de conferência.

A técnica aqui proposta foi voltada para aplicação no ambiente de importação, considerando sua importância à economia nacional. Diversas técnicas foram avaliadas, como, por exemplo, Sistemas de Aprendizado por Reforço, Redes Neurais, Fuzzy Logic e Redes de Crença (*Belief Networks*), também chamadas Redes Bayesianas (*Bayesian Networks*).

As Redes de Crença se caracterizam por representarem uma distribuição de probabilidade conjunta de várias variáveis que, entretanto, pode ser reduzida ao produto das probabilidades de cada variável condicionada às variáveis que lhe dão causa. Elas são utilizadas para estimar o estado de certas variáveis com base nos valores fornecidos de outras variáveis, consideradas evidências, bem como para achar a causa mais provável quando se conhecem apenas os efeitos gerados. É uma técnica apropriada nos casos em que se deseja decidir tendo apenas informações incompletas ou incertas.

O problema de diagnosticar a existência de uma infração aduaneira, tendo como base informações da DI, se assemelha muito a sistemas inteligentes de diagnóstico médico, os quais usam Redes de Crença e podem ser vistos nos trabalhos de Pradhan (1994) e Zhang (1995). Decidiu-se empenhar esforços para aplicar a teoria de Redes de Crença considerando a semelhança e a facilidade de lidar com conceitos probabilísticos.

Partindo do panorama geral apresentado, o trabalho foi desenvolvido com foco na obtenção de resultados práticos que permitissem a adoção da técnica aqui proposta. Para isso é feita uma explanação sobre a teoria de Redes de Crença no capítulo 2. Esse capítulo inicia-se com os conceitos básicos de probabilidade, introduz a finalidade da inferência probabilística e define o que são Redes de Crença. Também

discute o conceito de independência causal, fundamental para tornar tratável o problema do ponto de vista computacional.

O capítulo 3 descreve a modelagem do problema, considerando as condições para aplicação da Rede de Crença. A concepção do modelo final utiliza a teoria exposta no capítulo 2 e revela como o conceito de independência causal é empregado para atingir o objetivo proposto.

O capítulo 4 discute a implementação do modelo na prática e sua integração ao sistema atual. Estruturas de dados foram construídas para guardar as informações úteis da Rede de Crença, assim como programas para armazenar as infrações registradas, executar a inferência probabilística, estimar o potencial de lançamento de crédito tributário e ordenar e selecionar as DIs para conferência.

O capítulo 5 começa citando as limitações a que a solução está sujeita, analisa os resultados obtidos e comenta a eficiência da solução proposta.

Após perfazer todo esse caminho para cumprir o objetivo proposto, as conclusões do presente trabalho são feitas no capítulo 6.

CAPÍTULO 2

FUNDAMENTAÇÃO TEÓRICA

O conceito de probabilidade é natural para a maioria das pessoas e está intimamente ligado à observação de experimentos aleatórios, os quais podem ser definidos com experimentos na natureza que, mesmo efetuados sob condições uniformes, não repetem resultados idênticos. Exemplos de experimentos aleatórios podem ser: o lançamento de um dado, a jogada de uma moeda ou o sorteio de uma bola. A cada evento aleatório atribui-se um valor numérico entre 0 e 1, o qual demonstra a propensão para sua ocorrência.

A seguir é feito um esboço sobre a aplicação da Teoria de Probabilidade na modelagem de sistemas reais.

2.1 Aplicação da Teoria de Probabilidade

A análise de problemas de probabilidade geralmente tem como

foco principal uma determinada variável de interesse. Os cálculos são efetuados para determinar qual o valor de probabilidade dessa variável perante as condições estabelecidas e os valores observados das outras variáveis. Dessa forma, substitui-se a incerteza por uma medida numérica que representa quão próximo se está da afirmativa ou negativa de uma proposição. Para conhecimento, são relacionados a seguir os axiomas usados na Teoria de Probabilidade:

- i) os valores de probabilidade estão no intervalo entre 0 e 1, ou seja $0 \leq P(V) \leq 1$;
- ii) o evento certo tem $P(U)=1$, já o evento impossível tem probabilidade zero (0);
- iii) se X_1 e X_2 são eventos excludentes, então $P(X_1 \vee X_2) = P(X_1) + P(X_2)$.

Exemplificando, imagine um sistema automático de umidificação de verduras em um balcão refrigerado de um supermercado. Esse sistema visa manter a umidade do ambiente para que as verduras se mantenham com uma aparência de frescor. Embora o sistema não garanta a qualidade eterna das verduras, ele aumenta sobremaneira o período de vida do produto a um baixo custo. O aumento da umidade é feito pelo borrifo de água em pontos dispostos uniformemente acima das verduras. Um sensor de umidade percebe quando seu teor está abaixo do valor mínimo estabelecido e aciona o borrifo pelo tempo para restabelecer o teor de umidade. Desse sistema extraem-se os seguintes elementos:

- 1) um sistema de armazenamento de água para alimentar o borrifador;
- 2) um sensor que verifica a umidade do ar próximo às verduras;
- 3) um sensor que verifica se o borrifador está funcionando e controla seu tempo de acionamento; e
- 4) um sensor que mede a quantidade de água que existe no reservatório de água.

Esses elementos podem ser representados pelas seguintes variáveis:

X_1 (ÁGUA_OK); X_2 (UMIDIFICAR); X_3 (BORRIFO_ACIONADO); e X_4 (SENSOR_ÁGUA).

Imagine que o funcionamento desse sistema seja baseado em regras e, como exemplo de regra, suponha a seguinte: se existe água no reservatório e o sensor de umidade indica teor baixo, então o borrifo é acionado. Dessa maneira é perfeitamente possível inferir que, se X1 (ÁGUA_OK) e X2 (UMIDIFICAR) são verdadeiras, então X3 (BORRIFO_ACIONADO) também o será. Se X3 (BORRIFO_ACIONADO) é falsa, então X1 (ÁGUA_OK) é falsa ou X2 (UMIDIFICAR) é falsa. Se essa regra fosse escrita em lógica proposicional, resultaria em:

$$X1 (\text{ÁGUA_OK}) \wedge X2 (\text{UMIDIFICAR}) \Rightarrow X3 (\text{BORRIFO_ACIONADO})$$

O uso de regras limita a aplicação prática nos casos em que as variáveis sofrem de incertezas quanto ao valor falso ou verdadeiro (0 ou 1). Imagine que o sensor de umidade não seja confiável, o que significa que nem sempre a informação de que a umidade está baixa é correta. Portanto, existe uma chance de o sistema provocar o acionamento desnecessário do borrifo, pois o real valor da variável X2 é incerto.

Pesquisadores perceberam que sistemas especialistas baseados em regras não possuem bom desempenho para tratar sistemas sujeitos a incertezas. O funcionamento da Seleção Parametrizada é análogo a um sistema de regras: se as condições estão presentes, executa-se a ação correspondente; com uma pequena mudança nas condições, a ação não é executada.

O uso de probabilidade permite avaliar o sistema usando uma função de probabilidade conjunta. Essa função de probabilidade conjunta pode ser representada por uma tabela e é exemplificada a seguir, onde V significa verdadeiro e F, falso:

Tabela 1 - Exemplo de distribuição de Probabilidade Conjunta para as variáveis do Sistema de Umidificação.

A	U	B	S	Probabilidade Conjunta
V	V	V	V	0,6987
V	V	V	F	0,0307
V	V	F	V	0,0127
V	V	F	F	0,0068
...

Utilizando as variáveis A (ÁGUA_OK), U (UMIDIFICAR), S (SENSOR_ÁGUA) e B(BORRIFO_ACIONADO) do exemplo, o interesse primordial é estimar quais são as probabilidades para diversas situações envolvendo A, U, B e S. Essa tabela permite obter todos os valores de probabilidade marginal (probabilidade a uma variável), utilizá-los para avaliar as condições atuais do sistema e decidir que ação deve ser tomada para controlá-lo.

2.2 Inferência probabilística

Controlar um sistema significa obter valores de algumas variáveis, analisá-los e decidir que ação implementar. Esses valores podem ser representados por probabilidades de uma ou mais variáveis aleatórias, utilizando o prévio conhecimento do valor de outras variáveis. Por exemplo, se o borrifo não está funcionando, qual a probabilidade de que o sistema de água esteja funcionando e qual a decisão a ser tomada?

Esse cálculo pode ser denominado inferência probabilística, por analogia com os sistemas de inferência lógica. A inferência probabilística depende da definição de probabilidade condicional, que é a probabilidade de estimar a ocorrência de uma variável do sistema (X_i), conhecido o valor de outra variável (X_j), e que é representada pela seguinte expressão:

$$P(X_i | X_j) = \frac{P(X_i, X_j)}{P(X_j)} \quad (2.1)$$

Onde $P(X_i, X_j)$ é a probabilidade conjunta de X_i e X_j e $P(X_j)$ é a probabilidade de X_j .

Assim, a probabilidade de que o sistema de água esteja funcionando, dado que o borrifo não está funcionando, é fornecida por $P(A = V | B = F) = \frac{P(A = V, B = F)}{P(B = F)}$ ou, $P(A | \neg B) = \frac{P(A, \neg B)}{P(\neg B)}$ se utilizada a convenção de que variáveis sem símbolo significam assertivas verdadeiras (V) e, quando negadas (precedidas por \emptyset), significam assertivas falsas (F).

Considerando a definição de probabilidade condicional, é possível escrever:

$$P(X_i, X_j) = P(X_i | X_j) P(X_j) = P(X_j | X_i) P(X_i) \Rightarrow P(X_i | X_j) = \frac{P(X_j | X_i) P(X_i)}{P(X_j)} \quad (2.2)$$

Esta última equação foi formulada pelo Reverendo Thomas Bayes (PEARL, 1988; CRAMÉR, 1973) e recebeu o nome de Teorema de Bayes ou Regra de Bayes. Ela é fundamental para inversão probabilística e elemento-chave para responder a questões em sistemas probabilísticos.

Com base na expressão de probabilidade condicional, é possível obter uma expressão que lembra a regra da cadeia para funções (CRAMÉR, 1973, p. 40). Por meio dela, a probabilidade conjunta de $X_1, X_2, X_3, \dots, X_k$ é igual a probabilidade de ocorrência de X_1 vezes a probabilidade de ocorrência de X_2 , dado que X_1 ocorreu, vezes a probabilidade de ocorrência de X_3 dado que X_1 e X_2 ocorreram, até se chegar ao último dos fatores, que seria a probabilidade de ocorrência de X_k dado que $X_1, X_2, X_3, \dots, X_{k-1}$ ocorreram. Essa regra pode ser representada pela expressão a seguir:

$$P(X_1, X_2, X_3, \dots, X_{k-1}, X_k) = P(X_1) \cdot P(X_2|X_1) \cdot P(X_3|X_2, X_1) \dots P(X_k | X_{k-1}, \dots, X_3, X_2, X_1) \quad (2.3)$$

A forma geral para esse teorema é a seguinte:

$$P(X_1, X_2, X_3, \dots, X_k) = \prod_{i=1}^k P(X_i | X_{i-1}, \dots, X_2, X_1) \quad (2.4)$$

Aplicando essa fórmula ao exemplo do umidificador, a expressão seria: $P(S, B, U, A) = P(S) * P(B | S) * P(U | B, S) * P(A | U, B, S)$.

Para aplicar essa teoria ao problema, é preciso identificar as variáveis que representam a DI e como elas podem ajudar a identificar e estimar a probabilidade de ocorrência de alguma infração aduaneira.

A simplicidade até aqui apresentada pode conduzir à falsa idéia de que uma inferência probabilística não é complexa. Na verdade, ela se torna inadequada para um problema com n variáveis binárias e 2^n valores de probabilidade conjunta.

Entretanto, o autor e pesquisador Judea Pearl (1988) sustenta a hipótese de que a base de conhecimento sobre o domínio de um problema é construída de uma maneira que simplifica os cálculos de probabilidade condicional de certas variáveis, consideradas as evidências fornecidas. Essa simplificação é alcançada com o conceito de independência condicional, o qual é apresentado a seguir.

2.3 Independência condicional

Uma variável X possui a propriedade de ser condicionalmente independente de um conjunto de variáveis \mathcal{X}_i , dado um conjunto \mathcal{X}_j , se $P(X | \mathcal{X}_i, \mathcal{X}_j) = P(X | \mathcal{X}_j)$. A notação usada para representar esse fato é $I(X, \mathcal{X}_i | \mathcal{X}_j)$, o que significa que X é independente de \mathcal{X}_i se o valor de \mathcal{X}_j é conhecido.

O raciocínio básico sobre independência condicional é que, se o conhecimento sobre um conjunto de variáveis é dado, então é possível ignorar um outro conjunto de variáveis por ser a variável de interesse independente deste último.

No exemplo comentado anteriormente, se o sistema indica que X_1 (ÁGUA_OK) é verdadeiro, então a atenção fica voltada para saber se o borrifo está acionado, ou seja, X_3 (BORRIFO_ACIONADO) é verdadeiro. Logo, a informação sobre o sensor de água X_4 (SENSOR_ÁGUA) não tem a menor importância. Essa informação poderia ser descrita como $I(X_3, X_4 | X_1)$, o que quer dizer que dado X_1 , X_3 é independente de X_4 .

Se X_i é uma variável condicionalmente independente de outra variável X_j , dado um conjunto \mathcal{X} , então $P(X_i | X_j, \mathcal{X}) = P(X_i | \mathcal{X})$. Usando a definição de probabilidade condicional e que $I(X_i, X_j | \mathcal{X})$:

$$P(X_i, X_j | \div) = P(X_i | \div) P(X_j | \div)$$

Observe que a expressão $I(X_i, X_j | \mathcal{X})$ afirma que X_i é independente de X_j , dado \mathcal{X} , e que X_j também é independente de X_i , dado \mathcal{X} . Se \mathcal{X} é um conjunto vazio, então X_i e X_j são variáveis independentes.

As independências condicionais podem ser representadas por grafos acíclicos dirigidos (GAD), os quais são denominados Redes de Crença (*Belief Networks*) ou Redes Bayesianas (*Bayesian Networks*) e cuja teoria é apresentada a seguir.

2.4 Redes de Crença

Redes de Crença (PEARL, 1988) são um tópico de Inteligência Artificial na área de raciocínio sob incerteza e tiveram impulso na década de 80 e 90, quando vários trabalhos sobre o assunto foram publicados.

A Rede de Crença (exemplo na figura 1 a seguir) apresenta os nós como variáveis aleatórias proposicionais de interesse ao domínio do problema, e os arcos como relação de dependência causal ou funcional entre as variáveis. A quantificação dessa dependência é representada pelas probabilidades condicionais de cada nó da rede com relação àqueles que o provocam, denominados nós pais ou causais.

Uma Rede de Crença define que cada nó X_i no grafo é independente condicional de qualquer subconjunto de nós ascendentes do nó X_i se conhecidos os valores dos nós pais de X_i . Seja $A(X_i)$ o conjunto de nós ascendentes de X_i , isto é, o conjunto de todos os nós que estão acima de seus pais imediatos como causas diretas e indiretas destes, e seja $P(X_i)$ o conjunto de nós pais imediatos de X_i . Pela definição dada, pode-se dizer que ocorre $I(X_i, A(V_i) | P(V_i))$, ou seja X_i é independente do conjunto $A(X_i)$, dado $P(X_i)$, o que significa:

$$P(X_i | A(X_i), P(X_i)) = P(X_i | P(X_i)) \quad (2.5)$$

Se as variáveis $X_1, X_2, X_3, \dots, X_k$ são os nós de uma Rede de Crença e considerando a hipótese de independência condicional, é possível escrever a função de probabilidade conjunta para todos os nós na rede da seguinte maneira:

$$P(X_1, X_2, X_3, \dots, X_k) = \prod_{i=1}^k P(X_i | P(X_i)) \quad (2.6)$$

Essa expressão permite afirmar que a função de probabilidade conjunta de todas as variáveis pode ser representada pelo produto das probabilidades de cada variável condicionada por seus pais.

Exemplificando os conceitos expostos até aqui, é mostrada a seguir a Rede de Crença que modela o sistema automático de umidificação de verduras. Como já visto, as variáveis aleatórias proposicionais são A, B, S e U.

Raciocinando sobre o problema, se o sistema de fornecimento de água está OK (A) e o sensor de umidificação (U) informa que precisa umidificar, então o acionamento do borrifo pode ser efetuado, o que significa que B sofre influência de A e U. O sensor de água (S) sofre influência exclusivamente do sistema de fornecimento de água (A) e nenhuma do borrifo (B) ou do sensor de umidade (U). A Rede de Crença para esse problema pode ser vista na figura 1.

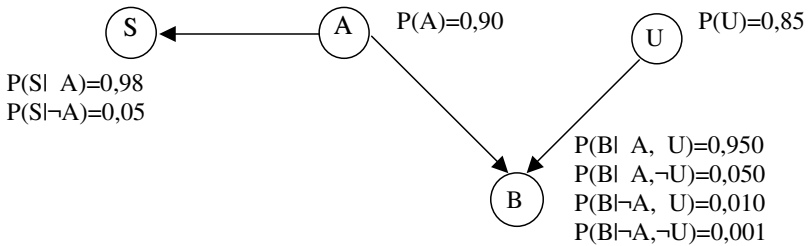


Figura 1 - Rede de Crença do Sistema Automático de Umidificação de Balcões Frigoríficos de Verduras.

Para o exemplo do sistema automático de umidificação, a função de probabilidade conjunta que representa essa Rede de Crença seria:

$$P(S,B,A,U) = P(S|B,A,U) P(B|A,U) P(A|U) P(U)$$

Pelo conceito de independência condicional, sabe-se que $P(S|B,A,U) = P(S|A)$ e $P(A|U) = P(A)$. Dessa forma, a expressão anterior pode ser reduzida para:

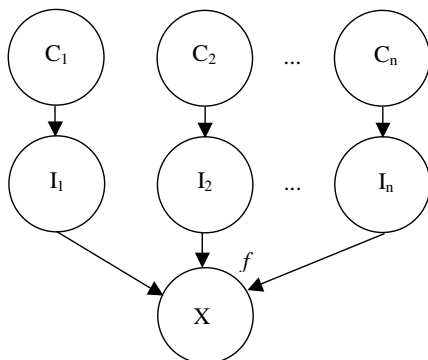
$$P(S,B,A,U) = P(S|A) P(B|A,U) P(A) P(U)$$

Essa mesma expressão pode ser obtida da Rede de Crença. De maneira simples, ela exhibe as independências condicionais e as relações de causa e efeito por meio de uma representação gráfica.

2.5 Independência causal

Durante a construção prática de Redes de Crença apareceram dificuldades que vários pesquisadores (PEARL, 1988; PRADHAN et al., 1994; HECKERMAN & BREESE, 1996) propuseram contornar com o uso de modelos baseados no conceito de independência causal.

Para compreender melhor esse conceito, imagine que a estrutura de uma dada Rede de Crença tenha uma variável X cuja ocorrência dependa de várias outras, ou seja, a produção do efeito X é proveniente de mais de uma causa. Entretanto, percebe-se que cada uma das causas exerce influência sobre a variável X sem concorrer com as outras. Um modelo para representar a independência causal é mostrado a seguir (HECKERMAN e BREESE, 1996; LUCAS, 2003):



Nesse modelo, cada uma das causas C_i influencia a variável X por intermédio das variáveis I_i e da função determinística f . A influência de cada causa C_i é independente das outras causas C_j para $j \neq i$, tendo a função f a tarefa de estabelecer o relacionamento entre as variáveis intermediárias e a variável comum X e, de forma indireta, levar a contribuição de cada causa C_i à variável X . *Figura 2 - Modelo geral de independência causal.*

O modelo convencional de Rede de Crença exige uma definição completa da tabela de probabilidades condicionais para o nó X . Se esta variável puder assumir m valores, e as n causas C_i puderem assumir t_i valores cada uma, então é preciso definir probabilidades; já o modelo com independência causal exigirá $(m-1) \prod_{i=1}^n t_i$ valores de probabilidade. Isso reduz a ordem exponencial do problema para polinomial em relação ao número de causas.

Noisy OR Gate (Porta OU com ruído – PEARL 1988; HECKERMAN e BREESE; 1996) foi o primeiro modelo a utilizar independência causal e será visto com mais detalhes a seguir.

2.6 *Noisy-OR Gate*

O modelo *Noisy-OR Gate* define a função f como a operação lógica OU e estabelece que todas as variáveis são binárias. A operação lógica $a \cup b$ (a ou b) é representada pela tabela verdade descrita a seguir:

Tabela 2 - Tabela verdade da operação $a \vee b$.

a	b	$a \vee b$
V	V	V
V	F	V
F	V	V
F	F	F

Suponha que a e b sejam causas de um efeito X , o que pode ser representado pela regra $a \vee b \rightarrow X$. Observando a tabela anterior, deduz-se que a única hipótese em que X não ocorre é quando ambas as causas não estão presentes. Isso significa que a probabilidade de X ocorrer é igual a 1 subtraído da probabilidade de nenhuma causa estar presente, o que pode ser representado pela seguinte equação:

$$P(X | I_1, I_2, \dots, I_n) = 1 - \prod_{i=1}^n P(\neg I_i | C_i) \quad (2.7)$$

Utilizando a representação gráfica para o modelo *Noisy-OR Gate* mostrada em (JENSEN, 1996), é possível simplificar a expressão anterior.

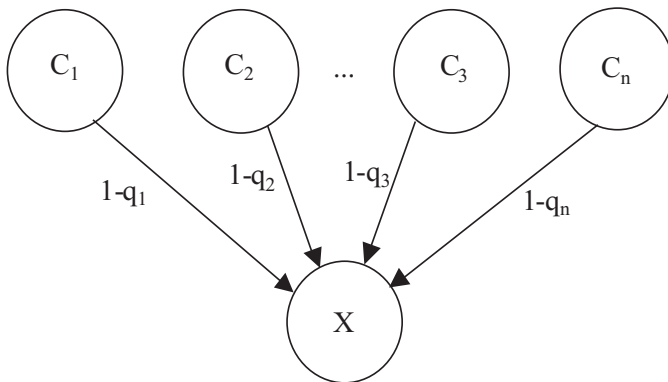


Figura 3 - Modelo *Noisy-OR Gate*.

Observe que o conjunto de variáveis $C_1, C_2, C_3, \dots, C_n$, são causas da variável binária X e que q_i representa a probabilidade de C_i não provocar o efeito X . Os valores de $P(X | C_1, C_2, C_3, \dots, C_n)$ e $P(\neg X | C_1, C_2, C_3, \dots, C_n)$ são dados por:.

$$P(X | C_1, C_2, C_3, \dots, C_n) = 1 - \prod_{i \in T} q_i \quad \text{e} \quad P(\neg X | C_1, C_2, C_3, \dots, C_n) = \prod_{i \in T} q_i \quad (2.8)$$

O conjunto T representa o conjunto de índices para os quais as variáveis causais foram evidenciadas, ou seja, é o conjunto em que $C_i = V$.

O uso desse modelo exige que dois requisitos sejam obedecidos: o primeiro é que, se todos os elementos C_i são falsos, ou seja, se nenhuma causa está presente, então $P(X)=0$, já que o conjunto C_i representa todas as causas que provocam X ; o segundo é que as influências devem atuar de forma independente, ou seja, se X pode ser causado por C_1 e C_2 , então o mecanismo de inibição de C_1 se manifesta independentemente do mecanismo de inibição de C_2 . Isso não quer dizer que C_1 é independente de C_2 , e sim que a influência de C_1 sobre X atua de forma independente da influência de C_2 sobre X .

Pode ser visto em Zhang (1995) que o uso de *Noisy-OR Gate* foi fundamental para a implantação prática da Rede CPSC. Algoritmos de inferências probabilísticas em Redes de Crenças usando *Noisy-OR* podem ser vistos nos trabalhos de Huang (1996) e Kearns (1998).

Como exemplo, observe a Rede de Crença, a seguir, com 5 variáveis aleatórias: V_1, V_2, V_3, V_4 e V_5 . As variáveis V_1 e V_2 são causas de V_3 e V_4 e estas, por sua vez, são causas de V_5 . As variáveis V_3, V_4 e V_5 são modeladas como *Noisy-OR*, o que significa que elas são influenciadas de forma independente por seus pais. Portanto, a ocorrência de V_3 e de V_4 depende de V_1 ou de V_2 , e a de V_5 depende de V_3 ou de V_4 .

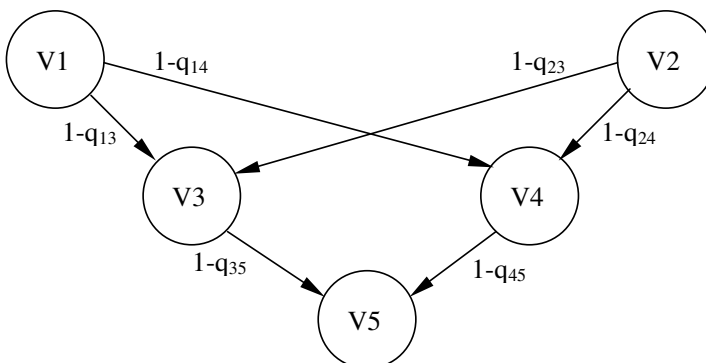


Figura 4 - Rede de Crença usando o modelo *Noisy-OR Gate* para os nós V_3, V_4 e V_5 .

Os valores q_{ij} representam a probabilidade de que, mesmo que a causa C_i esteja presente, o efeito X não é observado; portanto, a probabilidade $(1-q_{ij})$ representa a chance de ocorrer o efeito X . As tabelas de probabilidade condicional para as variáveis $V3$, $V4$ e $V5$ são construídas utilizando as equações vistas em 2.8.

A Tabela 3 exibe as probabilidades condicionais para essa Rede de Crença.

Tabela 3 - Tabela de probabilidades condicionais ($V3|V1,V2$), ($V4|V1,V2$) e ($V5|V3,V4$).

		V1	V		F	
		V2	V	F	V	F
V3	V		$1-q_{13} \cdot q_{23}$	$1-q_{13}$	$1-q_{23}$	0
	F		$q_{13} \cdot q_{23}$	q_{13}	q_{23}	1
V4	V		$1-q_{14} \cdot q_{24}$	$1-q_{14}$	$1-q_{24}$	0
	F		$q_{14} \cdot q_{24}$	q_{14}	q_{24}	1
		V3	V		F	
		V4	V	F	V	F
V5	V		$1-q_{35} \cdot q_{45}$	$1-q_{35}$	$1-q_{45}$	0
	F		$q_{35} \cdot q_{45}$	q_{35}	q_{45}	1

Pelo modelo tradicional é necessário obter os valores combinados das causas $V1$ e $V2$ influenciando as variáveis $V3$ e $V4$, assim como dessas duas variáveis combinadas para influenciar a variável $V5$. No modelo *Noisy-OR* é necessário apenas determinar a influência de cada causa separadamente para provocar o efeito, ou seja, qual a probabilidade de somente a variável C_i provocar o efeito X .

Suponha que se deseja calcular o valor de $P(V5|V1,V2)$, ou seja, a probabilidade de ocorrência de $V5$ considerando as variáveis $V1$ e $V2$ evidenciadas, o que pode ser feito indiretamente com o cálculo de $P(\neg V5|V1,V2)$. Incluindo os nós pais de $V5$ e aplicando uma variação da regra da cadeia (2.4), obtêm-se a seguinte expressão:

$$P(\neg V5 | V1, V2) = \sum_{V3,V4} P(\neg V5, V3, V4 | V1, V2) = \sum_{V3,V4} P(\neg V5 | V3, V4, V1, V2) P(V3, V4 | V1, V2)$$

Como $I(V5, [V1, V2] | [V3, V4])$ e $I(V3, V4 | [V1, V2])$, reduz-se a expressão para:

$$P(\neg V5 | V1, V2) = \sum_{V3,V4} P(\neg V5 | V3, V4) P(V3 | V1, V2) P(V4 | V1, V2)$$

Substituindo os valores de (V3,V4) (V3;¬V4) (¬V3,V4) e (¬V3,¬V4), obtém-se:

$$\begin{aligned}
 P(\neg V5 | V1, V2) = & P(\neg V5 | V3, V4) P(V3 | V1, V2) P(V4 | V1, V2) + \\
 & P(\neg V5 | V3, \neg V4) P(V3 | V1, V2) P(\neg V4 | V1, V2) + \\
 & P(\neg V5 | \neg V3, V4) P(\neg V3 | V1, V2) P(V4 | V1, V2) + \\
 & P(\neg V5 | \neg V3, \neg V4) P(\neg V3 | V1, V2) P(\neg V4 | V1, V2)
 \end{aligned}$$

Utilizando os valores q_{13} , q_{14} , q_{23} , q_{24} , q_{35} e q_{45} , a expressão se modifica para:

$$\begin{aligned}
 P(\neg V5 | V1, V2) = & q_{35} \cdot q_{45} (1 - q_{13} \cdot q_{23}) (1 - q_{14} \cdot q_{24}) + \\
 & q_{35} (1 - q_{13} \cdot q_{23}) (q_{14} \cdot q_{24}) + \\
 & q_{45} (q_{13} \cdot q_{23}) (1 - q_{14} \cdot q_{24}) + \\
 & 1 (q_{13} \cdot q_{23}) (q_{14} \cdot q_{24})
 \end{aligned}$$

Supondo o caso em que não exista mecanismo de inibição entre [V3, V4] e V5, portanto $q_{35}=0$ e $q_{45}=0$, a expressão final ficaria assim:

$$P(\neg V5 | V1, V2) = q_{13} \cdot q_{23} \cdot q_{14} \cdot q_{24} \quad \text{e} \quad P(V5 | V1, V2) = 1 - q_{13} \cdot q_{23} \cdot q_{14} \cdot q_{24} \quad \text{ou}$$

$$P(\neg V5 | V1, V2) = \prod_{i=1,2; j=3,4} q_{ij} \quad \text{e} \quad P(V5 | V1, V2) = 1 - \prod_{i=1,2; j=3,4} q_{ij} \quad (2.9)$$

A forma reduzida da expressão anterior será usada no capítulo seguinte para a modelagem de uma Rede de Crença representativa do problema.

Será visto como essa teoria pode ser aplicada ao problema de estimar a probabilidade de ocorrência de uma infração aduaneira em uma DI. O modelo emprega o conceito de independência causal baseado em *Noisy-OR Gate*.

CAPÍTULO 3

MODELAGEM

Este capítulo descreve a construção do modelo para estimar a probabilidade de ocorrência de infração aduaneira em um despacho de importação, baseado na teoria de Redes de Crença. Para que isso seja possível, é preciso que certos passos sejam seguidos (RUSSEL, 1995), os quais são enumerados abaixo:

- i) delimitar as questões de interesse que a rede deve ser capaz de responder por intermédio de uma estimativa probabilística de ocorrência;
- ii) identificar as variáveis que podem influenciar ou são influenciadas pelas questões de interesse definidas anteriormente, bem como o respectivo domínio;
- iii) estabelecer as relações de causa e efeito entre as variáveis que representam as questões de interesse e as demais variáveis;
- iv) determinar os valores de probabilidade condicional de cada variável e seus respectivos pais, bem como os valores de probabilidade *a priori* para as demais;
- v) decidir qual a estratégia a ser adotada para o cálculo de inferência probabilística. O algoritmo de cálculo dependerá da topologia de rede obtida.

No primeiro passo, decidiu-se que a questão primordial seria determinar a existência, ou não, de alguma infração aduaneira em uma DI.

Para a consecução do segundo passo, foram analisadas as principais variáveis de uma DI e sua provável relação com a ocorrência de uma infração aduaneira.

No terceiro passo foram determinados os principais efeitos e causas para ocorrência das infrações aduaneiras. O grafo resultante exhibe o relacionamento entre as variáveis que caracterizam uma DI e as infrações aduaneiras. A subjetividade na escolha das variáveis caracterizadoras da DI e na definição do relacionamento destas com as infrações aduaneiras em tempo de despacho tem peso significativo no modelo e, como não poderia deixar de sê-lo, resultou da experiência em

fiscalização aduaneira.

Para o quarto passo, decidiu-se que a abordagem para o levantamento das probabilidades condicionais seria freqüentativa, ou seja, determinam-se quantos casos tiveram infração e divide-se esse valor pelo número total de casos fiscalizados.

Em seguida é especificado o algoritmo a ser usado para o cálculo da probabilidade geral de ocorrência de alguma infração aduaneira para uma DI ainda não fiscalizada. O algoritmo obtido é simples e decorreu da simplificação adotada para redução da complexidade computacional.

Deve ficar claro que somente o valor da probabilidade de ocorrência para alguma infração não determinará a seleção desta para conferência aduaneira. Será mostrado que é preciso decidir, com base em um critério, quais DIs serão fiscalizadas e quais DIs serão liberadas. O critério escolhido possui aplicação objetiva e fácil cálculo computacional.

3.1 Variáveis causais

Decidido que o interesse do sistema seria identificar as probabilidades de ocorrência de alguma infração aduaneira, o segundo passo significa descobrir que informações prestadas na DI ajudam a caracterizar as infrações aduaneiras. Dessa análise foram estipuladas as seguintes variáveis causais:

- CNPJ - É o número identificador no Cadastro Nacional da Pessoa Jurídica composto por 14 dígitos (formato 99.999.999/0001-99). Esse trabalho focou a pessoa jurídica; entretanto, pode ser utilizado para a pessoa física.
- URF_D - Código numérico de 7 dígitos que identifica a Unidade da Receita Federal onde ocorre o Despacho Aduaneiro de Importação.
- URF_E - Código numérico de 7 dígitos que identifica a Unidade da Receita Federal onde ocorreu a entrada da mercadoria no território aduaneiro.
- RT - Código numérico de uma posição que estabelece o regime de tributação a que está sujeita a mercadoria declarada. O regime de tributação pode ser recolhimento

integral, suspensão, isenção, redução, imunidade e não-incidência.

- NCM - Código numérico de oito posições de acordo com a Nomenclatura Comum do Mercosul, obedecendo às regras gerais de interpretação do Sistema Harmonizado de Classificação de Mercadorias.
- PO - Código numérico de três posições que identifica o País de Origem. Considera-se origem da mercadoria o lugar onde foi produzida ou recebeu a maior parcela de transformação, beneficiamento ou montagem.
- PA - Código numérico de três posições que identifica o País de Aquisição. Representa o lugar onde o exportador possui domicílio.
- PP - Código numérico de três posições que identifica o País de Procedência. O conceito de procedência é o local de onde a mercadoria foi embarcada com destino ao território brasileiro.

Portanto, o sistema fica reduzido à observação de oito variáveis, com a ressalva de que as variáveis CNPJ, URF_D , URF_E e PP são dados gerais da Declaração de Importação e as outras variáveis, NCM, RT, PO e PA, são dados específicos, fazendo parte de cada adição da Declaração de Importação.

Embora o sistema registre a existência de aproximadamente dois terços das declarações de importação com apenas uma adição, isto é, apenas um tipo de mercadoria, é preciso levar em consideração os casos em que o número de adições é maior, a fim de se estabelecer um critério de comparação razoável e computável. A Rede de Crença será executada para cada adição e, obtidos os diversos resultados individuais, calcular-se-á o valor de probabilidade geral de ocorrência de alguma infração aduaneira.

3.2 Grafo da Rede de Crença

O terceiro passo para a modelagem tem como objetivo definir as relações entre as variáveis para a construção de um grafo acíclico

dirigido. A variável de interesse é representada por IA (Infração Aduaneira), e o cálculo do valor de sua probabilidade de ocorrência é feito observando os valores das variáveis causais definidas anteriormente.

Um desafio é obter a Tabela de Probabilidades Condicionais para a variável IA, considerando todas as outras 8 (oito) variáveis ($P(IA|CNPJ, URF_D, URF_E, RT, NCM, PO, PA, PP)$), o que exigiria um espaço de armazenamento muito alto, em torno de $2,33 \times 10^{20}$, levando em conta que há ± 20.000 CNPJs, ± 10.000 NCMs, 245 POs, 245 PAs, 245 PPs, 6 RTs, 100 URF_D s e 100 URF_E s.

Diante dessa situação, parece razoável lançar mão de modelos alternativos de Redes de Crença baseados no conceito de Independência Causal para reduzir a complexidade envolvida.

Analisando as variáveis causais, percebeu-se que era possível utilizar o conceito de independência causal com o uso do modelo *Noisy-OR Gate*. O uso desse modelo é recomendado por PEARL, (1988) para os casos em que a especificação da Rede de Crença não pode ser completa porque interações detalhadas entre as causas não estão disponíveis; para situações em que as relações são muito complexas; ou ainda quando a quantidade de valores de probabilidades a serem definidas é muito grande e o número de observações é pequeno.

Sabendo que as variáveis causais são: NCM, CNPJ, PO, PA, PP, URF_D , URF_E e RT, é possível estimar os valores de $P(IA|NCM)$, $P(IA|CNPJ)$, $P(IA|PO)$, $P(IA|PA)$, $P(IA|PP)$, $P(IA|URF_D)$, $P(IA|URF_E)$ e $P(IA|RT)$ e usá-los para calcular o valor de $P(IA|NCM, CNPJ, PO, PA, PP, URF_D, URF_E, RT)$.

A conversão do modelo de Rede de Crença original para o modelo canônico usando *Noisy-OR Gate* permitiu substituir variáveis causais com domínios representados por múltiplos valores discretos, por exemplo, a variável PO tem 245 valores possíveis, por variáveis causais binárias. Isso significa que o sistema vai avaliar a influência da NCM para provocar, ou não, a infração IA. Isso é diferente da interpretação tradicional que avaliaria a influência de todas as NCMs na infração IA.

Esse artifício é possível porque o foco não é obter a probabilidade geral de uma DI conter alguma infração aduaneira perante todos os casos possíveis; o que se deseja é prever a probabilidade confrontando casos idênticos ou semelhantes já fiscalizados, o que significa estimar

a probabilidade perante um universo bem menor de casos.

Assim, se todas as variáveis causais estão sempre presentes, pode-se avaliar a influência de cada uma delas na ocorrência do efeito com apenas dois valores, um que define a contribuição para a ocorrência do efeito e outro para a sua inibição. O sistema passa a se concentrar na estimativa da probabilidade de ocorrência da infração aduaneira apenas para o valor declarado da variável causal.

Definir q_i como o valor de probabilidade para não-ocorrência de IA quando a variável causal C_i está presente permite calcular a probabilidade de IA com a expressão $P(IA|NCM, CNPJ, PO, PA, PP, URF_D, URF_E, RT) = 1 - (q_1 \cdot q_2 \cdot q_3 \cdot q_4 \cdot q_5 \cdot q_6 \cdot q_7 \cdot q_8)$.

Cada valor q_i , o qual chamaremos de fator de inibição, é obtido subtraindo de 1 o valor resultante da divisão do número de despachos conferidos que apresentaram alguma infração aduaneira e que relatam a presença da respectiva variável causal pelo número total de despachos conferidos com essa mesma variável causal. Por exemplo, o valor que representa a probabilidade da ocorrência de IA, quando presente uma dada NCM, é obtido dividindo o número de despachos conferidos, contendo mercadorias classificadas nessa NCM e nos quais ficou constatada a ocorrência de alguma infração aduaneira, pelo número total de despachos aduaneiros conferidos contendo mercadorias classificadas nessa NCM.

A tabela de probabilidades condicionais exigiria apenas $2^8 = 256$ valores, o que é bem menor que a estimativa de $2,33 \times 10^{20}$ citada anteriormente. Assim, o cálculo de $P(IA|NCM, CNPJ, PO, PA, PP, URF_D, URF_E, RT)$ exigirá tempo e armazenamento menores.

Mesmo que existam casos que não permitam estimar o valor de $P(IA|C_i)$ porque, por exemplo, é a primeira vez que uma DI com determinado valor para C_i é fiscalizada, ainda é possível estimar os valores das outras variáveis causais. Contornar a falta desse valor pode ser feito atribuindo a q_i o valor 1, o que significa desconsiderar a influência da variável C_i na ocorrência de IA. Sob este ponto de vista, o modelo adotado mostrou a vantagem de permitir comparar situações semelhantes e determinar a probabilidade de o despacho aduaneiro conter uma infração aduaneira, mesmo que seja a primeira importação de uma pessoa jurídica, por exemplo.

De forma subjetiva, foi considerada a possibilidade de existir

um relacionamento entre as variáveis PP, PO e PA que provocaria o efeito de forma combinada, porém, independente das demais. Sem abandonar o conceito de independência causal, foi possível substituí-las por uma única variável denominada PP_PO_PA, a qual suporta melhor a hipótese de existência da infração e é independente causal das outras. Esse raciocínio aplica-se de forma análoga a outros dois conjuntos de variáveis: URF_D com URF_E ; e NCM com RT.

Portanto, o modelo final mostrará algumas variáveis causais aglutinadas, o que diminuiu a quantidade de variáveis causais para cada tipo de infração aduaneira.

Para que fosse possível discernir as diversas infrações aduaneiras, identificadas nos artigos do Decreto n.º 4.543/2002, foi confeccionada a Tabela 4. Essa tabela relaciona 18 tipos de infração aduaneira passíveis de ocorrerem em um despacho aduaneiro de importação, cada uma representada por uma variável de interesse IA_i . Reunindo estas com as variáveis causais, foi possível projetar um modelo de Rede de Crença, o qual é apresentado na figura 5.

Tabela 4 - Relação de Infrações Aduaneiras e respectivas penalidades.

Tipo	Infração	Penalidade
1	Falsificação ou adulteração de documentos	Perdimento da Mercadoria, ou seja, valor igual ao valor aduaneiro da mercadoria.
2	Falsificação ou adulteração de mercadorias	Perdimento da Mercadoria, ou seja, valor igual ao valor aduaneiro da mercadoria.
3	Falsa declaração de conteúdo	Perdimento da Mercadoria, ou seja, valor igual ao valor aduaneiro da mercadoria.
4	Mercadoria atentatória à moral, aos bons costumes, à saúde ou à ordem pública	Perdimento da Mercadoria, ou seja, valor igual ao valor aduaneiro da mercadoria.
5	Importação sem licença de importação com emissão vedada ou suspensa	Perdimento da Mercadoria, ou seja, valor igual ao valor aduaneiro da mercadoria.
6	Ocultação do real importador mediante fraude ou simulação	Perdimento da Mercadoria, ou seja, valor igual ao valor aduaneiro da mercadoria.
7	Extravio de mercadoria	Multa igual a 50% do II sem considerar isenção ou redução.
8	Inexistência da fatura comercial ou apresentação fora do prazo fixado	Multa igual a 10% do II sem considerar isenção ou redução.
9	Apresentação da fatura comercial sem o visto consular, quando exigido	Multa igual a 10% do II sem considerar isenção ou redução.
10	Diferença entre o preço declarado e o preço praticado ou arbitrado	Multa igual à diferença de preço.
11	Importação de mercadoria sem licença de importação de emissão permitida	Multa igual a 30% do valor aduaneiro. Valor mínimo de R\$ 137,60.
12	Embarque da mercadoria antes da emissão da licença de importação	Multa igual a 30% do valor aduaneiro. Valor mínimo de R\$ 137,60 e máximo de R\$ 1.376,00.
13	Embarque depois de vencido o prazo da licença de importação de mais de vinte até quarenta dias	Multa igual a 20% do valor aduaneiro. Valor mínimo de R\$ 137,60 e máximo de R\$ 1.376,00.
14	Embarque depois de vencido o prazo da licença de importação em até vinte dias	Multa igual a 10% do valor aduaneiro. Valor mínimo de R\$ 137,60 e máximo de R\$ 1.376,00.
15	Classificada incorretamente na Nomenclatura Comum do Mercosul	Multa igual a 1% do valor aduaneiro. Valor mínimo de R\$ 500,00.
16	Quantificada incorretamente na unidade de medida estatística	Multa igual a 1% do valor aduaneiro. Valor mínimo de R\$ 500,00.
17	Falta de pagamento, falta de declaração ou declaração inexata	Multa igual a 75% do tributo ou da diferença apurada, agravada para 112,5% se não for atendida intimação para prestar esclarecimentos.
18	Falta de pagamento, falta de declaração e declaração inexata quando comprovada ocorrência de fraude	Multa igual a 150% do tributo ou da diferença apurada, agravada para 225% se não for atendida intimação para prestar esclarecimentos.

O sufixo D nas variáveis causais, exceto para a variável URF_D , significa declarado e representa a informação prestada pelo importador na DI. O sufixo R significa real e foi apostado em variáveis diagnósticas

CNPJ, NCM e PO para indicar os possíveis valores reais dessas variáveis. Essas variáveis diagnósticas NCM_R , $CNPJ_R$ e PO_R serão usadas em trabalhos futuros.

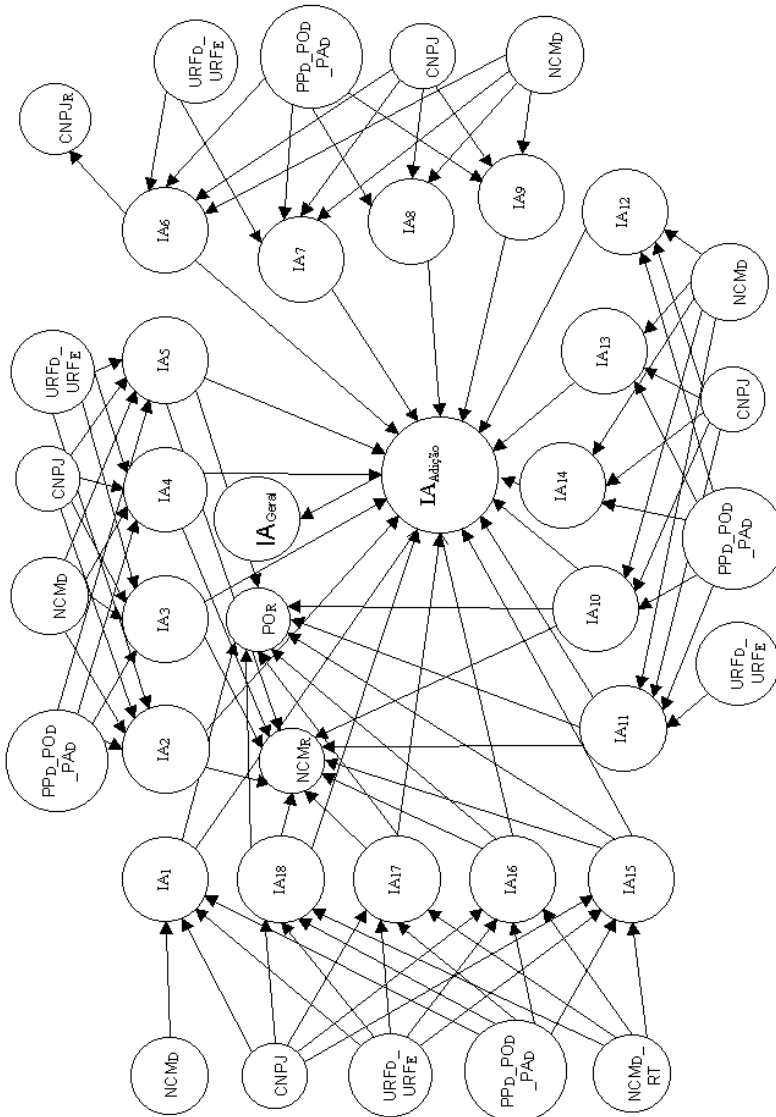


Figura 5 - Modelo de Rede de Crença.

A figura 5 apresenta as variáveis causais repetidas em cada um dos seus quatro cantos com o objetivo de diminuir o número de arcos cruzando o grafo. Esse grafo nasceu da experiência subjetiva com conferência aduaneira e da análise estatística dos dados coletados; por isso, está sujeito a críticas e revisões.

O modelo proposto permitirá obter o valor de probabilidade de ocorrência de infração e decidir qual o canal de conferência a ser aplicado. Valores de probabilidade próximos de 1 (um) selecionarão para o Canal Vermelho, valores pouco menores selecionarão para o Canal Amarelo e, abaixo desses, selecionarão para o Canal Verde.

Por sua natureza específica, o sistema não irá selecionar para o Canal Cinza, ficando a cargo do sistema atual de Seleção Parametrizada fazê-lo.

3.3 Probabilidades condicionais

O quarto passo da modelagem versa sobre o estabelecimento das tabelas de probabilidades condicionais de cada variável e seus respectivos pais. Isso pode ser feito de duas maneiras: a primeira seria pela atribuição subjetiva, ou seja, uma pessoa com base em sua experiência pessoal ditaria as probabilidades de ocorrência de cada causa; a segunda seria pela atribuição de frequência relativa, ou seja, calcular com base no histórico de infrações as diversas probabilidades condicionais. A opção recaiu sobre a segunda maneira, já que a primeira necessitaria de uma atribuição de valores muito grande.

Será visto no capítulo 4 que foram construídos programas para vasculhar a base de dados coletados e determinar os valores de $P(IA_i|NCM_D)$, $P(IA_i|CNPJ_D)$, $P(IA_i|PP_{D-}PO_{D-}PA_D)$, $P(IA_i|URF_{D-}URF_E)$ e $P(IA_i|NCM_{D-RT})$, onde IA_i representa cada um dos 18 tipos de infração aduaneira. No banco de dados foram armazenados os valores q_i que são iguais a $1-P(IA_i|C_i)$, onde C_i são as variáveis causais NCM_D , $CNPJ_D$, $PP_{D-}PO_{D-}PA_D$, $URF_{D-}URF_E$ e NCM_{D-RT} .

Nesse modelo, os valores de probabilidade *a priori* não são necessários porque todas as variáveis causais estão sempre presentes, e sua falta significa atribuir o valor unitário ao fator inibitório q_i . É importante ressaltar que o valor de $q_i=1$ não significa que a variável não influencia a causa, e sim que a probabilidade não pode ser determinada.

3.4 Algoritmo de inferência

No quinto passo da modelagem, é preciso examinar o algoritmo de inferência probabilística. Em função das condições do modelo será possível utilizar a expressão obtida em 2.9. Embora o caso mais comum seja de DIs com apenas uma adição, é preciso considerar que existem DIs com mais de uma adição que precisam ser avaliadas. O modelo, portanto, vai gerar, para cada adição de uma DI, 18 valores de probabilidade de ocorrência de infrações aduaneiras.

A expressão em 2.9 permite efetuar o produto desses 18 valores para obter probabilidade de ocorrência para a adição. Isso é possível porque o conceito de independência causal também é utilizado aqui, já que as infrações aduaneiras individuais influenciam a infração aduaneira na adição de forma independente e é pressuposto que, se não existe qualquer tipo de infração, então não existe infração para adição.

Já para o valor geral de ocorrência de alguma infração aduaneira para a DI, foi decidido que será o maior valor de probabilidade de ocorrência de infração aduaneira entre todas as suas adições. Dessa maneira, considera-se que a probabilidade de ocorrência de alguma infração aduaneira para uma DI não deve ser inferior à maior probabilidade entre todas as adições.

Ponderando o que foi discutido, a proposta para calcular a probabilidade de ocorrência e de não-ocorrência geral de alguma infração aduaneira de uma DI pode ser representada pelas seguintes expressões:

$$P(\text{IA}_{\text{GDI}} \mid \text{IA}_{\text{Ad1}}, \text{IA}_{\text{Ad2}}, \dots, \text{IA}_{\text{Adm}}) = 1 - \min \left(\prod_{j=1}^m \prod_{i=1}^{18} q_{ij} \right)$$

$$P(\neg \text{IA}_{\text{GDI}} \mid \text{IA}_{\text{Ad1}}, \text{IA}_{\text{Ad2}}, \dots, \text{IA}_{\text{Adm}}) = \min \left(\prod_{j=1}^m \prod_{i=1}^{18} q_{ij} \right)$$

Onde

IA_{GDI} representa a hipótese de ocorrência de alguma infração aduaneira para a DI e m representa a quantidade de adições;

IA_{Adj} representa o evento ocorrência de alguma infração aduaneira para a adição j ;

q_{ij} corresponde à probabilidade de ativação do mecanismo de inibição da infração aduaneira do tipo i na adição j .

3.5 Ordenação e decisão

Concluídos os cinco passos para a modelagem da Rede de Crença, incluindo a fórmula de cálculo da probabilidade geral de ocorrência de alguma infração aduaneira de uma DI, ainda resta um problema que não é de fácil solução: definir o critério de ordenação das DIs e decidir sobre a atribuição do canal de conferência. Até aqui o processo tinha observado os dados das DIs, usado esses dados para extrair informações sobre infrações e gerado o conhecimento probabilístico. Entretanto, falta usar esse conhecimento para decidir que DIs devem ser escolhidas. Ordenar e escolher as DIs pelo maior valor de probabilidade de ocorrência de uma infração não é uma boa solução porque desconsidera o potencial de lançamento de crédito tributário.

Procurando melhorar a eficiência do procedimento fiscal, uma estimativa de lançamento será feita com o objetivo de determinar as DIs com maior chance de recuperação de crédito fiscal. Essa estimativa será baseada na soma dos valores das multas e da diferença de impostos que seriam cobrados do importador caso as hipóteses de ocorrência de infrações viessem a se tornar realidade. Portanto, uma declaração contendo mercadorias de valores elevados e sujeita a um lançamento de crédito tributário de valor pífio não será escolhida se outra DI, mesmo contendo mercadorias de valores mais baixos, possuir um potencial de lançamento muito maior, presumindo que os valores de probabilidade de ocorrência de infração aduaneira estejam próximos.

A escala de probabilidade de 0 a 1 (de 0% a 100%) foi dividida em intervalos regulares de 20%; portanto, foram criadas 5 faixas. Num primeiro momento, alocam-se as DIs nesses intervalos de acordo com o valor de probabilidade geral calculado. Em seguida, dentro de cada intervalo, utilizam-se as estimativas de lançamento para ordená-las. O resultado final é uma lista de DIs ordenadas primeiro pelos intervalos e dentro deles pelas estimativas de lançamento ponderadas pelos valores de probabilidade, ou seja, multiplicadas por estes. Foi assumido o risco

de que discrepâncias possam ser introduzidas exatamente no corte de um intervalo para outro.

Para permitir que a Administração Aduaneira pudesse priorizar alguns tipos de infrações aduaneiras para a seleção de DIs sem modificar o conceito probabilístico da Rede de Crença, é criado um fator de agravo. Esse fator é uma atribuição subjetiva e o que ele altera é a estimativa ponderada de lançamento, o que acaba refletindo na ordem atribuída a cada DI dentro do seu respectivo intervalo. Dessa maneira, para que a infração aduaneira do tipo 1 influenciasse com peso maior a escolha de uma DI, a respectiva estimativa de lançamento poderia ser aumentada pelo sistema com um fator de agravo de, por exemplo, 100, enquanto outras poderiam ter fator de agravo menor que 10. Ressalte-se que a influência do fator de agravo entre as infrações reside no valor relativo entre eles, e não na imputação de valores altos ou baixos.

No capítulo seguinte, será mostrado como o modelo foi implementado e simulado. Serão mostradas as estruturas de dados e os bancos de dados construídos e as finalidades dos programas desenvolvidos.

CAPÍTULO 4

PROPOSTA DE IMPLEMENTAÇÃO

Neste capítulo é descrita de forma sucinta a implementação do modelo abordado no capítulo anterior. Os dados das DIs e os resultados da execução da Rede de Crença são armazenados em Banco de Dados Relacionais. Deve-se frisar que, embora seja importante em alguns aspectos, a tecnologia adotada não merece consideração, e sim a aplicação prática dos conceitos até aqui discutidos.

O primeiro passo foi replicar a estrutura de dados idêntica à existente no Siscomex para avaliar qualitativamente a dificuldade de implementação. Em seguida foram realizados estudos para decidir como armazenar as informações estáticas da Rede de Crença (tabelas de probabilidade condicional) e as informações dinâmicas (probabilidades calculadas e estimativas de lançamento).

Serão comentados os programas necessários para registrar

as infrações ocorridas, calcular as probabilidades, estimar os lançamentos e ordenar e efetuar a escolha das DIs.

4.1 Construção dos bancos de dados

O primeiro banco de dados construído replica a estrutura de dados de uma DI, e seu preenchimento foi feito com DIs registradas no período de janeiro de 1997 a setembro de 2002, escolhidas aleatoriamente do Siscomex. Concentrando o trabalho nas DIs de interesse relevante, foram extraídas 117.097 DIs despachadas e desembaraçadas para consumo e depois mais 20.734 DIs nas mesmas condições. O primeiro conjunto foi utilizado para efetuar o levantamento das probabilidades condicionais e o segundo, para validar o modelo. A relação das variáveis causais e os correspondentes campos e tabelas da DI são apresentados a seguir:

Tabela 5 - Relacionamento entre as variáveis causais e as tabelas e campos da estrutura de dados da DI.

Variável Causal	Tabela do Banco	Campo da Tabela
CNPJ	DECLARACAO_IMPORTACAO	NR_IMPORTADOR
URF _D	DECLARACAO_IMPORTACAO	CD_URF_DESPACHO
URF _E	DECLARACAO_IMPORTACAO	CD_URF_ENTR_CARGA
PP	DECLARACAO_IMPORTACAO	CD_PAIS_PROC_CARGA
PO	OP_ADICAO_DE_IMPORTACAO	CD_PAIS_ORIG_MERC
PA	OP_ADICAO_DE_IMPORTACAO	CD_PAIS_AQUIS_MERC
NCM	OP_ADICAO_DE_IMPORTACAO	CD_MERCADORIA_NCM
RT	OP_ADICAO_DE_IMPORTACAO	CD_REGIME_TRIBUTAR

Verificou-se que era necessário armazenar as informações referentes às fiscalizações efetuadas e as ocorrências confirmadas de cada um dos tipos de infrações aduaneiras. Para isso foi construído um banco de dados chamado INFRACAO, cuja estrutura pode ser vista na figura 6. A primeira tabela criada foi DI_VERIFICADAS com as informações necessárias para registrar os dados das DIs fiscalizadas, como, por exemplo, seu número, suas adições e os valores das variáveis causais, bem como dados adicionais com o objetivo de usá-los no futuro. Os seus campos são discriminados na tabela apresentada a seguir:

Tabela 6 - Estrutura da Tabela DI_VERIFICADAS.

Nome do Campo	Tabela de Origem	Nome do Campo	Tabela de Origem
NR_DECLARACAO_IMP	OP_ADICAO_DE_IMPORTACAO	NM_URF_ENTR_CARGA	DECLARACAO_IMPORTACAO
NR_SEQ_RETIFICACAO	OP_ADICAO_DE_IMPORTACAO	CD_PAIS_PROC_CARGA	DECLARACAO_IMPORTACAO
NR_OP_IMP	OP_ADICAO_DE_IMPORTACAO	CD_PAIS_ORIG_MERC	OP_ADICAO_DE_IMPORTACAO
NR_SEQ_DETALHE	OP_ADICAO_DETALHE_MERCADORIA	CD_PAIS_AQUIS_MERC	OP_ADICAO_DE_IMPORTACAO
NR_IMPORTADOR	DECLARACAO_IMPORTACAO	CD_MERCADORIA_NCM	OP_ADICAO_DE_IMPORTACAO
CD_URF_DESPACHO	DECLARACAO_IMPORTACAO	NM_MERCADORIA_NCM	OP_ADICAO_DE_IMPORTACAO
NM_URF_DESPACHO	DECLARACAO_IMPORTACAO	IN_MATERIAL_USADO	OP_ADICAO_DE_IMPORTACAO
CD_VIA_TRANSP_CARG	DECLARACAO_IMPORTACAO	CD_REGIME_TRIBUTAR	OP_ADICAO_DE_IMPORTACAO
CD_URF_ENTR_CARGA	DECLARACAO_IMPORTACAO		

A segunda tabela criada foi INFRACAO_OCORRIDA, a qual relaciona todas as infrações ocorridas para cada uma das adições. A estrutura dessa tabela é mostrada a seguir:

Tabela 7 - Estrutura da Tabela INFRACAO_OCORRIDA.

Nome do Campo	Tabela de Origem
NR_DECLARACAO_IMP	OP_ADICAO_DE_IMPORTACAO
NR_SEQ_RETIFICACAO	OP_ADICAO_DE_IMPORTACAO
NR_OP_IMP	OP_ADICAO_DE_IMPORTACAO
NR_SEQ_DETALHE	OP_ADICAO_DETALHE_MERCADORIA
CD_TIPO_INFRACAO	TIPOS_INFRACOES
NR_IMPORTADOR_REAL	DECLARACAO_IMPORTACAO
CD_PAIS_ORIG_MERC_DECL	OP_ADICAO_DE_IMPORTACAO
CD_PAIS_ORIG_MERC_REAL	Nenhuma
CD_MERCADORIA_NCM_DECL	OP_ADICAO_DE_IMPORTACAO
CD_MERCADORIA_NCM_REAL	Nenhuma
NM_MERCADORIA_NCM_DECL	OP_ADICAO_DE_IMPORTACAO
NM_MERCADORIA_NCM_REAL	Nenhuma
VULE_REAL	Nenhuma
NR_IMPORTADOR_REAL	Nenhuma

Essas tabelas são fundamentais para o cálculo das probabilidades de ocorrências das infrações aduaneiras. Cada registro contido na tabela DI_VERIFICADAS corresponde a uma adição de uma DI fiscalizada, independentemente da constatação ou não de infrações aduaneiras. A segunda tabela relaciona, para cada adição verificada, os tipos de infrações detectados.

As tabelas IA_CNPJ, IA_URF_D_URF_E, IA_PP_PO_PA, IA_NCM e IA_NCM_RT registram as probabilidades de não-ocorrência de uma infração aduaneira considerando as respectivas variáveis causais CNPJ, URF_D_URF_E, PP_PO_PA, NCM e NCM_RT. Por exemplo, a tabela IA_CNPJ armazena probabilidades do mecanismo de inibição estar ativo para cada um dos 18 tipos de infração aduaneira considerando um determinado CNPJ, ou seja, é a probabilidade de que não ocorra a infração para um dado CNPJ. Esse valor, o qual é representado por q_i , será denominado daqui por diante como “fator de inibição” e obtido com a seguinte fórmula:

$$q_i = \frac{\text{qtde de adições conferidas sem infração contendo a variável causal}}{\text{qtde de adições conferidas contendo a variável causal}}$$

A tabela VAR_EVIDENCIA descreve cada uma das variáveis causais ou de evidência existentes no modelo, e a tabela INFRACAO_EVIDENCIA relaciona, para cada tipo de infração, o conjunto de variáveis causais que a influenciam. A tabela TIPOS_INFRACOES descreve cada um dos tipos de infração. O diagrama Entidade-Relacionamento do banco de dados INFRACAO é mostrado a seguir.

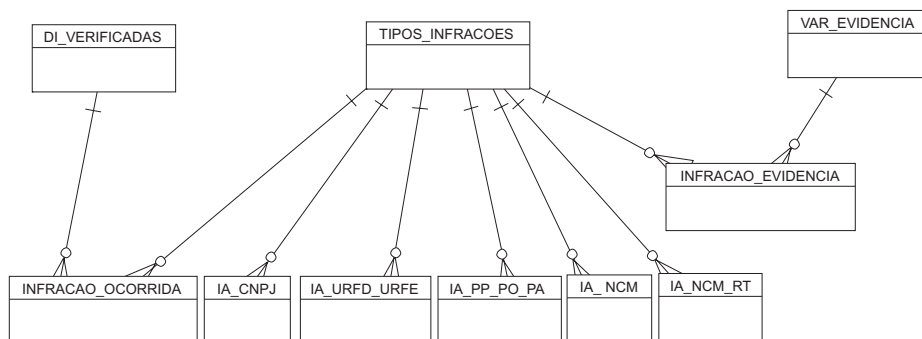


Figura 6 – Diagrama E-R do banco de dados INFRACAO.

Para armazenar as informações dinâmicas referentes à execução da Rede de Crença para cada DI, foi criado o banco de dados denominado IFA (Inteligência Fiscal Aduaneira), cujo diagrama E R é mostrado a seguir:

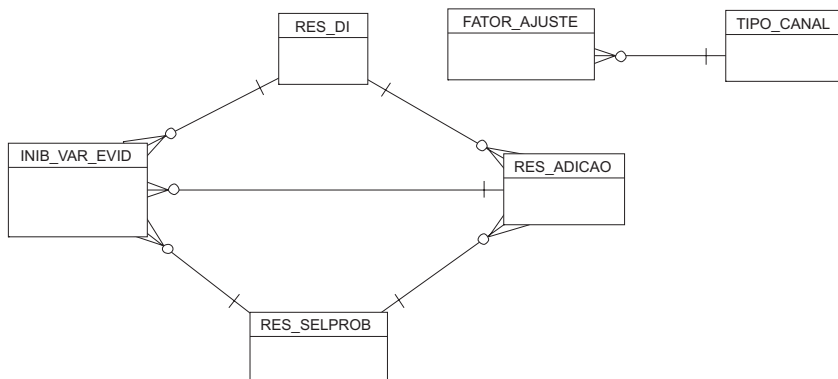


Figura 7 – Diagrama E-R do Banco de Dados IFA.

A tabela INIB_VAR_EVID contém os resultados de cada fator de inibição por variável causal, considerando cada adição e tipo de infração. A tabela RES_ADICAO contém o resultado dos produtos dos fatores de inibição e a estimativa de lançamento para cada adição e tipo de infração. A tabela RES_DI contém o resultado do fator de inibição geral para a DI e sua estimativa de lançamento total. A tabela RES_SELPROB armazena o resultado da seleção probabilística para cada DI, ou seja, o canal inicialmente atribuído à DI e o canal final de conferência após a execução do ajuste de lote de DIs, procedimento este que será comentado a seguir. A tabela TIPO_CANAL é auxiliar e apenas relaciona os diversos tipos de canais de conferência. O Canal Verde recebe o código '1', o Amarelo, código '2' e o Vermelho, código '3'. A tabela FATOR_AJUSTE também é auxiliar para efetuar o ajuste de lote, e ela armazena os percentuais de conferência para os canais amarelo e vermelho de cada unidade de despacho.

Uma vez construídos esses três bancos de dados, eles foram preenchidos com informações retiradas do sistema real, bem como com os resultados da execução da Rede de Crença.

A tabela DI_VERIFICADAS foi preenchida com os dados obtidos de informações reais. A tabela TIPOS_INFRACOES foi preenchida manualmente de acordo com os dados mostrados na tabela a seguir. Os fatores de agravo, cujos valores podem ser vistos a seguir, foram atribuídos de forma subjetiva.

Tabela 8 - Visão dos dados armazenados na tabela TIPOS_INFRACOES.

Tipo de Infração Aduaneira	Descrição Resumida	Penalidade e Mínima	Penalidade e Máxima	Fator de Agravo
1	Falsificação ou adulteração de documentos	N.A.	N.A.	1,2
2	Falsificação ou adulteração de mercadorias	N.A.	N.A.	1,1
3	Falsa declaração de conteúdo	N.A.	N.A.	1,0
4	Mercadoria atentatória à moral, aos bons costumes, à saúde ou à ordem pública	N.A.	N.A.	1,0
5	Importação sem licença de importação com emissão vedada ou suspensa	N.A.	N.A.	1,0
6	Ocultação do real importador mediante fraude ou simulação	N.A.	N.A.	1,2
7	Extravio de mercadoria	N.A.	N.A.	1,2
8	Inexistência da fatura comercial ou apresentação fora do prazo fixado	N.A.	N.A.	1,0
9	Apresentação da fatura comercial sem o visto consular, quando exigido	N.A.	N.A.	1,0
10	Diferença entre o preço declarado e o preço praticado ou arbitrado	N.A.	N.A.	0,8
11	Importação de mercadoria sem licença de importação de emissão permitida.	137,60	N.A.	1,0
12	Embarque da mercadoria antes da emissão da licença de importação	137,60	1.376,00	1,0
13	Embarque depois de vencido o prazo da licença de importação de mais de vinte até quarenta dias	137,60	1.376,00	1,0
14	Embarque depois de vencido o prazo da licença de importação em até vinte dias	137,60	1.376,00	1,0
15	Classificada incorretamente na Nomenclatura Comum do Mercosul	500,00	N.A.	1,0
16	Quantificada incorretamente na unidade de medida estatística	500,00	N.A.	1,0
17	Falta de pagamento, falta de declaração e declaração inexata	N.A.	N.A.	1,0
18	Falta de pagamento, falta de declaração e declaração inexata quando comprovada ocorrência de fraude	N.A.	N.A.	1,0

O registro das ocorrências que vão alimentar a tabela de Infrações Aduaneiras ocorridas será feito no Ambiente de Registro e Rastreamento da Atuação dos Intervenientes Aduaneiros (Radar). No sistema Radar a fiscalização declara as ocorrências no despacho aduaneiro ou em processo posterior de revisão do despacho.

Contudo, esse sistema somente entrou em operação no segundo semestre de 2002, e as DIs armazenadas são do período de janeiro de 1997 a setembro de 2002. Considerando que o histórico anterior é extenso, a alternativa foi povoar a tabela INFRACAO_OCORRIDA

com as informações obtidas pela comparação entre a versão original e final da DI e pela detecção de infrações lançadas em Auto de Infração.

Diversas infrações aduaneiras foram descobertas tomando por base a última versão da DI. Algumas infrações são verificadas observando, por exemplo, se existe o registro do pagamento de multas. Outras infrações podem ser inferidas pela comparação da última informação com a registrada originalmente. Acontecendo uma divergência de valores, fica demonstrada uma infração aduaneira. Por exemplo, a mudança da classificação fiscal (NCM) demonstra a ocorrência da infração tipo 15 (Classificada incorretamente na Nomenclatura Comum do Mercosul). A amostra do banco de dados do Siscomex contém todas as retificações de cada DI armazenada.

Após detecção das infrações e a inserção dos dados, o banco de dados INFRACAO possui as tabelas DI_VERIFICADAS, INFRACAO_OCORRIDA e TIPOS_INFRACOES corretamente preenchidas. Já as tabelas VAR_EVIDENCIA e INFRACAO_EVIDENCIA foram preenchidas, respectivamente, com informações sobre as variáveis causais e o relacionamento entre cada tipo de infração e suas respectivas variáveis causais, informações estas retiradas do modelo final da Rede de Crença. Os programas construídos são vistos a seguir.

4.2 Construção dos programas

As tabelas IA_CNPJ, IA_NCM, IA_PP_PO_PA, IA_URF_D_URF_E, IA_NCM_RT foram preenchidas por um programa que lia todas as DIs armazenadas em DI_VERIFICADAS e as infrações ocorridas para cada uma delas em INFRACAO_OCORRIDA. Esses programas foram denominados Preenche_IA_CNPJ, Preenche_IA_NCM, Preenche_IA_PP_PO_PA, Preenche_IA_URFD_URFE e Preenche_IA_NCM_RT.

O banco de dados IFA foi criado para armazenar os resultados obtidos da inferência probabilística. Na tabela INIB_VAR_EVID são guardados os fatores de inibição de cada variável causal presentes em cada DI, adição e tipo de infração aduaneira. A tabela RES_ADICAO armazena o fator de inibição e a estimativa de lançamento para cada DI e adição e tipo de infração aduaneira. A tabela RES_DI armazena o fator de inibi-

ção e a estimativa de lançamento para cada DI. Como já foi citado, o fator de inibição é o valor da probabilidade de não-ocorrência do evento, ou seja, é o valor de q_i .

O procedimento de verificação da existência de infrações aduaneiras e da determinação dos fatores de inibição foi executado uma única vez nesta simulação; porém, uma implementação prática exigirá uma atualização freqüente para permitir ao sistema se adaptar a novos casos.

Em seguida foram construídos os programas que lêem os valores das variáveis causais das DIs, calculam o valor de probabilidade geral de ocorrência de infração aduaneira e efetuam a estimativa de lançamento. As descrições desses programas são feitas na tabela a seguir:

Tabela 9 - Relação dos programas que calculam e armazenam a probabilidade geral de infração aduaneira para a DI e efetuam a seleção probabilística.

Nome do Programa	Descrição
probDI.java	Esta classe cria objetos probDI em que os valores de probabilidade e de lançamento são estimados para cada uma das DIs armazenadas no banco de dados Modelo. Alguns dados da DI, importantes para o cálculo, são armazenados nesses objetos. Para que os objetos tenham persistência, as informações serão armazenadas nas tabelas do banco de dados IFA.
setProbMod.java	Classe que lê o banco de dados modelo de DIs e cria um objeto probDI para cada uma. Ao final da Execução, existirá no banco de dados IFA o resultado de cada DI nas tabelas INIB_VAR_EVID, RES_ADICAO e RES_DI.
SeleProb.java	Este programa efetua a Seleção Probabilística. A Seleção Probabilística é feita em todas as DIs que tenham resultado já registrado no banco de dados IFA. Este programa determina todas as URFs distintas e as DIs para elas registradas, obtém os valores de probabilidades e efetua o ajuste de lote dessas DIs. O resultado final é gravado na tabela RES_SELPROB do banco de dados IFA.

Utilizando a probabilidade de ocorrência de infração para a DI e a estimativa de lançamento de crédito, é atribuído inicialmente um canal de conferência. Esse resultado é armazenado na tabela RES_SELPROB. Se a DI tiver probabilidade geral de ocorrência de alguma infração entre 80% e 100%, ela receberá o Canal Vermelho; caso esteja entre 60% e 80%, receberá o Canal Amarelo e, abaixo de 60%, receberá Canal Verde.

Entretanto, o canal atribuído inicialmente pode não ser o definitivo, pois um ajuste de lote será feito em seguida. Esse ajuste de lote é

necessário para adequar o número de DIs a serem conferidas à capacidade de mão-de-obra da unidade de despacho. Quando sobram DIs no Canal Vermelho, as de probabilidade menores serão ajustadas para amarelo. Se necessário, algumas DIs amarelas serão ajustadas para verde. A operação inversa também é verdadeira. O efeito colateral do ajuste do lote é que existe o risco de se liberar DIs com infração e conferir DIs sem infração.

Após a execução desses passos, a probabilidade de ocorrência de infração de cada DI e sua estimativa de lançamento, cuja forma de cálculo será vista a seguir, são conhecidos. Isso permite ordenar as DIs de cada URF_D e aplicar o ajuste de canais vermelhos e amarelos, o que é feito separando primeiro as DIs por unidade de despacho. Os valores de corte escolhidos (80% e 60%) para os canais também foram fruto de uma escolha subjetiva, observando os resultados parciais até então obtidos.

A estimativa de lançamento é efetuada com base nos valores fornecidos pelo importador na DI e pela aplicação da penalidade prevista para cada tipo de infração. As penalidades e os fatores de agravamento aplicados estão relacionados na tabela a seguir:

Tabela 10 - Fórmulas de cálculos dos lançamentos para cada tipo de infração Aduaneira.

Infração Aduaneira Tipo	Cálculo da Penalidade por adição	Mínimo	Máximo	Fator de Agravamento
1	VAL_PENAL=VALOR_ADUANEIRO	N.A.	N.A.	1,2
2	VAL_PENAL=VALOR_ADUANEIRO	N.A.	N.A.	1,1
3	VAL_PENAL=VALOR_ADUANEIRO	N.A.	N.A.	1,0
4	VAL_PENAL=VALOR_ADUANEIRO	N.A.	N.A.	1,0
5	VAL_PENAL=VALOR_ADUANEIRO	N.A.	N.A.	1,0
6	VAL_PENAL=VALOR_ADUANEIRO	N.A.	N.A.	1,2
7	VAL_PENAL=50%*VL_II_DEVIDO	N.A.	N.A.	1,2
8	VAL_PENAL=10%*VL_II_DEVIDO	N.A.	N.A.	1,0
9	VAL_PENAL=10%*VL_II_DEVIDO	N.A.	N.A.	1,0
10	VAL_PENAL=10%*VAL_ADU	N.A.	N.A.	0,8
11	VAL_PENAL=30%*VAL_ADU	137,60	N.A.	1,0
12	VAL_PENAL=30%*VAL_ADU	137,60	1.376,00	1,0
13	VAL_PENAL=20%*VAL_ADU	137,60	1.376,00	1,0
14	VAL_PENAL=20%*VAL_ADU	137,60	1.376,00	1,0
15	VAL_PENAL=1%*VAL_ADU	500,00	N.A.	1,0
16	VAL_PENAL=1%*VAL_ADU	500,00	N.A.	1,0
17	VAL_PENAL=75%*VL_II_DEVIDO +75%*VL_IPI_DEVIDO +75%*VL_DUMP_DEVIDO	N.A.	N.A.	1,0
18	VAL_PENAL=150%*VL_II_DEVIDO +150%*VL_IPI_DEVIDO +150%*VL_DUMP_DEVIDO	N.A.	N.A.	1,0

Obs.: N.A. – Não aplicável.

Dessa maneira, com a estimativa de lançamento de crédito tributário e a probabilidade de ocorrência, foi possível efetuar a seleção das DIs para os canais de conferência aduaneira. No capítulo a seguir são apresentados e analisados os resultados colhidos com a simulação efetuada para as 20.734 DIs.

CAPÍTULO 5

RESULTADOS E ANÁLISE

Neste capítulo são apresentados os resultados, bem como são efetuadas as análises pertinentes ao modelo proposto. Ele começa definindo as limitações que tanto a Seleção Parametrizada quanto a Seleção Probabilística possuem.

São mostrados os resultados obtidos com a Seleção Parametrizada e com a Seleção Probabilística. O percentual de erro é quantificado para indicar qual a precisão do sistema ao selecionar ou liberar as DIs em ambos os casos.

5.1 Limitações

Independentemente da técnica de seleção empreendida, seja aquela efetuada por meio de parâmetros fixados pelo usuário, seja calculando o valor de probabilidade de ocorrência de uma infração aduaneira, ela sempre estará sujeita a dois tipos de erros, discriminados a seguir:

Erro Tipo I - Falso positivo. A DI não contém nenhuma infração aduaneira a despeito do sistema afirmar que possui.

Erro Tipo II - Falso negativo. A DI contém alguma infração aduaneira, entretanto o sistema não é capaz de percebê-la.

O primeiro tipo de erro induzirá a seleção da DI para conferência, seja ela documental, seja física. Após o procedimento fiscal de conferência, a fiscalização prestará uma informação indicando a existência ou não de infrações aduaneiras. Em princípio, a seleção para conferência de uma DI sem infração mostra-se inútil; porém, não pode ser considerada uma perda completa, pois o sistema no futuro atualizará a probabilidade de ocorrência daquela infração, reduzindo a chance de o sistema selecionar casos idênticos ou semelhantes a esse.

O segundo tipo de erro, em tese, não poderá ser descoberto nem pelo sistema nem pela fiscalização. As informações existentes conduzem a um cálculo que resulta em um baixo valor de probabilidade de ocorrência de alguma infração aduaneira em tal DI. Para que o sistema fosse capaz de descobrir tal caso, seria necessário que ele pudesse

escolher alguns dos despachos liberados pelo sistema. Uma proposta é permitir uma seleção aleatória com um determinado percentual. Assim, o sistema adquire a possibilidade de detectar infrações em DIs, mesmo diante de uma baixa probabilidade de ocorrência de infração aduaneira.

5.2 Resultados obtidos

Com o objetivo de medir a eficiência do sistema, um banco de dados com 20.734 DIs escolhidas aleatoriamente foi utilizado. Dentre as 20.734 DIs foram encontradas 5.596 DIs fiscalizadas durante o despacho aduaneiro ou posteriormente a este. Nesse conjunto de 5.596 DIs, existem 2.272 DIs com pelo menos uma infração aduaneira relatada e 3.324 DIs sem registro de infração.

Os resultados da execução da Seleção Parametrizada são apresentados nas tabelas a seguir. A Tabela 11 mostra a quantidade de DIs selecionadas para os canais vermelho, amarelo e verde, uma parte com infração e outra sem. A Tabela 12 exhibe os resultados dos lançamentos de crédito tributário e os impostos pagos para cada canal de conferência aduaneira. A Tabela 13 mostra a eficiência da Seleção Parametrizada na seleção e liberação das DIs. A linha que representa o percentual de erro na Tabela 13 está associada aos dois tipos de erros citados anteriormente. O primeiro erro é denominado falso positivo, ou seja, o sistema seleciona a DI prevendo a existência de infração aduaneira que não se confirma. O segundo tipo é o falso negativo, ou seja, o sistema libera a DI que possui infração pressupondo que esta não existe.

Os resultados para a Seleção Parametrizada foram obtidos aplicando para todas as URFs_D um fator de ajuste igual a 10% no Canal Vermelho e 10% no Canal Amarelo.

Tabela 11 - Resultados da Seleção Parametrizada para o número de DIs.

Seleção Parametrizada						
Com Ajuste do Lote, Somente DIs fiscalizadas						
Canais		Com Infração Aduaneira		Sem Infração Aduaneira		Total de DIs
		Qtde. DIs	Percentual	Qtde. DIs	Percentual	
Selecionadas	Vermelho	527	23,19%	1.090	32,79%	1.617
	Amarelo	378	16,64%	630	18,95%	1.008
Subtotal		905	39,83%	1.720	51,74%	2.625
Liberadas	Verde	1.367	60,17%	1.604	48,26%	2.971
Total		2.272	100,00%	3.324	100,00%	5.596

Na Tabela 11, a coluna denominada “Com Infração Aduaneira” informa a quantidade de DIs que possuíam infração aduaneira e que o sistema selecionou para os canais de conferência vermelho e amarelo ou liberou pelo Canal Verde. Os dados demonstram que o sistema conseguiu detectar 39,83% das infrações constatadas. O restante das infrações detectadas, 60,17%, foi descoberto pela fiscalização aduaneira ou foi apontado espontaneamente pelo próprio importador.

Tabela 12 - Resultados da Seleção Parametrizada para o lançamento de crédito tributário.

Seleção Parametrizada				
Com Ajuste do Lote, Somente DIs fiscalizadas				
Canais		Com Infração Aduaneira		
		Lançamento (R\$)	II Pago (R\$)	IPI Pago (R\$)
Selecionadas	Vermelho	601.738,32	4.426.490,73	1.822.978,03
	Amarelo	304.111,81	3.382.713,24	1.494.515,64
Subtotal		905.850,13	7.809.203,97	5.794.369,00
Liberadas	Verde	3.628.757,77	15.844.285,97	5.902.676,66
Total		4.534.607,90	23.653.489,94	9.220.193,10

Na Tabela 12, a coluna “Lançamento (R\$)” indica o crédito tributário hipotético recuperado pela cobrança de diferença de impostos, multas e juros; já as colunas “II Pago” e “IPI Pago” indicam os valores de impostos pagos no registro da DI. O que essa tabela demonstra é um retorno hipotético muito baixo, em torno de 20% do valor total, considerando apenas as DIs selecionadas, ou seja, mais de 80% do valor lançado não poderia ser recuperado, ressaltando que isso é apenas uma simulação.

Tabela 13 - Percentuais de Acerto e Erro da Seleção Parametrizada com ajuste.

Percentuais(%)	Sistema Selecciona DI	Sistema Libera DI
Acerto	34,48	53,99
Erro	65,52 (Falso-positivo)	46,01 (Falso-negativo)

Os resultados a seguir mostram o resultado para a Seleção Probabilística executada para o mesmo banco de dados de 20.734 DIs, efetuado o mesmo ajuste de lote. A Tabela 14 mostra a quantidade de DIs selecionadas para os canais vermelho, amarelo e verde. A Tabela 15 exhibe os resultados dos lançamentos de crédito tributário para DIs com infração e os impostos pagos para cada canal de conferência aduaneira. A Tabela 16 mostra a eficiência percentual da Seleção Probabilística na seleção e liberação das DIs.

Tabela 14 - Resultados para o número de DIs da Seleção Probabilística com ajuste.

Seleção Probabilística						
Com Ajuste do Lote, Somente DIs fiscalizadas						
Canais		Com Infração Aduaneira		Sem Infração Aduaneira		Total de DIs
		Qtde. DIs	Percentual	Qtde. DIs	Percentual	
Selecionadas	Vermelho	1.072	47,18%	183	5,50%	1.255
	Amarelo	576	25,35%	304	9,15%	880
Subtotal		1.648	72,53%	487	14,65%	2.135
Liberadas	Verde	624	27,47%	2.837	85,35%	3.461
Total		2.272	100,00%	3.324	100,00%	5.596

A Tabela 14 mostra que a técnica conseguiu detectar 72% das DIs com infrações constatadas, com uma perda de 27% das DIs com infrações. A perda de receita gerada pelo lançamento de crédito tributário, mostrada na Tabela 15, é menor que o da Seleção Parametrizada e fruto da decisão de avaliar o potencial de lançamento de crédito tributário para ordenar as DIs.

Tabela 15 - Resultados para o lançamento de crédito da Seleção Probabilística com ajuste

Seleção Probabilística				
Com Ajuste do Lote, Somente DIs fiscalizadas				
Canais		Com Infração Aduaneira		
		Lançamento (R\$)	II Pago (R\$)	IPI Pago (R\$)
Selecionadas	Vermelho	2.660.513,27	10.923.387,95	3.388.884,04
	Amarelo	730.655,64	5.572.162,66	2.405.462,19
Subtotal		3.391.168,91	16.495.550,61	5.794.346,23
Liberadas	Verde	1.143.438,99	7.157.939,33	3.425.824,10
Total		4.534.607,90	23.653.489,94	9.220.170,33

Tabela 16 - Percentuais de Acerto e Erro da Seleção Probabilística com ajuste.

Percentuais(%)	Sistema Seleciona DI	Sistema Libera DI
Acerto	77,19	81,97
Erro	22,81 (Falso-positivo)	18,03 (Falso-negativo)

Comparando os resultados quanto ao número de DIs das Tabelas 11 e 14, verifica-se que o percentual de acerto da Seleção Parametrizada é menor que o da Seleção Probabilística. Seu desempenho foi prejudicado pela escolha aleatória efetuada no ajuste de lote que foi de 10% para o Canal Vermelho e de 10% para o Canal Amarelo. A figura 8 mostra qualitativamente a quantidade de DIs com infração selecionadas e de DIs sem infração liberadas para cada canal de conferência, considerando o tipo de seleção efetuado. Observe que cada coluna é dividida em três partes, e cada parte representa, de cima para baixo, os canais verde, amarelo e vermelho. Ele mostra que a Seleção

Parametrizada não é tão eficiente quanto a Seleção Probabilística para selecionar DIs com infração, e regularmente libera a maior parte. Analisando quanto ao aspecto de liberar DIs sem infração, a Seleção Parametrizada apresentou um desempenho melhor; porém, a Seleção Probabilística selecionou menos DIs sem infração para conferência.

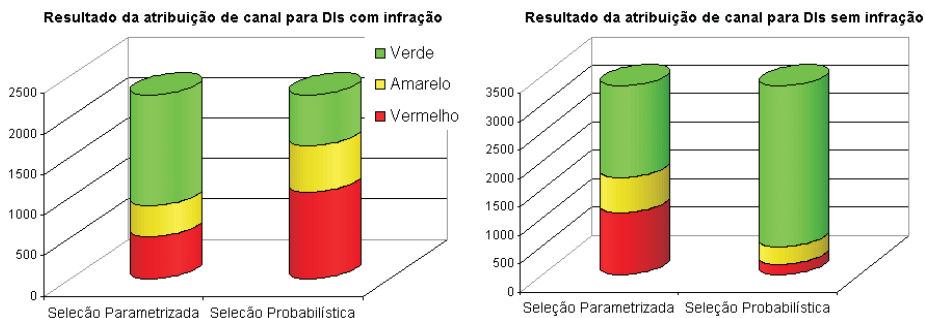


Figura 8 - Gráfico comparando a quantidade de DIs com infração e sem infração por canal.

Confrontando as Tabelas 12 e 15, nota-se que a Seleção Probabilística apresentou uma perda hipotética de receita de R\$ 1.143.438,99; já a perda para a Seleção Parametrizada seria de R\$ 3.628.757,77, um valor bem maior. O gráfico a seguir mostra qualitativamente o lançamento de crédito tributário de DIs com infração por canal de conferência atribuído ao tipo de seleção efetuada.

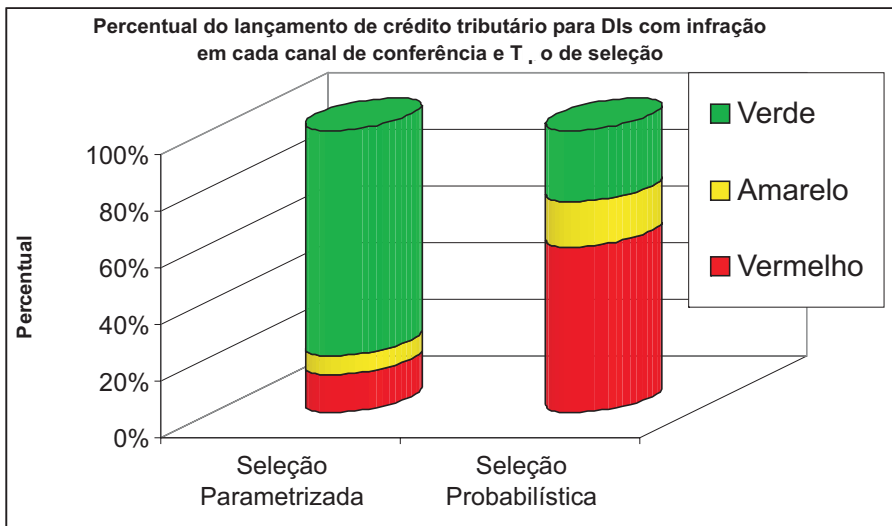


Figura 9 - Gráfico que mostra o percentual relativo de lançamento de crédito para DIs com infração por canal de conferência e tipo de seleção.

Observe que cada coluna no gráfico da figura 9 é dividida em três partes e cada parte representa, de cima para baixo, os canais verde, amarelo e vermelho. A Seleção Parametrizada apresenta uma perda maior quando comparada com a perda da Seleção Probabilística. Isso é explicado, principalmente, pela aleatoriedade com que a Seleção Parametrizada agrava ou abranda o canal de conferência de um despacho aduaneiro, enquanto a Seleção Probabilística utiliza o potencial de lançamento de crédito tributário para descartar ou incluir as DIs nos canais amarelo e vermelho.

Comparando qualitativamente os percentuais de erro entre as técnicas na figura 10, verifica-se que a Seleção Probabilística apresenta desempenho melhor, seja selecionando para conferência, seja liberando as DIs.

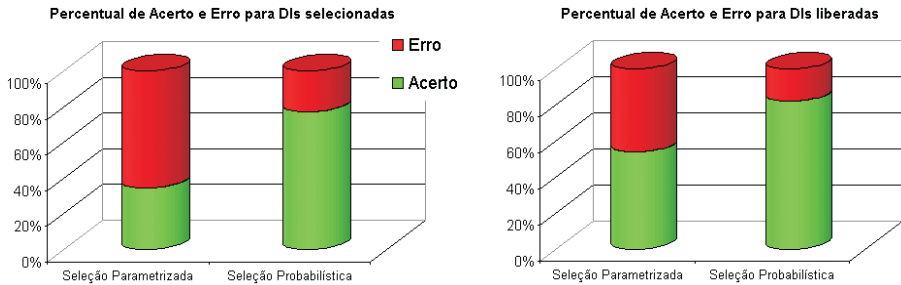


Figura 10 - Gráfico de comparação de acerto e erro para DIs selecionadas e liberadas.

A grande vantagem da Seleção Probabilística é a liberdade em estimar quão próximo se está da detecção de um tipo de infração aduaneira. Por não estar presa a parâmetros fixos e por característica do modelo *Noisy-OR Gate*, ela consegue estimar valores de probabilidade de casos ainda não totalmente conhecidos. Além disso, a Seleção Probabilística é mais eficiente quando é necessário o ajuste para agravar ou abrandar o canal de conferência da DI, já que a escolha não é feita de forma aleatória.

No capítulo a seguir serão feitas conclusões sobre o presente trabalho, bem como comentários que podem nortear trabalhos futuros sobre o tema.

CAPÍTULO 6

CONCLUSÃO

Os resultados apresentados no capítulo anterior mostraram que a Seleção Probabilística pode melhorar a eficiência da conferência aduaneira por adotar uma abordagem diferente para o problema.

A modelagem e implementação foram feitas com base nas informações atuais e testadas em um ambiente que simulou as condições reais. A solução permite que seja integrada ao sistema atual sem muitas modificações; basta acrescentar dois bancos de dados e as rotinas que fazem a inferência probabilística, a estimativa de lançamento de crédito tributário e a ordenação das DIs.

A implantação da Seleção Probabilística não implica a desativação completa da Seleção Parametrizada, muito pelo contrário. Deve ser reconhecido que esta é uma ferramenta gerencial para situações com que a primeira não pode lidar. Nada impede que seja feita uma combinação das duas, o que garantirá que parâmetros fixados pela Seleção Parametrizada sejam obedecidos e casos de DIs desconhecidas possam ser avaliados pela Seleção Probabilística. A escolha aleatória para agravar ou desagravar o canal de uma DI, feita pela Seleção Parametrizada, seria substituída pela ordenação feita pela Seleção Probabilística.

Difícilmente qualquer solução proposta para o problema apresentado seria perfeita; todas possuem algumas vantagens e desvantagens. A principal fraqueza na solução proposta reside no critério escolhido para ordenar as DIs. A ordem das DIs foi essencial para decidir sobre sua seleção ou liberação; por isso, diferentes técnicas de ordenação foram experimentadas, e a escolhida foi a que providenciou o melhor retorno financeiro. Uma proposta é seguir adiante com um estudo que considere a aplicação da Teoria de Decisão.

Não se pode deixar de registrar que o sistema é capaz de adquirir conhecimento probabilístico mediante a atualização das tabelas de probabilidade condicional. Assim, o próximo passo será avaliar o impacto resultante das alterações incrementais dos fatores de inibição, com base nas informações prestadas pela fiscalização.

Finalizando, a Seleção Probabilística apresentou nos testes um retorno financeiro muito bom e provou ser uma técnica alternativa viável. Pode-se dizer que o consumo de recursos computacionais é baixo, ainda que não tenham sido feitos testes reais. Ela pode não representar a solução perfeita para o problema; porém, pode significar o aumento da eficiência da seleção para conferência aduaneira e da recuperação de crédito tributário, com a vantagem de ser implantada com pouco esforço.

REFERÊNCIAS

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] BRASIL. Decreto n.º 4.543, de 26 de dezembro de 2002. Regulamenta a administração das atividades aduaneiras e a fiscalização, o controle e a tributação das operações de comércio exterior. **Diário Oficial da República Federativa do Brasil**, Poder Executivo, Brasília, DF, 27 dez. 2002. Seção 1, p. 151.
- [2] CRAMÉR, Harald. **Elementos da teoria da probabilidade e algumas de suas aplicações**. São Paulo: Mestre Jou, 1973. 330 p.
- [3] DÍEZ, F.J. Parameter adjustment in Bayes networks. The generalized Noisy-OR Gate. In: CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 9, July 1993, Washington D.C. **Proceedings...** San Francisco: Morgan Kaufmann, 1993, p. 99-105.
- [4] HECKERMAN, David; BREESE, John S. Causal independence for probabilistic assessment and inference using Bayesian networks. **IEEE Transactions on Systems, Man and Cybernetics**, [S.l.], v. 26, n. 6, p. 826-831, jun. 1996.
- [5] HUANG, Kurt; HENRION, Max. Efficient search-based inference for Noisy-OR Belief Networks: TopEpsilon. In: CONFERENCE ON

UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 12, Aug. 1996, Portland. **Proceedings...** San Francisco: Morgan Kaufmann, 1996. p. 446-454.

[6] JENSEN, Finn Verner. **An introduction to Bayesian networks**. New York: Springer Verlag, 1996. 178 p.

[7] KEARNS, Michael; MANSOUR, Yishay. Exact inference of hidden structure from sample data in Noisy-OR networks. In: CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 14., July 1998, Wisconsin. **Proceedings...** San Francisco: Morgan Kaufmann, 1998. p. 304-310.

[8] LUCAS, Peter. **Bayesian network modelling by qualitative patterns**. Disponível em: <<http://www.cs.kun.nl/~peterl/interact-ecai.ps.gz>>. Acesso em: 19 jan. 2003.

[9] NILSSON, Nils J. **Artificial Intelligence: a new synthesis**. San Francisco: Morgan Kaufmann, 1998. 513 p.

[10] PEARL, Judea. **Probabilistic reasoning in intelligent systems: networks of plausible inference**. San Francisco: Morgan Kaufmann, 1988. 552 p.

[11] PRADHAN, M. et al. Knowledge engineering for large belief networks. In: CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 10, July 1994, Seattle. **Proceedings...** San Francisco: Morgan Kaufmann, 1994. p. 484-490.

[12] RUSSEL, Stuart Jonathan; NORVIG, Peter. **Artificial Intelligence: a modern approach**. New Jersey: Prentice-Hall, 1995. 932 p.

[13] ZHANG, Nevin Lianwen. Inference with causal independence in the CPSC network. In: CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE, 11., Aug. 1995, Montreal. **Proceedings...** San Francisco: Morgan Kaufmann, 1995. p. 582-589.