

5

O mecanismo de identificação de fraudadores

Na seção 1.2 apresentamos de forma sintética o esquema de funcionamento do jogo “pega ladrão”. Agora vamos apresentar o *mecanismo de identificação de fraudadores*, que consiste em um sistema de computação humana com o objetivo de prever a ocorrência de fraudes. Esse mecanismo é composto de dois submecanismos:

- O mecanismo de classificação de vendedores, cujo objeto é prever, dado um conjunto de vendedores, quais cometerão fraudes.
- O mecanismo de incentivo aos jogadores, que visa alcançar um equilíbrio entre os gastos feitos para incentivá-los a colaborarem na classificação de vendedores e o retorno obtido com a detecção antecipada de fraudadores.

Na seção 5.1 apresentaremos o modelo da atuação dos fraudadores que será utilizado daqui em diante. Esse modelo baseia-se nos resultados do Capítulo 3. Na seção 5.2 apresentaremos o modelo de atuação dos jogadores que será utilizado nos mecanismos. Nas seções 5.3 e 5.4 apresentaremos os dois submecanismos que formam o mecanismo de identificação de fraudadores. Na seção 5.5 avaliaremos o funcionamento do mecanismo através de simulações, mostrando a viabilidade do uso do mesmo com parâmetro realistas.

5.1.

Modelo da atuação fraudulenta

O esquema fraudulento no qual nos detivemos foi o da fraude de não-entrega, no qual o fraudador anuncia muitos produtos simultaneamente. Esse esquema pode ser percebido com facilidade *a posteriori*, já que os vários clientes enganados vão denunciá-lo com suas qualificações, especialmente nos comentários textuais. Na Figura 9 apresentamos ciclo de vida típico de uma identidade fraudulenta usada para perpetrar esta fraude.

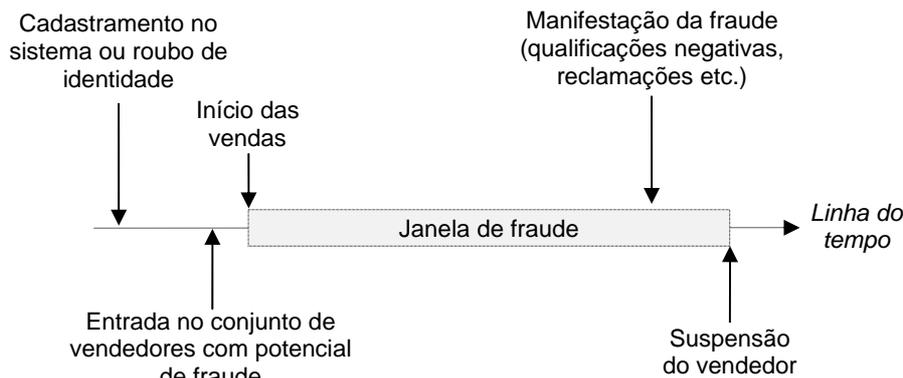


Figura 9: ciclo de vida de uma identidade fraudulenta

O primeiro passo é a obtenção de uma identidade com possibilidade de anunciar produtos. O fraudador pode cadastrar-se diretamente ou então tentar roubar senhas de contas já existentes. No momento oportuno, ele anuncia várias unidades de produtos fictícios: é o momento em que ele passa a ter *potencial de fraude*. Como normalmente os produtos são anunciados em condições atraentes de preço, prazo de entrega etc., pouco tempo depois as vendas começam, abrindo então a *janela de fraude*: o tempo decorrido entre a primeira venda e a suspensão do vendedor por parte do operador, seja por conta das reclamações dos compradores, seja porque o operador dispõe de outras evidências de fraude. Um fraudador habilidoso pode prolongar a janela de fraude oferecendo explicações plausíveis para a “demora” na entrega, dissuadindo os compradores de reclamarem.

Em algum momento os primeiros compradores percebem o engodo e qualificam negativamente o fraudador, possivelmente recorrendo também aos mecanismos que o sítio disponha para denunciar fraudes. Nesse momento torna-se muito mais difícil vender, já que os possíveis clientes verão os comentários negativos recentes e ficarão receosos. Pouco tempo depois ocorre a suspensão da identidade fraudulenta, fechando a janela de fraude e encerrando seu ciclo de vida¹⁶.

Do ponto de vista do funcionamento dos mecanismos, há dois momentos relevantes nesse ciclo: (i) a obtenção do potencial de fraude e (ii) o término da janela de fraude. O primeiro momento é aquele em que o operador tem que tomar uma decisão se considera o vendedor em questão suspeito ou não, recorrendo ao mecanismo de identificação de fraudadores. No segundo momento obtêm-se evidências de que o vendedor de fato era um fraudador.

¹⁶ No caso de ter ocorrido um roubo de identidade, o operador pode posteriormente apagar as qualificações negativas e reabilitar a mesma.

Na prática, a janela de fraude variará de golpe para golpe. Para concluir que um determinado vendedor é normal, o operador pode utilizar o valor máximo esperado da janela de fraude: se após esse período não houver qualificações negativas o suficiente, então o vendedor será considerado normal.

O operador pode interferir no ciclo de vida, suspendendo a identidade fraudulenta antes de haver reclamações. Isto ocorre quando ele dispõe de outras informações comprometedoras acerca da identidade em questão. Caso já tenham sido efetuadas algumas vendas verdadeiras¹⁷, então o ciclo de vida passa a ser o mostrado na Figura 10. Os clientes que compraram não vão receber e alguns reclamarão; ocorrerá apenas uma redução “forçada” da janela de fraude.

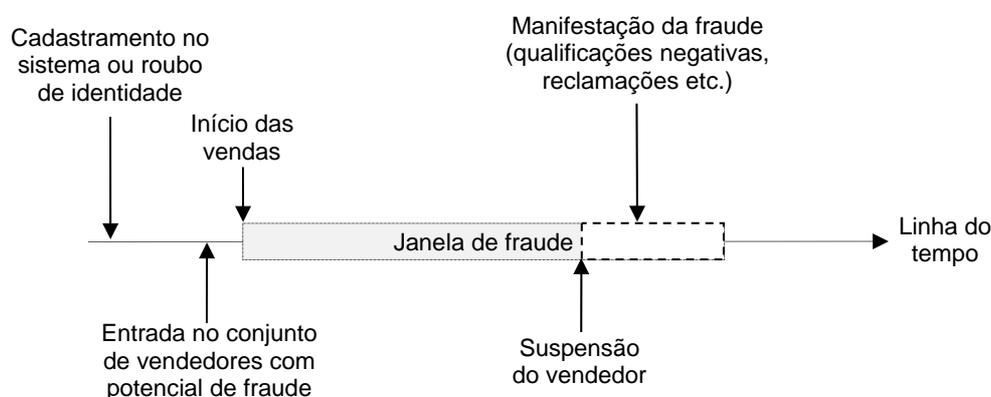


Figura 10: ciclo de vida com suspensão prematura do vendedor

Caso o vendedor seja suspenso *antes* de serem feitas vendas verdadeiras, então os ciclos mostrados já não serão mais válidos: faltarão manifestações de que ocorreu algo errado. O operador ficará sem uma confirmação mais cabal de que a identidade suspensa de fato pertencia a um fraudador. Ele poderá utilizar outros indícios para decidir se acertou ou errou ao suspender o vendedor; por exemplo, se o vendedor em questão não responder aos questionamentos do operador, então ele passa a ser considerado um fraudador. Contudo, são evidências mais frágeis que as proporcionadas pelas reclamações.

Com relação à redução dos prejuízos do operador, interessa antecipar ao máximo a suspensão das identidades fraudulentas. Todavia, o jogo precisa da confirmação de se ocorreu fraude ou não, já que a premiação dos jogadores depende do quanto eles acertaram. Assim sendo, o momento da suspensão das identidades deve ser escolhido levando em conta a necessidade de coletar evidências de fraude

¹⁷ Usamos a expressão *vendas verdadeiras* para diferenciar de possíveis vendas “fictícias”, simuladas pelo fraudador para aumentar a sensação de segurança dos clientes verdadeiros.

para informar ao jogo, através da mensagem “comportamento dos vendedores avaliados”, quais vendedores cometeram fraude. Vamos assumir que todos os fraudadores efetivamente cometem fraude, ou seja, todos os produtos vendidos por eles não são entregues.

Outro elemento importante da atuação fraudulenta é a *quantidade de fraudadores presentes no meio dos vendedores normais*. Vamos assumir que a probabilidade de um vendedor escolhido ao acaso ser um fraudador varia lentamente, o que permite que tratemos essa probabilidade como uma constante, ao menos durante certo período de tempo.

5.2.

Modelo da atuação dos jogadores

Agora nos deteremos na análise do comportamento dos jogadores frente à tarefa proposta, que é a de classificar um dado vendedor como normal ou fraudador. Vamos modelar um jogador genérico como um *classificador probabilístico*, isto é, que acerta a classificação dos vendedores com determinada probabilidade. Daqui em diante vamos chamar os jogadores de *classificadores*.

Para tornar o mecanismo o mais genérico o possível, vamos assumir que o conjunto de classificadores é arbitrário, isto é, não há restrições de entrada. Essa suposição é utilizada no sítio Mechanical Turk, onde muitas tarefas podem ser desempenhadas por pessoas sem qualquer uso prévio do sistema. Nada impede, no entanto, que o mecanismo registre informações acerca do desempenho dos classificadores e que tome decisões com base nessas informações.

A modelagem probabilística que vamos propor inspira-se em trabalhos da área de sistemas de reputação, de recomendação e de projeto de mecanismos (Dellarocas, 2005; Jurca, 2007; Miller et al., 2005). Vamos assumir que todo vendedor possui um tipo $\theta \in \{N, F\}$, onde N significa normal e F , fraudador, sendo que o vendedor sempre se comporta de acordo com seu tipo (no caso, o vendedor normal entrega os produtos vendidos e o fraudador, não). Ao analisarem um vendedor, os classificadores percebem um *signal* (imperfeito) do tipo do vendedor. Esse sinal é uma variável aleatória. Dado um agente humano i e um vendedor v , o agente humano percebe um sinal $S_i(v)$ do verdadeiro tipo do vendedor, $\theta(v)$. Vamos assumir as seguintes premissas, que são análogas às adotadas nos trabalhos supracitados:

- O tipo $\theta(v)$ do vendedor a ser analisado é uma variável aleatória que segue a distribuição de Bernoulli.
- O sinal $S_i(v)$ é uma variável aleatória de Bernoulli, com $S_i(v) = 1$ quando o classificador considera o vendedor como sendo do tipo fraudador e $S_i(v) = 0$, caso contrário. O classificador acertará o verdadeiro tipo do vendedor com probabilidades possivelmente diferentes, condicionadas pelo tipo do vendedor em questão. Por exemplo, o classificador pode acertar mais facilmente o tipo de um vendedor normal que o de um fraudador.

Na Tabela 10 sintetizamos em notação matemática as informações acima. Para simplificar as expressões, daqui em diante utilizaremos a notação mostrada na Tabela 11 para os eventos e variáveis aleatórias em questão.

Tabela 10: premissas acerca dos classificadores

Tipo de vendedor	Distribuição	Probabilidade em questão
$\theta(v) = N$	$S_i(v) \sim \text{Bernoulli}(p)$	$\Pr[S_i(v) = 1 \theta(v) = N] = p$
$\theta(v) = F$	$S_i(v) \sim \text{Bernoulli}(p')$	$\Pr[S_i(v) = 1 \theta(v) = F] = p'$

Tabela 11: notação utilizada

Variável / evento / probabilidade	Abreviação
$S_i(v)$	S_i
$\theta(v)$	θ
$\Pr[\theta(v) = F]$	$\Pr(F)$
$\Pr[\theta(v) = N]$	$\Pr(N)$
$\Pr[S_i(v) = x]$	$\Pr(S_i)$
$\Pr[\theta(v) = y]$	$\Pr(\theta)$

Dados os dois tipos de vendedores, temos que:

$$\Pr(S_i) = \Pr(S_i | N) \Pr(N) + \Pr(S_i | F) \Pr(F) \quad (5.1)$$

No entanto, a informação que nos interessa é a probabilidade do vendedor ter determinado tipo dado o sinal recebido pelo agente humano: $\Pr(\theta | S_i)$.

$$\Pr(\theta | S_i) = \frac{\Pr(S_i | \theta) \Pr(\theta)}{\Pr(S_i)} \quad (5.2)$$

Fazendo a substituição:

$$\Pr(\theta | S_i) = \frac{\Pr(S_i | \theta) \Pr(\theta)}{\Pr(S_i | N) \Pr(N) + \Pr(S_i | F) \Pr(F)} \quad (5.3)$$

É interessante notar que o classificador só traz uma nova informação quando $\Pr(S_i | N) \neq \Pr(S_i | F)$.

Como assumimos que não há barreiras de entrada, levanta-se o problema de como obter os parâmetros de cada classificador. Uma possível solução é utilizar inferência bayesiana: inicialmente utilizam-se valores-padrão para os parâmetros dos novos classificadores (probabilidades *a priori*) e depois se atualizam essas probabilidades a cada nova resposta do classificador junto com o tipo verdadeiro do vendedor. O problema concreto fica sendo o de estimar o parâmetro de uma distribuição binomial a partir das sucessivas observações. Pode-se utilizar como valor-padrão de cada parâmetro o valor médio desse mesmo parâmetro na população de classificadores mais experientes, isto é, dos quais já se possui uma quantidade grande de respostas.

5.3. Mecanismo de classificação de vendedores

O mecanismo de classificação de vendedores especifica um *classificador binário de vendedores a partir dos seus perfis*, onde um perfil é uma identidade virtual junto com um determinado estado (produtos anunciados, produtos vendidos, reputação etc.). Cada vendedor pode ser classificado como *aprovado* ou *reprovado*. O objetivo do classificador é o de reprovar todos os fraudadores e aprovar todos os vendedores normais.

Na Figura 11 vemos o esquema do mecanismo de classificação de vendedores utilizando computação humana. O operador escolhe um conjunto de vendedores a serem analisados através jogo, procurando utilizar critérios que maximizem a chance de encontrar fraudadores. Por exemplo, o mecanismo pode escolher os vendedores dentre as dez categorias de produtos com mais fraudes. A seleção de vendedores pode ser feita periodicamente: por exemplo, quando um vendedor passa a ter potencial de fraude pela primeira vez, ele é incluído na próxima lista a ser analisada. Essa lista é enviada ao jogo através da mensagem “vendedores a analisar”.

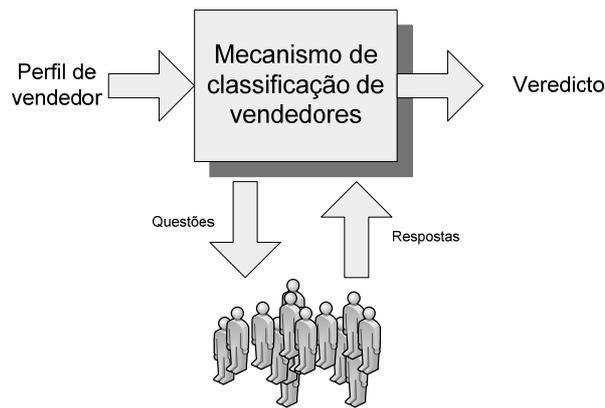


Figura 11: mecanismo de classificação baseado em computação humana

O mecanismo conta com a colaboração de certo número de agentes humanos, os *classificadores*, para efetuar a classificação dos perfis de vendedores. O jogo distribui os perfis de vendedores a serem analisados entre os classificadores (mensagem “distribuição entre os jogadores”). Cada classificador analisa os perfis que lhe foram atribuídos e informa sua resposta com relação a cada um deles, aprovando-o (classificando-o como normal) ou reprovando-o (classificando-o como fraudador). Note-se que há uma diferença importante com relação aos sistemas de reputação e de recomendação: o classificador não relata uma experiência que teve com o vendedor e sim tenta prever seu comportamento futuro.

O mecanismo então fica responsável por agregar as respostas dos diversos classificadores (mensagem “avaliações dos jogadores”) e produzir um *veredicto*, isto é, uma classificação final de cada vendedor analisado, de forma que o conjunto de classificações feitas (mensagem “avaliação do jogo sobre os vendedores”) esteja de acordo com os parâmetros do operador em termos de precisão e recuperação. O veredicto acerca de cada vendedor será *aprovado*, quando o vendedor for classificado como normal, ou então *reprovado*, quando o vendedor for classificado como fraudador.

Uma premissa importante é a de que a quantidade de fraudadores presente no conjunto de vendedores a serem classificados não deve ser muito pequena; caso contrário, os classificadores terão que analisar muitos vendedores para achar algum, o que seria desmotivador. Se a fração de fraudadores no conjunto analisado for de 0,1%, os classificadores podem ter que analisar centenas de vendedores para achar um fraudador. Isso é inviável: além de tornar o jogo maçante, pode acabar fazendo com que os classificadores se “acostumem” com o fato de que recebem muitos vendedores normais para analisar, o que aumenta a chance de

deixarem escapar um fraudador quando este aparecer. É necessário que o conjunto de vendedores a ser analisado tenha uma porcentagem mínima de fraudadores (na média) para evitar esses inconvenientes. Essa tarefa pode não ser tão difícil, já que os operadores de mercados eletrônicos têm acesso a muitas informações históricas sobre os perfis dos fraudadores, além da própria competência para detectar fraudes, o que permite achar esses subconjuntos de vendedores com maior probabilidade de serem fraudadores.

5.3.1. Integrando as respostas de múltiplos classificadores

Vamos agora examinar a questão de como integrar as respostas de diversos classificadores para chegar ao veredicto do mecanismo. Uma primeira possibilidade seria a de atualizar as probabilidades *a priori* de um vendedor ser de um determinado tipo a cada novo sinal recebido dos classificadores. Vamos assumir que as distribuições de probabilidade dos sinais recebidos por diferentes classificadores são independentes. Do ponto de vista teórico observamos que é uma hipótese frequentemente utilizada (Jurca, 2007; Miller et al., 2005).

Dados dois sinais S_i e S_j recebidos nessa seqüência, temos que:

$$\Pr(\theta | S_j) = \frac{\Pr(S_j | \theta) \Pr(\theta | S_i)}{\Pr(S_j | N) \Pr(N | S_i) + \Pr(S_j | F) \Pr(F | S_i)} \quad (5.4)$$

O mesmo processo pode ser repetido diversas vezes, tantos quantos forem os sinais disponíveis. Para visualizar o resultado que pode ser obtido, faremos um pequeno exemplo numérico. Dadas as probabilidades na Tabela 12, vamos calcular a probabilidade *a posteriori* de F dado que $S_i = 1$ e $S_j = 0$. As probabilidades omitidas são complementares às mostradas.

Tabela 12: probabilidades do exemplo

$\Pr(S_i = 1 F)$	30%
$\Pr(S_i = 1 N)$	10%
$\Pr(S_j = 1 F)$	40%
$\Pr(S_j = 1 N)$	30%

Inicialmente calculamos a nova probabilidade baseados no primeiro sinal:

$$\Pr(F | S_i = 1) = \frac{\Pr(S_i = 1 | F) \Pr(F)}{\Pr(S_i = 1 | N) [1 - \Pr(F)] + \Pr(S_i = 1 | F) \Pr(F)} \quad (5.5)$$

$$\Pr(F | S_i = 1) = \frac{0,3 \cdot 0,01}{0,1 \cdot (1 - 0,01) + 0,3 \cdot 0,01} = 2,94\%$$

Agora podemos fazer o cálculo com o segundo sinal:

$$\Pr(F | S_j = 0) = \frac{\Pr(S_j = 0 | F) \Pr(F | S_i = 1)}{\Pr(S_j = 0 | N) \Pr(N | S_i = 1) + \Pr(S_j = 0 | F) \Pr(F | S_i = 1)} \quad (5.6)$$

$$\Pr(F | S_j = 0) = \frac{0,6 \cdot 0,0294}{0,7 \cdot (1 - 0,0294) + 0,6 \cdot 0,0294} = 2,53\%$$

Podemos também calcular as probabilidades em questão utilizando os sinais simultaneamente:

$$\Pr(F | S_i = 1, S_j = 0) = \frac{\Pr(S_i = 1, S_j = 0 | F) \Pr(F)}{\Pr(S_i = 1, S_j = 0 | N) \Pr(N) + \Pr(S_i = 1, S_j = 0 | F) \Pr(F)}$$

$$\Pr(F | S_i = 1, S_j = 0) = \frac{\Pr(S_i = 1 | F) \Pr(S_j = 0 | F) \Pr(F)}{\Pr(S_i = 1 | N) \Pr(S_j = 0 | N) \Pr(N) + \Pr(S_i = 1 | F) \Pr(S_j = 0 | F) \Pr(F)} \quad (5.7)$$

$$\Pr(F | S_i = 1, S_j = 0) = \frac{0,3 \cdot 0,6 \cdot 0,01}{0,1 \cdot 0,7 \cdot 0,99 + 0,3 \cdot 0,6 \cdot 0,01} = 2,53\%$$

Vamos agora especificar como o mecanismo irá classificar o vendedor em análise integrando as múltiplas respostas recebidas. Nosso espaço amostral passa a ser o conjunto dos vetores na forma $\hat{s} = (s_1, \dots, s_n)$, $\hat{s} \in \{0,1\}^n$, onde \hat{s} é uma realização da variável aleatória \hat{S} . Com isso, temos que:

$$\Pr(\theta, \hat{S}) = \Pr(\theta | \hat{S}) \Pr(\hat{S}) \quad (5.8)$$

O mecanismo terá que associar a cada veredicto um evento, isto é, um conjunto de valores \hat{s} . Chamaremos de S^F o evento que será utilizado para reprovar um vendedor¹⁸. Temos então que:

$$\Pr(S^F | \theta) = \sum_{\hat{s} \in S^F} \Pr(\hat{S} = \hat{s} | \theta)$$

$$= \sum_{\hat{s} \in S^F} \prod_{i=1}^n \Pr(S_i = s_i | \theta) \quad (5.9)$$

¹⁸ Naturalmente o outro evento (o utilizado para aprovar um vendedor) será o complementar deste.

Retomando o exemplo, se utilizarmos $S^F = \{(1,1)\}$, teremos que $\Pr(S^F | F) = 0,3 \cdot 0,4 = 12\%$. Analogamente teremos que $\Pr(S^F | N) = 0,1 \cdot 0,3 = 3\%$. Ou seja: com esse conjunto de agentes e com esse conjunto de sinais “negativos” (que levam à reprovação), seria possível reprovar 12% dos fraudadores à custa de reprovar 3% dos vendedores normais. Se ao invés disso utilizássemos $S^F = \{(0,1), (1,0), (1,1)\}$, teríamos que $\Pr(S^F | F) = 0,3 \cdot 0,4 + 0,3 \cdot 0,6 + 0,7 \cdot 0,4 = 58\%$ e $\Pr(S^F | N) = 37\%$.

Esses exemplos numéricos ilustram o fato de que a escolha de S^F depende da quantidade de vendedores normais que o operador tolera que sejam reprovados. Quanto maior a cardinalidade de S^F , maior a chance de reprovar fraudadores e também maior a chance de reprovar normais. O desafio está em achar a medida adequada. Aqui entra em cena a *tolerância a erros*, que definiremos como sendo o *valor esperado da fração de vendedores normais que o operador aceita que sejam erroneamente reprovados pelo mecanismo*. Esse parâmetro será denotado por P_{RN}^{\max} e a restrição que o mecanismo deve ater-se é a de que $\Pr(S^F | N) \leq P_{RN}^{\max}$. A tarefa do mecanismo será então a de resolver o seguinte problema:

$$S^F = \arg \max_{\hat{S} \subset 2^{\{0,1\}^n}, P(\hat{S}|N) \leq P_{RN}^{\max}} \Pr(\hat{S} | F) \quad (5.10)$$

Em outras palavras, o mecanismo teria que encontrar o evento que maximiza a reprovação de fraudadores *dentro do conjunto de eventos que reprovam uma quantidade aceitável de vendedores normais*.

Uma busca exaustiva implicaria testar 2^{2^n} combinações, o que é inviável para um número um pouco maior de classificadores. Para reduzir a quantidade de alternativas a serem testadas, vamos alterar o espaço amostral, utilizando uma função que mapeia os eventos do espaço antigo no novo espaço, sendo $\hat{S} = (S_1, \dots, S_n)$ uma variável aleatória com domínio no espaço antigo e V uma variável aleatória com domínio no novo espaço:

$$V = f(\hat{S}) = \sum_{i=1}^n S_i \quad (5.11)$$

Intuitivamente, passamos a fazer uma *votação*: levamos em conta apenas a quantidade de sinais pró-reprovação. Na Figura 12 vemos um exemplo dessa mudança quando há dois classificadores.

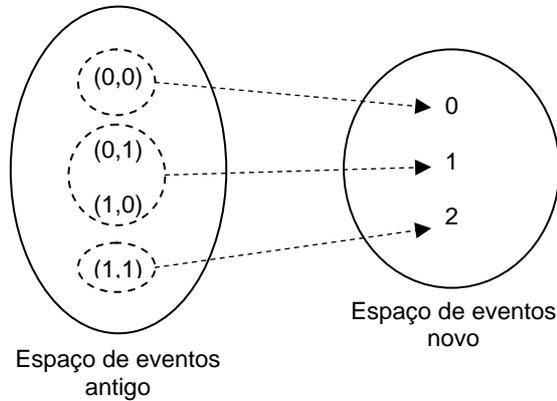


Figura 12: mudança no espaço de eventos

Com essa modificação no espaço amostral, a quantidade de eventos diferentes cai de 2^{2^n} para 2^n . Obviamente essa redução tem um custo: pode haver eventos no espaço amostral original que não estão representados no novo espaço (e.g. $\{(0,0), (1,1)\}$) e que ofereceriam um resultado melhor para o mecanismo que quaisquer dos eventos no novo espaço.

Como o valor da nova variável aleatória V é dado pela soma dos valores de outras variáveis aleatórias, podemos calcular sua distribuição através da *convolução* das funções de densidade de probabilidade (Grinstead & Snell, 1997):

$$\Pr(V = x | \theta) = \Pr\left(\sum_{i=1}^n S_i = x | \theta\right) = f_{S_1}(x | \theta) * \dots * f_{S_n}(x | \theta) \quad (5.12)$$

Onde f_{S_i} é a função densidade de probabilidade da variável aleatória S_i , dada por:

$$f_{S_i}(x | \theta) = f_{Bernoulli}(x, p_{i,\theta}) = \begin{cases} p_{i,\theta} & \text{se } x = 1 \\ 1 - p_{i,\theta} & \text{se } x = 0 \end{cases} \quad (5.13)$$

Nessa equação, $p_{i,\theta} = \Pr(S_i = 1 | \theta)$, isto é, a probabilidade do i -ésimo classificador apontar um vendedor como fraudador *dado que ele é do tipo* θ . A convolução entre duas funções de densidade de probabilidade é dada por:

$$f(x) * g(x) = \sum_{z=-\infty}^{+\infty} f(z-x)g(x) \quad (5.14)$$

Quando os sinais são identicamente distribuídos, isto é, todos os classificadores acertam com a mesma probabilidade, então V tem uma distribuição binomial (Grinstead & Snell, 1997).

Com o novo espaço amostral, os eventos a serem considerados pelo mecanismo para decidir a classificação do vendedor passam a ter a seguinte forma:

$$S^F \subset 2^{\{0, \dots, n\}} \quad (5.15)$$

No caso de haver cinco classificadores, um exemplo seria $S^F = \{3, 5\}$: o vendedor seria reprovado se tivesse três ou cinco votos contra.

Agora podemos introduzir uma nova simplificação, visando não só reduzir o espaço de busca, mas também tornar a atuação do mecanismo mais intuitiva. Trata-se de decidir o veredicto com base em uma quantidade máxima de votos que um vendedor pode receber sem ser reprovado, quantidade que chamaremos *limiar de reprovação* – l . Com isso, as decisões e as probabilidades de interesse do mecanismo passam a ser as mostradas na Tabela 13.

Tabela 13: probabilidades com votação

Evento	Veredicto	Probabilidades	
$V \leq l$	Aprovado	$\Pr(V \leq l F)$	Probabilidade de aprovar um fraudador
		$\Pr(V \leq l N)$	Probabilidade de aprovar um vendedor normal
$V > l$	Reprovado	$\Pr(V > l F)$	Probabilidade de reprovar um fraudador
		$\Pr(V > l N)$	Probabilidade de reprovar um vendedor normal

Agora o cálculo das probabilidades se reduz ao uso das distribuições acumuladas e suas complementares – $\Pr(V \leq l)$ e $\Pr(V > l)$. Uma consequência importante disso é que as probabilidades de reprovação passam a ser *funções monótonas decrescentes* de l : aumentam quando l diminui. Essa nova simplificação tem efeito análogo à anterior: alguns eventos que ofereceriam um resultado melhor para o mecanismo podem não ser contemplados.

Daqui em diante utilizaremos no modelo as probabilidades de reprovação – $\Pr(V > l | F)$ e $\Pr(V > l | N)$, que podem ser obtidas através das fórmulas a seguir, onde F_{ξ} e G_{ξ} são funções de distribuição de probabilidade acumuladas e f_{ξ} é uma função densidade de probabilidade:

$$\begin{aligned}
\hat{p}_\theta &= (p_{1,\theta}, \dots, p_{n,\theta}) \\
\Pr(V \leq l | \theta) &= F_{\hat{s}}(l, \hat{p}_\theta) \\
F_{\hat{s}}(l, \hat{p}_\theta) &= \sum_{x=0}^l f_{\hat{s}}(x, \hat{p}_\theta) \\
f_{\hat{s}}(x, \hat{p}_\theta) &= f_1(x, p_{1,\theta}) * \dots * f_n(x, p_{n,\theta}) \\
\Pr(V > l | \theta) &= G_{\hat{s}}(x, \hat{p}_\theta) = 1 - \Pr(V \leq l | \theta) = 1 - F_{\hat{s}}(x, \hat{p}_\theta)
\end{aligned} \tag{5.16}$$

Definiremos agora como fica a decisão do mecanismo com base nesse novo critério. Seja $p_{RF_i} = \Pr(S_i = 1 | F)$ a probabilidade do i -ésimo classificador reprovar um fraudador e $\hat{p}_{RF} = (p_{RF_1}, \dots, p_{RF_n})$; analogamente, seja $p_{RN_i} = \Pr(S_i = 1 | N)$ e $\hat{p}_{RN} = (p_{RN_1}, \dots, p_{RN_n})$. Dados esses parâmetros e dada a monotonicidade das probabilidades de reprovação, o problema do mecanismo passa a ser então o de encontrar o menor valor de l tal que $\Pr(V > l | N) \leq P_{RN}^{\max}$. Exemplificando, se o mecanismo dispuser de dez classificadores, todos com $p_{RF} = 40\%$ e $p_{RN} = 30\%$, e $P_{RN}^{\max} = 10\%$, então o mecanismo reprova todos os vendedores que receberem mais de cinco votos, com $\Pr(V > 5 | N) = 4,73\%$ e $\Pr(V > 5 | F) = 16,62\%$. Se repetirmos o cálculo com vinte classificadores, então teremos que o mecanismo reprova os vendedores com mais de nove votos, com $\Pr(V > 9 | N) = 4,8\%$ e $\Pr(V > 9 | F) = 24,5\%$. Note-se que acrescentar novos agentes com o mesmo desempenho permite aumentar arbitrariamente a reprovação de fraudadores mantendo o mesmo limite de reprovação de vendedores normais.

Agora já é possível descrever um roteiro de funcionamento do mecanismo para classificar os vendedores:

- Para cada vendedor, seleciona o maior número possível de classificadores para avaliarem-no.
- Obtém os vetores \hat{p}_{RF} e \hat{p}_{RN} .
- Totaliza os votos dos classificadores.
- Calcula a quantidade de votos l que faz com que $\Pr(V > l | N)$ mais se aproxime de P_{RN}^{\max} , dados os parâmetros dos classificadores, e utiliza esse valor como limiar de reprovação:

$$l = \min \{x \in \mathbb{N} \mid G_{\hat{s}}(x, \hat{p}_{RN}) \leq P_{RN}^{\max}\} \quad (5.17)$$

- Reprova todos os vendedores que tenham recebido mais votos que esse limiar.

A aplicação dessa sequência trará os resultados mostrados na Tabela 14, que se baseiam na equação (5.16). O valor esperado da quantidade de fraudadores é dado por $\Pr(F)V$, onde V é a quantidade de vendedores sendo analisada; já o de vendedores normais é dado por $\Pr(N)V = [1 - \Pr(F)]V$.

Tabela 14: resultados do mecanismo

Percentual médio de vendedores reprovados	$\Pr(V > l) = \Pr(V > l \mid N)\Pr(N) + \Pr(V > l \mid F)\Pr(F)$ $= G_{\hat{s}}(l, \hat{p}_{RN})\Pr(N) + G_{\hat{s}}(l, \hat{p}_{RF})\Pr(F)$
Percentual médio de vendedores normais reprovados	$\Pr(V > l \mid N) = G_{\hat{s}}(l, \hat{p}_{RN})$
Percentual médio de fraudadores reprovados	$\Pr(V > l \mid F) = G_{\hat{s}}(l, \hat{p}_{RF})$
Percentual médio de fraudadores entre os vendedores reprovados	$\Pr(F \mid V > l) = \frac{\Pr(V > l \mid F)\Pr(F)}{\Pr(V > l \mid F)\Pr(F) + \Pr(V > l \mid N)\Pr(N)}$ $= \frac{G_{\hat{s}}(l, \hat{p}_{RF})\Pr(F)}{G_{\hat{s}}(l, \hat{p}_{RF})\Pr(F) + G_{\hat{s}}(l, \hat{p}_{RN})\Pr(N)}$

O exemplo a seguir, com os parâmetros da Tabela 15, ilustra esses resultados. Os parâmetros dos classificadores foram escolhidos de forma a refletir casos extremos, incluindo aqueles em que são reprovados mais normais que fraudadores e em que os índices de reprovação são muito elevados. Além disso, os valores são baixos na média, refletindo a dificuldade da tarefa de distinguir perfis de fraudadores dos de vendedores normais.

Tabela 15: parâmetros do exemplo

Quantidade de classificadores	8
\hat{p}_{RN}	[10% 10% 20% 20% 30% 30% 50% 50%]
\hat{p}_{RF}	[5% 30% 40% 45% 45% 50% 50% 80%]
$\Pr(F)$	1%
P_{RN}^{\max}	10%

Com esses parâmetros, temos os resultados mostrados na Tabela 16. O mecanismo identifica um conjunto contendo 3,08% dos vendedores dos quais 6,30%

são fraudadores, mais de seis vezes a fração do conjunto inicial. Do total de fraudadores existentes no conjunto inicial, cerca de um quinto está no conjunto dos reprovados. Se duplicássemos o conjunto de classificadores, mantendo os mesmos parâmetros, os resultados seriam os mostrados na Tabela 17.

Tabela 16: resultados da aplicação do mecanismo

Limiar de reprovação	5
Percentual médio de vendedores reprovados	3,26%
Percentual médio de vendedores normais reprovados	3,09%
Percentual médio de fraudadores reprovados	20,6%
Percentual médio de fraudadores entre os vendedores reprovados	6,30%

Tabela 17: resultados com o dobro de classificadores

Limiar de reprovação	8
Percentual médio de vendedores reprovados	4,00%
Percentual médio de vendedores normais reprovados	3,67%
Percentual médio de fraudadores reprovados	36,8%
Percentual médio de fraudadores entre os vendedores reprovados	9,20%

5.3.2.

Melhorando o desempenho do mecanismo

Reparando nos percentuais de reprovação médios do exemplo anterior, podemos notar que são bem inferiores ao limite imposto (10%). No segundo caso, se o limiar de reprovação fosse 7, o percentual médio de reprovados seria 11,27%, pouco acima do limite. Neste mesmo caso, o percentual médio de fraudadores reprovados seria de 58,41%, bem melhor que o anterior.

É possível contornar essa ineficiência imposta pela grande variação da probabilidade de reprovação: ao invés de trabalhar sempre com a *maior* probabilidade de reprovação que é *menor* que P_{RN}^{\max} , podemos testar se não seria melhor trabalhar com a *menor* probabilidade de reprovação que é *maior* que P_{RN}^{\max} . Nesse caso, para adequar a resposta ao limite imposto, o mecanismo escolheria aleatoriamente entre os reprovados a quantidade máxima permitida. Essa escolha manteria o percentual de fraudadores presente nos reprovados, mas diminuiria sua quantidade absoluta.

Aplicando essa técnica ao exemplo anterior, a quantidade de fraudadores sobe para 51,83%, com 9,58% de vendedores normais reprovados. A concentração de fraudadores no conjunto de reprovados cai para 5,18%, o que ainda assim é bem superior ao 1% do conjunto original.

Utilizando essa técnica, o cálculo do limiar passa a ser:

$$l = \arg \max_{x \in \mathbb{N} | x \leq n} G_{\hat{s}}(x, \hat{p}_{RF}) \Pr(F) \frac{\min \left[P_{RN}^{\max}, G_{\hat{s}}(x, \hat{p}_{RF}) \Pr(F) + G_{\hat{s}}(x, \hat{p}_{RN}) \Pr(N) \right]}{G_{\hat{s}}(x, \hat{p}_{RF}) \Pr(F) + G_{\hat{s}}(x, \hat{p}_{RN}) \Pr(N)} \quad (5.18)$$

Como $\Pr(F)$ é uma constante positiva, podemos eliminá-lo:

$$l = \arg \max_{x \in \mathbb{N} | x \leq n} G_{\hat{s}}(x, \hat{p}_{RF}) \frac{\min \left[P_{RN}^{\max} G_{\hat{s}}(x, \hat{p}_{RF}) \Pr(F) + G_{\hat{s}}(x, \hat{p}_{RN}) \Pr(N) \right]}{G_{\hat{s}}(x, \hat{p}_{RF}) \Pr(F) + G_{\hat{s}}(x, \hat{p}_{RN}) \Pr(N)} \quad (5.19)$$

O percentual de fraudadores efetivamente reprovado é dado por:

$$\Pr(S^F | F) = G_{\hat{s}}(l, \hat{p}_{RF}) \frac{\min \left[P_{RN}^{\max}, G_{\hat{s}}(l, \hat{p}_{RF}) \Pr(F) + G_{\hat{s}}(l, \hat{p}_{RN}) \Pr(N) \right]}{G_{\hat{s}}(l, \hat{p}_{RF}) \Pr(F) + G_{\hat{s}}(l, \hat{p}_{RN}) \Pr(N)} \quad (5.20)$$

Agora o percentual de vendedores reprovados passa a ser dependente de $\Pr(F)$. A razão é simples: se a quantidade absoluta de fraudadores aumenta, também vai aumentar a quantidade de fraudadores reprovados; como a quantidade de reprovados está sendo limitada, chega o momento em que esta atinge o teto e a proporção de fraudadores reprovados começa a cair.

Esse resultado ainda pode ser melhorado se, ao invés de limitarmos o total de reprovados, impusermos o limite sobre a *total de normais reprovados* e então aplicar o mesmo raciocínio acima: se a quantidade de normais reprovados for maior que o limite, então diminui o tamanho do conjunto de reprovados para adequá-la. Assim sendo, o novo percentual de fraudadores reprovados seria:

$$\Pr(S^F | F) = G_{\hat{s}}(x, \hat{p}_{RF}) \frac{\min \left[P_{RN}^{\max}, G_{\hat{s}}(x, \hat{p}_{RN}) \right]}{G_{\hat{s}}(x, \hat{p}_{RN})} \quad (5.21)$$

O cálculo de l passaria a ser o seguinte:

$$l = \arg \max_{x \in \mathbb{N} | x \leq n} G_{\hat{s}}(x, \hat{p}_{RF}) \frac{\min \left[P_{RN}^{\max}, G_{\hat{s}}(x, \hat{p}_{RN}) \right]}{G_{\hat{s}}(x, \hat{p}_{RN})} \quad (5.22)$$

O percentual de vendedores reprovados deixa de ser constante:

$$\Pr(S^F) = \left[G_{\hat{s}}(l, \hat{p}_{RF}) \Pr(F) + G_{\hat{s}}(l, \hat{p}_{RN}) \Pr(N) \right] \frac{\min \left[P_{RN}^{\max}, G_{\hat{s}}(x, \hat{p}_{RN}) \right]}{G_{\hat{s}}(x, \hat{p}_{RN})} \quad (5.23)$$

É possível melhorar ainda mais o desempenho se, ao invés de escolher aleatoriamente os vendedores a serem excluídos do conjunto de reprovados, escolhermos os que receberam menos votos. No entanto, não vamos calcular aqui esse resultado.

5.3.3. Exemplo

Para ilustrar o funcionamento do mecanismo de classificação de fraudadores, vamos apresentar um exemplo simples com os seguintes parâmetros:

- $P_{RN}^{\max} = 5\%$
- $\hat{p}_{RF} = (p_{RF_1}, \dots, p_{RF_n}), p_{RF_i} \sim \text{Beta}(9;11)$, com média de 45%.
- $\hat{p}_{RN} = (p_{RN_1}, \dots, p_{RN_n}), p_{RN_i} \sim \text{Beta}(2;8)$, com média de 20%

As distribuições de probabilidade dos parâmetros dos classificadores estão mostradas graficamente na Figura 13.

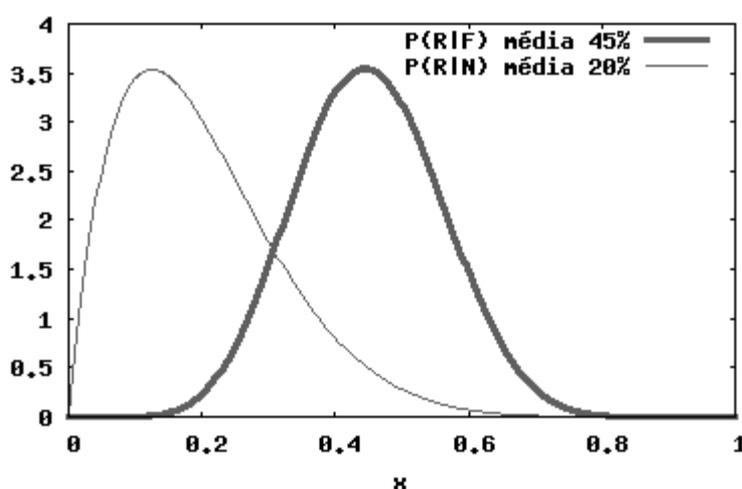


Figura 13: distribuições de probabilidade dos parâmetros dos classificadores

Na Figura 14 vemos o valor médio que $\Pr(S^F | F)$ assume conforme aumenta o número de classificadores disponíveis. Esse valor médio é calculado através da equação (5.21). Como esse resultado varia conforme a amostra de classificadores utilizada, para cada ponto simulamos 200 amostras de classificadores com base nas distribuições dos parâmetros. O resultado final mostrado é a média aritmética dos 200 resultados obtidos com a aplicação do mecanismo com cada um desses conjuntos. Também vemos a evolução do valor do limiar de reprovação (indicado acima dos pontos do gráfico). Esse limiar é calculado usando a equação (5.22). Na ausência de um número, subentende-se que o valor de l é igual ao do ponto anterior.

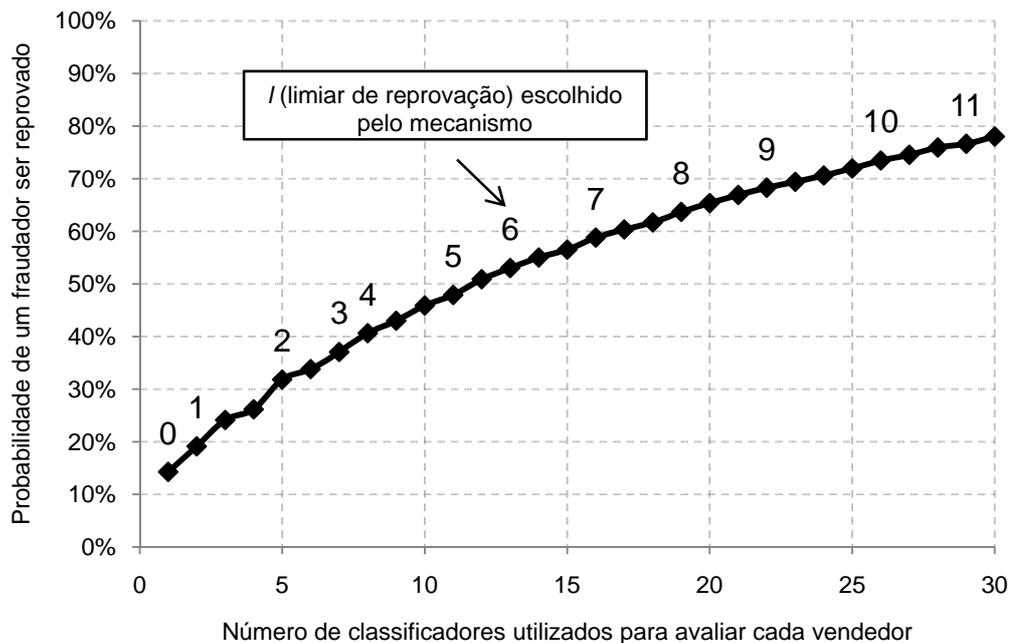


Figura 14: desempenho do mecanismo vs número de classificadores

5.4. Mecanismo de incentivo aos classificadores

Para funcionar, o mecanismo de classificação proposto depende da qualidade das classificações feitas e também da quantidade de classificadores, que permite compensar um menor desempenho destes a partir da agregação de múltiplas opiniões. Na seção anterior, assumimos a existência de populações de classificadores com determinadas características; também assumimos que o desempenho dos classificadores não varia muito ao longo do tempo.

Contudo, essas premissas dificilmente se sustentam num cenário de participação voluntária de agentes humanos. As pessoas precisam de algum tipo de motivação para participar. Logo, faz falta um *mecanismo de incentivo*, que torne a atuação como classificador proveitosa para eles e que atraia um bom número de agentes humanos querendo participar. A qualidade de uma resposta depende do esforço que o classificador faz para analisar o vendedor em questão. Assim, é também necessário que haja incentivo para exercer o maior esforço possível.

Vamos delinear neste capítulo um mecanismo de incentivo através do qual os bons classificadores recebem do operador algum tipo de benefício: remuneração, descontos em produtos, prêmios etc. Esse benefício traz um custo extra para o operador, que deve ser levado em conta.

Vamos inicialmente analisar o comportamento estratégico dos classificadores, visando identificar quais as restrições que o mecanismo de incentivo deve obedecer para que atinja seu objetivo. Sendo o desempenho dos classificadores crucial para o funcionamento do mecanismo de classificação de vendedores, analisaremos alternativas para medir esse desempenho. Com os elementos anteriormente estudados, será possível então propor um mecanismo de remuneração vinculado ao desempenho e que coíba atuações indesejáveis dos classificadores. Por fim, analisaremos o impacto do mecanismo de remuneração no orçamento do operador e mostraremos um exemplo.

5.4.1. Comportamentos estratégicos

Nesta seção vamos analisar as interações entre os classificadores e o operador como jogos nos quais tanto uns quanto outros atuam como *agentes racionais*, no sentido de que buscam maximizar a utilidade obtida a partir das classificações feitas. Aqui entendemos utilidade como o benefício com o qual o operador retribui o classificador pelo seu trabalho: pode ser dinheiro, prêmios, descontos etc. No caso do operador, essa utilidade pode ser dada pelo quanto ele deixa de perder ao utilizar as informações fornecidas pelos classificadores.

As análises serão feitas de modo qualitativo, visando derivar políticas do operador para maximizar a *sua* utilidade. Uma extensão natural do trabalho seria efetuar-las quantitativamente, de modo a identificar a existência de equilíbrios favoráveis ao operador, na linha da teoria de projeto de mecanismos.

5.4.1.1. Atuação dos classificadores

Os classificadores dispõem de duas ações básicas:

- *Cadastrar-se*, adquirindo uma identidade única frente ao mecanismo.
- *Classificar* um conjunto de vendedores, dizendo para cada um deles se o considera suspeito ou não de ser fraudador.

A primeira ação é pré-requisito para efetuar a segunda; esta pode ser repetida um número arbitrário de vezes. Tanto uma quanto outra podem ter custos associados.

O operador espera que o classificador faça esforço e responda depois de analisar os vendedores. Contudo, não há nada que vincule *a priori* o classificador a fazer isso: sendo um agente racional, ele decidirá em função da utilidade que obtém com cada resposta.

Podemos identificar três tipos de resposta, juntamente com a utilidade que o agente pode derivar de cada uma delas:

- *Resposta com esforço*: essa é aquela obtida após analisar os vendedores. Há um custo associado a esse esforço e quanto maior o esforço, maior o custo e também maior a taxa de acerto nas respostas. Todavia, é razoável supor que o custo é uma função que cresce mais rapidamente conforme aumenta a precisão. Também é razoável supor que há um limite na precisão que pode ser atingida, dado que o classificador conta com informações e tempo limitados para responder.
- *Resposta aleatória*: nesse caso, o classificador não exerce esforço e responde aleatoriamente. É o caso particular da anterior quando é escolhida uma resposta com custo zero.
- *Resposta manipulada*: é aquela na qual as respostas são dadas em função de algum benefício externo ao sistema. Nesse caso, podemos supor que o classificador não está interessado na remuneração que o sistema poderia lhe dar: sua utilidade depende exclusivamente de que o sistema atribua determinados veredictos a determinados vendedores.

O classificador pode escolher que tipo de resposta dar a cada vendedor analisado: ele pode manipular a resposta de uns, responder aleatoriamente outros e outros ainda com esforço.

Uma vez dada a resposta ao mecanismo, este anuncia o resultado da ação (*outcome*), que consiste na utilidade que o mecanismo transfere do operador para cada um dos classificadores. Essa utilidade é uma função das respostas e da atuação observada dos vendedores classificados. No caso do operador, essa utilidade é dada pelo quanto ele deixou de perder ao utilizar as respostas do conjunto de classificadores. Teoricamente poderia haver transferência de utilidade dos classificadores para o operador. No entanto, parece-nos uma condição pouco viável para

um sistema de computação humana distribuído exigir que os classificadores paguem ao mecanismo. Logo, assumimos que não há transferências nesse sentido.

Quanto à decisão do classificador de participar ou não do jogo, utilizamos o critério de *participação voluntária*: ele só participa se o valor esperado da utilidade (resultado menos custos de participação) for maior ou igual ao que ele ganharia não participando. Dito de outro modo: o classificador só participa se houver *possibilidade* de lucrar algo, dado que ele não sabe ainda qual a precisão esperada das suas respostas e, com isso, o quanto pode ganhar do operador. A cada rodada o classificador obtém mais informação acerca da sua precisão, o que pode fazer com que deixe de participar nas rodadas seguintes.

Vamos assumir que os classificadores não conseguem interferir no comportamento do vendedor, isto é, não conseguem fazer com que um vendedor analisado exiba uma atuação fraudulenta quando na verdade não era um fraudador e vice-versa.

5.4.1.2. Atuação do operador

O operador só tem uma decisão a tomar: se participa ou não do jogo. Ou seja: ele decide em função do valor esperado das variáveis que afetam a sua utilidade (precisão dos classificadores, perda com fraudes e gasto com a remuneração dos classificadores).

5.4.1.3. Tornando o mecanismo compatível com incentivo

Um dos objetivos básicos da teoria de projeto de mecanismos é o de encontrar as condições para que um mecanismo de revelação direta seja *compatível com incentivo*, isto é, que todos os agentes participantes tenham como melhor estratégia revelar de modo veraz suas informações privadas.

Para tornar o mecanismo proposto compatível com incentivo, um primeiro passo é o de remunerar cada classificador apenas em função do seu desempenho individual, sem levar em conta as ações dos outros classificadores. Desse modo, não haverá incentivo para um classificador influenciar as ações dos outros com relação à utilidade que pode ser obtida do operador.

No entanto, as ações dos outros classificadores podem influenciar a utilidade individual caso esta também dependa de algum benefício externo que, por sua

vez, dependa do veredicto acerca dos vendedores analisados. Um classificador nessa situação teria incentivo para tentar fazer com que outros classificadores dêem opiniões que lhe ajudem.

Ainda assim, podemos continuar assumindo que os classificadores não terão incentivos para influenciarem-se, já que os benefícios advindos de tal estratégia teriam que ser divididos entre eles, sendo que poderiam obter o mesmo resultado através do “ataque de sibila” (Douceur, 2002), que consiste na obtenção de um número arbitrário de identidades como classificador.

Para completar, a remuneração do classificador deve ser uma função monótona crescente do seu desempenho. Também deve levar em conta as funções de custo dos classificadores, já que estes só aceitarão participar se a remuneração superar determinado limite. Com essas restrições, garante-se que estejam alinhados o interesse do classificador (maximizar sua utilidade) com o do operador (maximizar a descoberta de fraudadores).

Um segundo passo para tornar o mecanismo compatível com incentivo é o de contrabalançar a utilidade que os agentes possam obter a partir da manipulação dos veredictos do mecanismo. Isso pode ser feito adotando algumas medidas:

- Impor um custo de entrada suficientemente elevado. Isso significa que a ação de cadastrar-se deve ter algum tipo de custo associado que não implique transferências do classificador para o operador. Um exemplo disso é a exigência de que o classificador analise certo número de vendedores sem ser remunerado por isso.
- Excluir os classificadores com um desempenho abaixo de determinado limite, o que implica para o agente humano ter que arcar novamente com o custo de cadastramento.
- Limitar o tamanho do conjunto de vendedores a ser analisado por unidade de tempo. Essa medida limita a quantidade de opiniões manipuladas que podem ser dadas antes do mecanismo perceber, pela atuação dos vendedores anteriormente classificados, que o limite inferior de desempenho foi violado.
- Distribuir os vendedores aleatoriamente entre os classificadores e serializar a avaliação dos vendedores, só permitindo que um classificador avalie um novo vendedor quando tiver finalizado a análise do

anterior. Com isso, torna-se custoso “escolher” que vendedores avaliar. Para que de fato isso ocorra, também é necessário que haja uma quantidade de classificadores suficientemente grande, de modo que seja pouco provável a coincidência de um grupo de classificadores avaliarem com frequência os mesmos vendedores. Note-se que essa medida também remove o incentivo para simular fraudes, já que será difícil conseguir classificar a identidade que perpetrou a fraude simulada.

Com isso, o sistema associa à tarefa de manipular cada resultado certo custo, já que o classificador malicioso terá que reentrar no sistema várias vezes e com várias identidades para conseguir influenciar algum resultado. Uma escolha adequada dos valores associados às restrições acima pode remover o incentivo à manipulação de resultados. O cálculo desse custo mínimo envolve as restrições acima mais o benefício advindo da manipulação de um resultado: um fraudador pode, por exemplo, pagar alguém para tentar “aprová-lo”. Ou então um vendedor pode pagar para difamar determinados concorrentes. Esse pagamento seria o benefício advindo da manipulação. Por fim, também depende da probabilidade de encontrar o vendedor cujo veredicto se deseja manipular, o que depende da quantidade de vendedores, de classificadores e de quantas opiniões o mecanismo solicita para cada vendedor.

Um modo de impor um custo inicial que também seja útil para o mecanismo é o de exigir que o candidato a classificador faça um teste: avalie alguns vendedores para que o sistema tenha alguma idéia do seu desempenho. Para acelerar esse teste, ao invés de seguir o processo normal de classificação (aguardar certo tempo para verificar se o vendedor cometeu ou não fraude), o mecanismo pode apresentar ao novo usuário vendedores que estejam sob avaliação. Caso a resposta do usuário coincida com o veredicto do mecanismo, então o mecanismo considera que ele “acertou” o resultado. O mecanismo pode então exigir certo número de acertos para deixar o novo classificador entrar.

Como o que interessa para o mecanismo é a habilidade do classificador em descobrir fraudadores, esse teste inicial pode ser feito aproveitando o próprio conjunto de vendedores que estão próximos de ser reprovados. Assim garante-se que haja uma proporção significativa de fraudadores nesse teste. Outra solução ainda é

a de comparar a resposta do candidato não com o veredicto do sistema, mas com a resposta dos melhores classificadores.

Tendo essas restrições em mente, um trabalho futuro seria o de calcular quais deveriam ser os valores associados às restrições acima (custo de entrada, custo de classificar, precisão mínima etc.) para levar o sistema a um único equilíbrio no qual a ação mais proveitosa para os classificadores seria a de esforçar-se na análise dos vendedores e dar respostas verazes.

5.4.2. Avaliando classificadores

O próximo passo para a definição do mecanismo de incentivo é a escolha de uma métrica para avaliar os classificadores. Conforme vimos na seção anterior, interessa que a função de remuneração acompanhe o desempenho deste. Um primeiro passo então passa a ser o de modelar esse desempenho.

O interesse do operador é o de que seja reprovada a maior quantidade possível de fraudadores mantendo a reprovação de vendedores normais dentro do limite, o que na prática significa que os classificadores têm que procurar minimizar a reprovação de normais. Tendo isso em vista, vamos calcular qual o resultado da atuação do classificador ao avaliar um conjunto de vendedores, utilizando as variáveis na Tabela 18 (relacionadas ao desempenho de um classificador).

Tabela 18: variáveis utilizadas para avaliar um classificador

Quantidade de vendedores sendo analisada	V
Quantidade de fraudadores	$V_F = P_F V$
Quantidade de vendedores normais	$V_N = (1 - P_F) V$
Quantidade de vendedores normais reprovados	V_{RN}
Quantidade de fraudadores reprovados	V_{RF}
Quantidade de vendedores reprovados	$V_R = V_{RN} + V_{RF}$

O comportamento do classificador “melhora” quando ele reprova mais fraudadores e quando reprova menos normais. Assim, podemos esboçar uma primeira fórmula para calcular o desempenho do classificador de forma a refletir essas melhoras:

$$V_{RF} (V_N - V_{RN})$$

No entanto, essa fórmula depende da quantidade de fraudadores presente na amostra: quanto menos fraudadores, menos valores V_{RF} pode assumir. Para remover essa dependência, podemos trabalhar com as *proporções* de normais e fraudadores na amostra:

$$\frac{V_{RF}}{V_F} \left(1 - \frac{V_{RN}}{V_N} \right)$$

Outra característica importante é que o desempenho do classificador só é positivo quando se dá a seguinte condição:

$$\frac{V_{RF}}{V_F} > \frac{V_{RN}}{V_N}$$

Caso contrário, o desempenho do classificador seria aleatório ou pior. Para embutir isso na fórmula, podemos ajustá-la, de modo que seja maior que zero quando a condição acima é respeitada:

$$\frac{V_{RF}}{V_F} \left(1 - \frac{V_{RN}}{V_N} \right) - \frac{V_{RN}}{V_N} \left(1 - \frac{V_{RN}}{V_N} \right)$$

Simplificando:

$$\left(\frac{V_{RF}}{V_F} - \frac{V_{RN}}{V_N} \right) \left(1 - \frac{V_{RN}}{V_N} \right)$$

Por conta da modelagem probabilística adotada, se a quantidade de vendedores avaliada for suficientemente grande, podemos substituir as frações pelos parâmetros do classificador e escrever o desempenho de um classificador i com a seguinte fórmula:

$$D_i = (p_{RF_i} - p_{RN_i}) (1 - p_{RN_i})$$

5.4.3. Remunerando os classificadores

Agora que temos um modo de calcular o desempenho de cada classificador, podemos definir uma função de remuneração que atenda às restrições apresentadas na seção 5.4.1.3. A solução mais simples seria remunerar proporcionalmente ao desempenho:

$$R_i = r_{\max} D_i \quad (5.24)$$

Desse modo, a remuneração vai de zero até r_{\max} , respeitando a condição de não haver remuneração “negativa”, isto é, pagamentos do classificador para o ope-

rador. No entanto, é necessário levar em conta o mínimo que o classificador espera receber para participar. Uma maneira de fazer isso é:

$$R_i = r_{\max} D_i + r_{\min} (1 - D_i) \quad (5.25)$$

Quanto à r_{\max} , o operador deve determiná-lo de modo a motivar os classificadores a fazerem mais esforço, já que se o acréscimo de remuneração por esforço for muito pequeno, não haverá incentivo para exercer trabalho adicional além do mínimo. Também deve ser levado em conta o desempenho esperado dos *melhores* classificadores da população, já que dificilmente estes se aproximam do máximo teórico e, no entanto, são os que serão mais bem remunerados. O operador então pode calcular r_{\max} a partir daí. Seja D_{\max} o desempenho médio dos melhores classificadores (e.g. dos 10% melhores); podemos então reescrever a equação anterior em função disso:

$$R_i = \frac{r_{\max} - r_{\min}}{D_{\max}} D_i + r_{\min} \quad (5.26)$$

Nesse caso, teremos um novo máximo teórico:

$$r'_{\max} = \frac{r_{\max} - r_{\min} (1 - D_{\max})}{D_{\max}} \quad (5.27)$$

Podemos dar uma definição mais precisa de D_{\max} ; por exemplo, podemos utilizar o valor esperado da maior estatística de ordem, isto é, do maior elemento de uma amostra de classificadores:

$$D_{\max} = E[D_{(N_r)}] \quad (5.28)$$

Temos então que:

$$R_i = \frac{r_{\max} - r_{\min}}{E[D_{(N_r)}]} D_i + r_{\min} \quad (5.29)$$

Por outro lado, o operador pode especificar um desempenho mínimo para que seja efetuado o pagamento:

$$R_i = \frac{r_{\max} - r_{\min}}{E[D_{(N_r)}] - D_{\min}} (D_i - D_{\min}) + r_{\min} \quad (5.30)$$

Ao contrário do caso do valor máximo, o valor de D_{\min} não será calculado a partir da distribuição: será um parâmetro a mais que o mecanismo disponibiliza para controlar a remuneração.

Agora podemos calcular o valor esperado da remuneração de um classificador escolhido aleatoriamente dentro da população de classificadores:

$$E(R) = E \left\{ \frac{r_{\max} - r_{\min}}{E[D_{(N_r)}] - D_{\min}} [E(D - D_{\min})] + r_{\min} \right\} \quad (5.31)$$

$$= \frac{r_{\max} - r_{\min}}{E[D_{(N_r)}] - D_{\min}} [E(D) - D_{\min}] + r_{\min}$$

Definida a remuneração do classificador, podemos agora analisar o custo necessário para incentivar os classificadores. Seja V a quantidade total de vendedores a serem analisados, n_v a quantidade de classificadores que analisam cada vendedor. O valor esperado do custo total do operador pode ser descrito com a seguinte fórmula:

$$C = Vn_v E(R)$$

$$= Vn_v \left\{ \frac{r_{\max} - r_{\min}}{E[D_{(N_r)}] - D_{\min}} [E(D) - D_{\min}] + r_{\min} \right\} \quad (5.32)$$

$$= Vn_v \left\{ \frac{r_{\max} - r_{\min}}{E[D_{(N_r)}] - D_{\min}} \left\{ E[(p_{RF} - p_{RN})(1 - p_{RN})] - D_{\min} \right\} + r_{\min} \right\}$$

Para exemplificar, dados os parâmetros mostrados na Tabela 19, temos que o valor esperado do custo total para analisar os vendedores seria de R\$ 7.724,00: uma média de R\$ 7,72 gastos por vendedor. O valor médio pago a cada classificador por vendedor analisado seria de R\$ 0,39. Se os vendedores forem distribuídos igualmente entre todos os classificadores, cada um avaliará 200 vendedores e o valor esperado do seu pagamento será de R\$ 77,24. O melhor classificador receberia em média R\$ 200,00 por essa tarefa.

Tabela 19: exemplos de parâmetros do mecanismo

V	1000 vendedores
N_r	100 classificadores disponíveis
n_v	20 classificadores analisando cada vendedor
r_{\max}	R\$ 1,00
r_{\min}	R\$ 0,01
População de classificadores	É tal que os parâmetros de um classificador escolhido aleatoriamente são dados pelas seguintes variáveis aleatórias: $p_{RF_i} \sim \text{Beta}(33 \frac{3}{7}; 10)$, média = 65% $p_{RN_i} \sim \text{Beta}(5 \frac{3}{5}; 12 \frac{3}{5})$, média = 30%
$E[(p_{RF} - p_{RN})(1 - p_{RN})]$	0,26 (calculado a partir da população acima)
$E[D_{\max}] = E[D_{(N_r)}]$	0,57 (idem)
$E(D_{\min})$	0,07 (equivalente a $p_{RF} = 40\%$ e $p_{RN} = 30\%$)

Com a equação (5.32), o mecanismo pode adequar-se ao orçamento fornecido pelo operador (mensagem “orçamento máximo”) ajustando os parâmetros n_v , D_{\min} , r_{\min} e r_{\max} .

5.4.4. Equilibrando o orçamento

O operador deve equilibrar o custo advindo com o uso do jogo com a redução de perdas ao descobrir mais rapidamente quem são os fraudadores. Podemos incluir esse balanceamento na própria operação do mecanismo. Para isso, temos que quantificar o benefício que a reprovação do fraudador traz para o operador. Vamos desenvolver um modelo simplificado do impacto da atuação fraudulenta e então incluiremos o impacto da reprovação de fraudadores e do custo de remuneração dos classificadores.

O primeiro elemento do modelo é a perda média causada por um fraudador atuando livremente (isto é, sem que o operador esteja usando o mecanismo). Chamaremos esse parâmetro de L_F (“loss”). É o valor esperado dos pagamentos que o operador deixa de receber por conta da atuação de um fraudador (cfr. seção 3.4).

V é a quantidade de vendedores por unidade de tempo (e.g. por dia) que passam a ter potencial de fraude. P_F é percentual de fraudadores que iniciam ou reiniciam suas atividades a cada unidade de tempo, valor obtido pelos dados históricos. Com esses parâmetros, é possível estimar a perda total causada pelos fraudadores *por unidade de tempo*:

$$L = L_F P_F V \quad (5.33)$$

O mecanismo proposto permite identificar um subconjunto de vendedores que alberga, em média, $P_F V \Pr(S^F | F)$ fraudadores. Se aplicarmos essa idéia ao conjunto de vendedores que iniciam sua atividade num determinado momento, tempo que:

$$V_{RF} = \Pr(S^F | F) P_F V \quad (5.34)$$

Se aplicarmos a idéia de medidas coercitivas que estabeleçam limites máximos no prejuízo que cada fraudador pode causar (cfr. seção 5.1), podemos calcular a perda levando em conta a reprovação de fraudadores, assumindo que o prejuízo

ízo causado por um fraudador reprovado, L_{RF} , é menor que o causado por um fraudador atuando livremente:

$$\begin{aligned} L' &= [P_F V - V_{RF}] L_F + V_{RF} L_{RF} + C \\ L' &= [1 - \Pr(S^F | F)] L_F P_F V + \Pr(S^F | F) L_{RF} P_F V + C \end{aligned} \quad (5.35)$$

O parâmetro L_{RF} é o prejuízo que é causado por um fraudador que foi reprovado pelo mecanismo e sofreu restrições a sua atuação; C é o custo de operar o mecanismo de incentivo, dado pela equação (5.32). Podemos agora calcular quanto é recuperado pelo mecanismo com relação à perda original do operador: $\frac{L-L'}{L}$. Para diminuir o tamanho da equação, vamos substituir $\Pr(S^F | F)$ por \bar{P}_{RF} e C por Vc , onde $c = n_v E(R)$ é o custo para avaliar um vendedor:

$$\begin{aligned} \frac{L-L'}{L} &= \frac{L_F P_F V - (1 - \bar{P}_{RF}) L_F P_F V - \bar{P}_{RF} L_{RF} P_F V - V n_v E(R)}{L_F P_F V} \\ &= \frac{L_F - (1 - \bar{P}_{RF}) L_F - \bar{P}_{RF} L_{RF} - \frac{n_v E(R)}{P_F}}{L_F} \\ &= 1 - (1 - \bar{P}_{RF}) - \bar{P}_{RF} \frac{L_{RF}}{L_F} - \frac{n_v E(R)}{L_F P_F} \\ &= \bar{P}_{RF} \left(1 - \frac{L_{RF}}{L_F} \right) - \frac{n_v E(R)}{L_F P_F} \end{aligned} \quad (5.36)$$

Vemos que o valor recuperado depende de duas razões relacionadas com o prejuízo que um fraudador causa individualmente. Como as variáveis acima não são todas independentes, para ficar mais claro podemos reescrever alguns parâmetros como funções de outros:

$$\frac{L-L'}{L} = \bar{P}_{RF} (n_v, \hat{p}_{FN}) \left(1 - \frac{L_{RF}}{L_F} \right) - \frac{n_v E[R(\hat{p}_{FN})]}{L_F P_F} \quad (5.37)$$

Onde $\hat{p}_{FN} = (\hat{p}_{RF}, \hat{p}_{RN})$ é o conjunto de parâmetros da população de classificadores. As variáveis relacionadas ao funcionamento do mecanismo (n_v e \hat{p}_{FN}) aparecem nas duas parcelas, indicando que há um problema subjacente de otimização. Podemos encontrar os parâmetros que maximizam o *valor recuperado* (lado esquerdo da equação): essa seria a otimização mais óbvia. No entanto, o operador pode estar mais preocupado com os efeitos indiretos da presença dos fraudadores. Nesse caso, podemos maximizar a reprovação de fraudadores (\bar{P}_{RF}), seja

num cenário de equilíbrio (recuperação zero), seja num cenário de recuperação limitada (total gasto dentro de certo limite). No primeiro cenário, temos que maximizar \bar{P}_{RF} sujeitos à seguinte condição:

$$\bar{P}_{RF}(n_v, \hat{p}_{FN}) \left(1 - \frac{L_{RF}}{L_F} \right) - \frac{n_v E[R(\hat{p}_{FN})]}{L_F P_F} \geq 0 \quad (5.38)$$

No segundo cenário, temos que voltar a equação e maximizar \bar{P}_{RF} sujeitos a essa outra condição:

$$\left[1 - \bar{P}_{RF}(n_v, \hat{p}_{NF}) \right] L_F P_F V + \bar{P}_{RF}(n_v, \hat{p}_{NF}) L_{RF} P_F V + V n_v E[R(\hat{p}_{NF})] \leq L_{\max} \quad (5.39)$$

Simplificando:

$$n_v E[R(\hat{p}_{NF})] - L_{RF} P_F \bar{P}_{RF}(n_v, \hat{p}_{NF}) \leq \frac{L_{\max}}{V} \quad (5.40)$$

Note-se que o fator que imediatamente impacta \bar{P}_{RF} é o aumento de n_v , e o que importa com relação ao limite de orçamento é o quanto o operador está disposto a gastar por vendedor.

No processo de equilibrar o orçamento, o mecanismo de incentivo obterá um valor ótimo de n_v , valor esse que é usado pelo mecanismo de classificação. Ou seja, a conexão entre os dois mecanismos se dá através desse parâmetro.

5.4.5. Exemplo

Para ilustrar os resultados, vamos simular o funcionamento do mecanismo de identificação de fraudadores, que consiste na operação conjunta do mecanismo de classificação de vendedores com o mecanismo de incentivo aos classificadores. A simulação é necessária porque os resultados variam dependendo do conjunto de classificadores escolhido. Por conta disso, dada uma população de classificadores e dado um valor de n_v , vamos aplicar o mecanismo diversas vezes, escolhendo aleatoriamente subconjuntos da população com n_v classificadores. No final, calcularemos a média aritmética dos resultados obtidos, que será uma aproximação do valor esperado do resultado do mecanismo dada a população em questão e dado n_v . Deter-nos-emos em dois resultados: o valor recuperado – $\frac{L-L'}{L}$ – e a fração de fraudadores reprovados – $\Pr(S^F | F)$.

Vamos aproveitar os parâmetros do exemplo anterior (cfr. Tabela 19), acrescidos dos parâmetros mostrados na Tabela 20.

Tabela 20: parâmetros adicionais do mec. de identificação de fraudadores

P_F	5% de fraudadores
L_F	R\$ 300,00
L_{RF}	20% de $L_F =$ R\$ 60,00
P_{RN}^{\max}	5% de tolerância

A Figura 15 mostra como varia a média do valor recuperado em função do número de classificadores empregado para avaliar cada vendedor. Utilizamos 200 iterações para calcular a média; escolhemos esse número empiricamente, observando a variação dos resultados. O percentual recuperado máximo é de 31% da perda original, com 12 classificadores avaliando cada vendedor e com uma taxa de reprovação de fraudadores de 75%. Aumentando o número de classificadores, a recuperação decresce e vai a praticamente zero com 31 classificadores avaliando cada vendedor. Nessa situação, o mecanismo consegue reprovar, em média, 99% dos fraudadores.

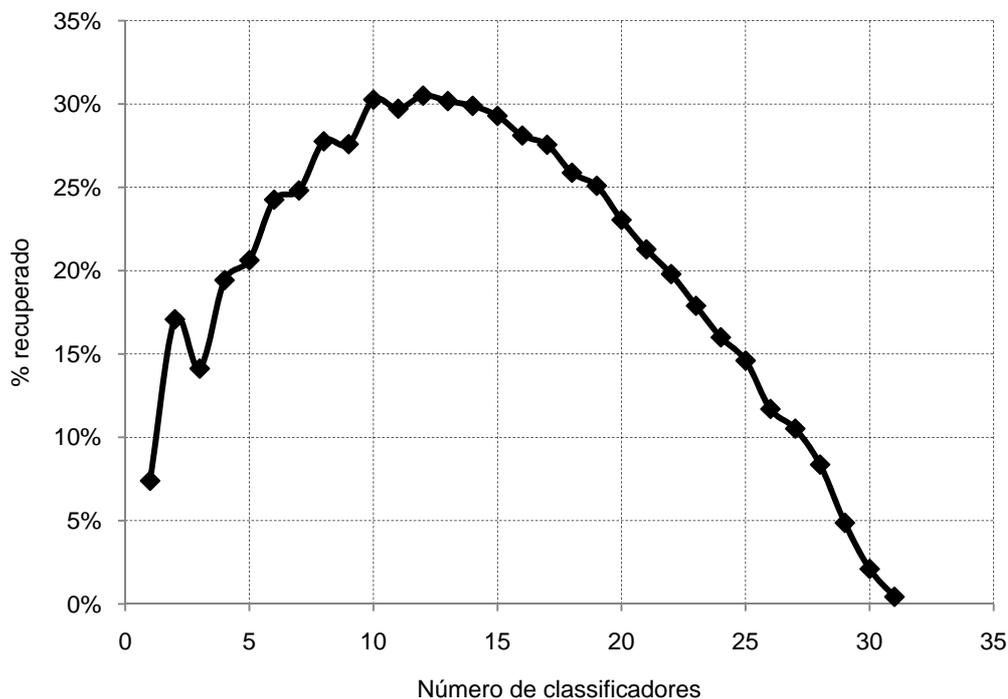


Figura 15: valor recuperado médio vs número de classificadores

5.5.

Avaliação do mecanismo de identificação de fraudadores

Nesta seção vamos verificar a viabilidade do mecanismo de identificação de fraudadores. Para isso, simularemos o funcionamento do mesmo com diversos parâmetros e argumentaremos que os resultados obtidos são relevantes.

Nas próximas seções vamos descrever:

- Os critérios usados para escolher os valores dos parâmetros.
- Quais os resultados do funcionamento do mecanismo que serão analisados.
- Os resultados da operação do mecanismo (obtidos através de simulação).

5.5.1.

Estimando parâmetros de funcionamento do mecanismo

Nesta seção, vamos indicar como foram fixados os valores dos parâmetros, tendo em conta que desejamos avaliar o mecanismo em cenários que se aproximem da realidade.

Prejuízo médio causado por um fraudador atuando normalmente – L_F . Vamos utilizar para esse parâmetro o valor de R\$ 157,00, que aparece na pesquisa exploratória (cfr. seção 3.4) como sendo o prejuízo médio que os fraudadores encontrados causaram.

Redução de prejuízo com a reprovação de um fraudador – r_F . Esse parâmetro não é mais que $\frac{L_{RF}}{L_F}$, ou seja, a razão entre a perda causada por um fraudador que sofreu restrições com a que é causada por um fraudador que atua livremente. Não temos subsídios para escolher esses valores. Vamos utilizar dois valores: um otimista – 20% – e outro pessimista – 40%.

Percentual médio de fraudadores entre os vendedores selecionados – P_F . Esse percentual médio depende dos dados históricos e do critério de escolha dos vendedores a serem analisados pelos classificadores. Na seleção de perfis de vendedores no teste piloto, conseguimos um percentual de fraudadores de pouco mais de 10% (cfr. seção 4.3.1). Vamos utilizar esse valor como uma estimativa otimista e a sua metade – 5% – como uma estimativa pessimista. Como o operador do sistema possui um conhecimento profundo dos perfis dos vendedores, esperamos

que ele de fato consiga selecionar conjuntos de vendedores com essas porcentagens de fraudadores (na média).

Pagamentos a serem efetuados aos classificadores – r_{\min} , r_{\max} . A escolha desses parâmetros deve ser feita levando em conta que (i) se forem muito baixos ou muito pouco diferenciados, não atrairão classificadores o suficiente ou só atrairão maus classificadores, (ii) se forem muito elevados, acabarão comprometendo a diminuição do prejuízo direto com fraudes. Balizamos a escolha dos parâmetros verificando os pagamentos feitos no sítio Mechanical Turk: procuramos tarefas similares em termos de tempo máximo de execução e utilizamos os valores mínimos e máximos pagos por essas tarefas. Visitamos o sítio no dia 23/08/2008 e verificamos as tarefas cujo tempo de execução máximo era menor que 10 minutos, bem próximo do tempo máximo gasto pelos classificadores no teste piloto (12 minutos). Verificamos que o menor pagamento era de US\$ 0.01 e o maior de US\$ 1.00. Utilizaremos os mesmos valores em todos os cenários, só que em moeda local: pagamentos entre R\$ 0,01 e R\$ 1,00.

Limite inferior da qualidade do classificador – D_{\min} . Esse limite deve ser escolhido em função da população de classificadores, já que quanto maior, menos classificadores recebem pagamento. Vamos utilizar o pior caso: $D_{\min} = 0$, o que significa que todos os classificadores com desempenho positivo, i.e. $D_i > 0$, são remunerados.

Desempenho máximo esperado – D_{\max} . Vamos utilizar o desempenho do melhor classificador que participou no teste piloto ($p_{RF} = 1$ e $p_{RN} = 0,15$), com o qual temos $D_{\max} = 0,72$.

Tolerância com a reprovação de vendedores normais – P_{RN}^{\max} . Esse parâmetro está relacionado com as regras de negócio do operador que, de acordo com suas estratégias, avalia quão rigoroso deve ser no processo de descoberta de fraudadores. Na pesquisa exploratória feita com o MercadoLivre (cfr. Capítulo 3), reparamos que 1,1% dos vendedores que analisamos foram suspensos e depois reabilitados. Vamos utilizar o valor de 1% como uma estimativa pessimista de P_{RN}^{\max} e o seu dobro – 2% – como uma estimativa otimista.

Parâmetros dos classificadores – \hat{p}_{RF} , \hat{p}_{RN} . Vamos utilizar o desempenho das pessoas que participaram do teste piloto (cfr. Tabela 8, colunas FCF e NCF)

como parâmetros do mecanismo. Mais especificamente, os parâmetros dos classificadores serão escolhidos através de amostragem aleatória com repetição dentre os pares que aparecem na Tabela 8, com $p_{RF} = FCF$ e $p_{RN} = NCF$.

5.5.2.

Medindo o resultado da operação do mecanismo

A saída do mecanismo de identificação de fraudadores consiste em uma classificação dos vendedores em aprovados ou reprovados e em uma lista de quanto deve ser pago a cada classificador. Utilizaremos as métricas a seguir para avaliar a saída do mecanismo.

Percentual médio de fraudadores reprovados – $\Pr(S^F | F)$. É o valor esperado da fração de fraudadores que são efetivamente reprovados pelo mecanismo. Calculado através da equação (5.21).

Percentual da perda original que é recuperado – $L-L'/L$. É o valor esperado da fração da perda que o operador deixa de ter com a atuação dos fraudadores por conta da reprovação de fraudadores e a consequente restrição que sofrem. É calculado através da equação (5.37).

Pagamento médio aos classificadores por vendedor analisado – $E(R)$. É o valor médio dos pagamentos calculados pelo mecanismo para cada vendedor avaliado, dada uma população de classificadores. É calculado através da equação (5.31).

5.5.3.

Simulando a operação do mecanismo

Mostraremos agora o funcionamento do mecanismo com diferentes conjuntos de parâmetros. Vamos variar os três parâmetros para os quais escolhemos dois valores, um otimista e um pessimista (cfr. seção 5.5.1). Os demais permanecerão fixos com os valores indicados na mesma seção.

Os valores dos resultados do mecanismo que serão apresentados foram obtidos por simulação, do mesmo modo que nos exemplos mostrados nas seções 5.3.3 e 5.4.5. A única diferença será o número de iterações, que aumentamos para 10^4 , de forma a obtermos resultados mais precisos, com desvio padrão menor que 10^{-5} .

Na Tabela 21 mostramos os resultados do mecanismo para as diversas combinações de valores dos parâmetros variáveis. A simulação foi feita com a finalidade de maximizar a *recuperação de perdas*. Ou seja, os resultados apresentados referem-se ao uso do mecanismo com um número ótimo de classificadores avaliando cada vendedor, calculado pelo mecanismo de incentivo usando a equação (5.38). Esse número é mostrado na última linha da tabela.

Tabela 21: resultados da simulação

Percentual médio de fraudadores	5%				10%			
	40%		20%		40%		20%	
Redução do prejuízo								
Tolerância	1%	2%	1%	2%	1%	2%	1%	2%
% médio de fraudadores reprovados	13%	16%	20%	45%	64%	72%	76%	79%
% da perda original recuperado	2%	4%	5%	9%	11%	16%	25%	30%
Pagamento médio aos classificadores por vendedor (em R\$)	0,43							
Número ótimo de classificadores avaliando cada vendedor	1	1	2	5	10	10	13	12

Os valores dos parâmetros foram ordenados de forma a realçar o impacto que cada um deles tem sobre a reprovação de fraudadores e sobre o valor recuperado. O parâmetro mais importante é o percentual de fraudadores entre os vendedores analisados, seguido pela redução percentual do prejuízo por conta das restrições aos fraudadores reprovados. Por último vem a tolerância com a reprovação de vendedores normais.

Observamos que a combinação mais pessimista traz um pequeno ganho para o operador; todavia, torna-o muito vulnerável aos classificadores maliciosos, já que a classificação do vendedor dependerá de apenas um classificador. O mesmo ocorre com o resultado seguinte. Já a combinação mais otimista redundará num benefício real: redução de quase um terço nas perdas diretas, com a reprovação de quase 80% dos fraudadores. As combinações intermediárias também resultam em ganhos sensíveis para o operador. Destacamos a combinação da 5ª coluna, que junta os parâmetros extraídos da pesquisa exploratória ($P_{RN}^{\max} = 1\%$ e $P_F = 10\%$) com o valor pessimista para r_F (40%), e consegue recuperar 11% das perdas reprovando 64% dos fraudadores (em média).

O impacto positivo do mecanismo não se limita a reduzir a perda direta: uma elevada reprovação de fraudadores diminui o lucro esperado dos mesmos, reduzindo conseqüentemente o incentivo para tentar fraudar.