Revista da CGU

ANO V - Nº 8
Outubro/2010
ISSN 1981-674X

Controladoria-Geral da União

Controladoria-Geral da União

Revista da CGU

Brasília, DF
Outubro /2010

CONTROLADORIA-GERAL DA UNIÃO – CGU
SAS, Quadra 01, Bloco A, Edifício Darcy Ribeiro
70070-905 - Brasília /DF
cgu@cgu.gov.br

A Revista da CGU é editada pela Controladoria-Geral da União.

Tiragem: 1.500 exemplares
Diagramação e arte: Assessoria de Comunicação Social da CGU
Distribuição gratuita da versão impressa
Disponível também no site www.cgu.gov.br

Revista da CGU / Presidência da República, Controladoria-Geral da União. - Ano V, nº 8, Outubro/2010. Brasília: CGU, 2010.

120 p. Coletânea de artigos.

1.Prevenção e Combate da corrupção. I. Controladoria-Geral da União.

ISSN 1981- 674X
CDD 352.17

Sumário

Nota do editor.....5

Artigos

A aplicação do princípio da proporcionalidade no processo administrativo disciplinar.....8
Alexandro Mariano Pastore e Márcio de Aguiar Ribeiro

Medidas cautelares no processo administrativo sancionador: uma análise da possibilidade de suspensão cautelar do direito de uma pessoa licitar e contratar com a Administração Pública.....20
Luiz Henrique Pandolfi Miranda

Técnicas de mineração de dados como apoio às auditorias governamentais.....28
Carlos Vinícius Sarmento Silva e Henrique Aparecido da Rocha

Eficiência, proporcionalidade e escolha do procedimento disciplinar.....40
Carlos Higino Ribeiro de Alencar

Utilização de pregão nas contratações de obras e serviços de engenharia.....49
Lucimar Cezar Fernandes Silva

Auditoria de TI: proposta de modelo de implementação de auditoria de tecnologia da informação no âmbito da Secretaria Federal de Controle.....60
Maíra Hanashiro

Corrupção na Administração Pública e crimes de "lavagem" ou ocultação de bens, direitos e valores.....70
Paulo Roberto de Araújo Ramos

Legislação

Atos normativos.....88
Legislação em destaque.....91

Jurisprudência

Julgados recentes do TCU - Acórdãos.....96
Julgados recentes de tribunais – Acórdãos.....98

Técnicas de mineração de dados como apoio às auditorias governamentais

Carlos Vinícius Sarmento Silva, mestrando e bacharel em Ciência da Computação (UnB), Analista de Finanças e Controle da CGU, lotado na SIINF/DSI/CGU.

Henrique Aparecido da Rocha, mestre em Ciência da Computação (Unicamp), Analista de Finança e Controle e Gerente de Pesquisas estratégicas/DIE/SPCI.

Resumo

O trabalho de auditoria governamental tem sido realizado no âmbito do Poder Executivo Federal pela Controladoria-Geral da União. Várias estratégias são utilizadas visando a prevenção e o combate à corrupção. No entanto, devido ao crescente aumento de informações nos bancos de dados governamentais, a tarefa de exploração desses dados para geração de conhecimento útil na atividade de auditoria se torna cada vez mais árdua. As técnicas de Mineração de Dados, estudadas na área de Inteligência Artificial, têm sido alvo de várias pesquisas por causa de seus bons resultados no processo de descoberta de conhecimento em grandes volumes de dados. Este artigo trata da aplicação de técnicas de Mineração de Dados em um conjunto de dados reais de licitações realizadas pelo Governo Federal. O objetivo é verificar o potencial das técnicas para lidar com conjuntos de dados provenientes dos sistemas de informação do Governo, procurando assim identificar padrões de interesse que possam subsidiar ações de controle.

1. Introdução

São inquestionáveis a relevância, a importância e o poder da informação na sociedade contemporânea. O elevado número de atividades produtivas que dependem da gestão de fluxos informacionais aliado ao uso intenso das novas tecnologias de informação e comunicação nos introduziu em um novo modelo de organização: a Sociedade da Informação.

E não é diferente com a Administração Pública. Atualmente a maioria dos seus processos é suportada por sistemas computacionais que registram, de forma detalhada, informações sobre programas de governo, finanças, transferências, orçamentos, servidores, diárias, viagens, entre outras.

O SIAFI, por exemplo, registrou mais de um bilhão de transações financeiras

de 24 mil unidades gestoras no ano passado [Portal SIAFI - http://www.tesouro.fazenda.gov.br/siafi/index.asp]. O SIAPE armazena os registros de mais de um milhão de servidores, entre ativos, aposentados e pensionistas [Apresentação SRH/MP - http://www.planejamento.gov.br/secretarias/upload/Arquivos/srh/palestras_apre/090600_politica_gestao.pdf]. Nesses registros estão armazenadas informações sobre pagamentos mensais, afastamentos, progressões e diversas outras ocorrências dos assentos funcionais. O mais recente dos chamados sistemas estruturadores, o SICONV, em pouco mais de um ano de funcionamento, já registra cerca de cinco mil convênios celebrados pelo Governo Federal [Portal SICONV - https://www.convenios.gov.br/portal/arquivos/Boletim_Gerencial_SICONV_n2.pdf]. Essa pequena amostra de sistemas utilizados pelo Governo Federal demonstra a dimensão do conjunto de informações disponíveis atualmente em meio eletrônico.

O Controle Interno utiliza os dados provenientes dos sistemas de informação para planejar e executar auditorias e fiscalizações dos recursos públicos. À medida que esses sistemas incorporam mais informações, antes disponíveis apenas em papel, o trabalho executado pelos auditores da CGU tende a apresentar melhores resultados, com menos recursos logísticos. A maior dificuldade, porém, reside em correlacioná-los para gerar informação útil para os auditores. As alternativas atualmente se restringem a consultas aos sistemas em casos pontuais ou preparação de amostras estatísticas que diminuem o universo para um conjunto reduzido de informações, tratável pela capacidade operacional do Órgão.

Entretanto, existem técnicas computacionais que podem auxiliar na tarefa de produzir conhecimento a partir de grandes volumes de dados, como o são as bases de dados governamentais. Técnicas baseadas em inteligência artificial são amplamente utilizadas para esse fim por empresas, para identificar padrões ou informações úteis para seus negócios.

No Brasil temos um exemplo recente de utilização pelo Banco Nossa Caixa. O banco desenvolveu um sistema de prevenção de transações financeiras fraudulentas baseado em redes neurais, outra técnica disponível na matéria de Inteligência Artificial. Esse sistema tem a missão de correlacionar dados provenientes dos canais de atendimento, tipos de transações e locais comumente usados pelos clientes para identificar e interromper transações suspeitas em tempo real [Portal do Governo do Estado de São Paulo - http://www.saopaulo.sp.gov.br/spnoticias/lenoticia.php?id=103153].

Nos Estados Unidos, temos um grande exemplo na esfera pública. Em 2004, o General Accounting Office – órgão de controle externo – realizou pesquisa para identificar as iniciativas de utilização de mineração de dados pelos órgãos do governo federal. Os resultados indicaram a utilização da técnica para uma variedade de propósitos, desde a melhoria dos serviços prestados até a detecção de padrões ou atividades terroristas. A pesquisa alcançou 128 agências federais e revelou que 52 agências utilizavam ou planejavam utilizar mineração de dados. Essas agências juntas reportaram 199 iniciativas, das quais 68 estavam em planejamento, e 131, ope-

racionais [GAO – http://www.gao.gov/products/GAO-04-548].

Este artigo demonstra a aplicação da técnica de mineração de dados em um conjunto de dados reais de licitações realizadas pelo Governo Federal. O objetivo é verificar se a técnica tem potencial para lidar com conjuntos de dados provenientes dos sistemas de informação do Governo e capacidade para identificar padrões de interesse que possam subsidiar ações de controle. Os dados apresentados neste artigo foram obtidos como resultado preliminar de pesquisa de mestrado em andamento na Universidade de Brasília, com o apoio da Diretoria de Informações Estratégicas da CGU.

2. Mineração de dados

Para lidar com grandes volumes de informação, a utilização de técnicas de mineração de dados tem-se mostrado de grande valia na obtenção de informações potencialmente úteis. Essas técnicas pertencem a um ramo da Ciência da Computação conhecido como Descoberta de Conhecimento em Base de Dados, ou *Knowledge Discovery in Database* (KDD). Na definição de Frawley et al. (1992), KDD é uma extração não trivial de informações implícitas, previamente desconhecidas e potencialmente úteis de uma base de dados. É por isso que aplicações de técnicas de KDD têm sido vistas em diversas áreas, tanto no campo da pesquisa quanto dos negócios e do governo (Fayyad et al., 1996b).

Segundo Fayyad et al. (1996a), o processo de KDD pode ser definido como o processo não trivial de identificar padrões válidos, originais, potencialmente úteis e compreensíveis em determinados bancos de dados. O processo é classificado como não trivial porque envolve decisões que estão além da aplicação das técnicas, como a de definir exatamente o problema que se tem para que assim se possa encontrar um caminho de otimização dos algoritmos de determinado método de mineração de dados. Ainda segundo Fayyad et al. (1996a), o processo de KDD é interativo e iterativo (com muitas decisões tomadas pelo usuário). O processo torna-se iterativo na medida em que os resultados obtidos por meio da mineração de dados fazem pouco ou nenhum sentindo, exigindo assim um recalibramento das funções de mineração relacionando a parte interativa do processo na qual se obtém dados que sejam de fato úteis ao negócio.

Segundo Tan et al. (2005), as tarefas de mineração de dados são geralmente divididas em duas categorias principais:

Tarefas Preditivas: têm como objetivo predizer o valor de um atributo particular baseado nos valores de outros atributos. O atributo a ser predito é conhecido como *alvo ou variável dependente*, enquanto os atributos usados para fazer a predição são conhecidos como *explanatórios ou variáveis independentes*.

Tarefas Descritivas: têm como objetivo derivar padrões como correlações, tendências, grupos, trajetórias e anomalias, as quais sumarizam as relações subjacentes nos dados. Tarefas de mineração de dados descritivas são frequentemente exploratórias e frequentemente requerem a utilização de técnicas

para validar e explanar o resultado (pós-processamento).

Algumas das principais técnicas de mineração de dados são classificação, clusterização e regras de associação.

2.1 Classificação

A técnica de mineração de dados por classificação é uma técnica preditiva. Segundo Tan et al. (2005), classificação é a tarefa de aprender uma função alvo *f* que mapeia cada conjunto de atributos *A* para um dos rótulos de classificação *y*.

Por exemplo, dado um conjunto de características de diversos animais, tais como sistema de temperatura, habitat, sistema ósseo, cobertura da pele, e um atributo de rotulação de classe y (mamífero, réptil, peixe, anfíbio), o algoritmo de classificação aprende a função *f*, chamada também de modelo de classificação, tal que essa função seja capaz de definir regras de classificação a partir das características dada. Exemplo: um animal homeotermo, que tem o corpo coberto de pelo e quadrúpede, é um mamífero. A técnica de Regressão segue a mesma ideia da classificação, com a diferença de que, na classificação, as classes são discretas, e, na regressão, as classes são contínuas.

A técnica de classificação constrói um modelo de classificação (função *f*) baseado em algoritmos de aprendizagem, sendo que o algoritmo empregado tenta identificar um modelo que melhor adapta a reação entre o conjunto de atributos e o rótulo de classe selecionado dos dados de entrada. Existem diversas técnicas de classificação, tais como árvore de decisão, classificadores baseado em regras e redes neurais, bayesianos, entre outros (Tan et al., 2005).

2.2 Clusterização

Segundo and Dubes (1988), *clusterização* é a tarefa descritiva na qual se procura identificar um conjunto finito de categorias ou "clusters" para descrever uma informação. Essas categorias podem ser mutuamente exclusivas ou não.

A análise de *cluster* está relacionada com outras técnicas que são usadas para dividir objetos de dados em grupos. Por exemplo, a *clusterização* pode ser considerada como a forma de classificação em que se cria uma rotularização de objetos com rótulos de classe (que são os *clusters*). Entretanto, esses rótulos são derivados unicamente dos dados.

Em contraste, o processo propriamente dito de classificação é uma classificação supervisionada, isto é, objetos novos e não rotulados recebem um rótulo de classe usando um modelo desenvolvido a partir de objetos com rótulos de classes já conhecidos. Por essa razão, análise de *clusters* é algumas vezes referida como uma espécie de classificação não supervisionada (Tan et al., 2005).

2.3 Regras de associação

Essa técnica de mineração de dados consiste em descobrir relações fortes entre determinadas informações. Tem a capacidade de detectar padrões em forma de regras que associam valores de atributos num determinado conjunto de dados. Essas regras são expressas por meio de cláusulas da seguinte forma:

IF $atrib_1 = valor_1$ **AND** $abrib_2 = valor_2$ **AND**...
THEN $atrib_n = valor_n$ AND $atrib_{n+1} = valor_{n+1}$...

Em que *atrib* é um atributo do conjunto de dados e *valor* é o valor do atributo identificado na regra.

Segundo Witten and Frank (2005), diferença entre classificação e regras de associação é que estas podem prever padrões com qualquer atributo, e não só da classe selecionada. Diferentes regras de associação expressam diferentes regularidades subjacentes no conjunto de dados, cada uma prevendo coisas diferentes.

A cobertura das regras de associação é o número de instâncias em que a regra se repete, e é chamada de *suporte*. A acurácia da regra, chamada de *confiança*, é o número de instâncias que a regra prediz corretamente, e é expressa como uma proporção de todas as instâncias em que a regra se aplica.

Por exemplo, na seguinte regra:

temperatura=frio > umidade=normal

O suporte será o número de instâncias na base de dados em que o atributo *temperatura* seja *frio* e o atributo *umidade* seja *normal*. Já a confiança será a proporção de instâncias com temperatura fria que tenham umidade normal.

3. Estudo de caso

Foram realizadas atividades de mineração de dados numa base de licitações extraída do sistema ComprasNet, em que são realizados os pregões eletrônicos do Governo Federal. Os dados são relativos a todas as licitações para contratação de um determinado tipo de serviço na modalidade de Pregão para órgãos do Poder Executivo Federal, durante os anos de 2005 a 2008, em todos os estados da Federação. Os testes foram executados utilizando a ferramenta Weka (Environment for Knowledge Analysis) e reúnem vários algoritmos para execução de tarefas de mineração de dados (Witten and Frank, 2005).

A Tabela 1 mostra algumas informações da base de dados utilizada nos experimentos. Cada registro da base de dados representa uma participação de uma empresa numa determinada licitação.

Tabela 1: Base de dados utilizada nos experimentos preliminares	
Informações	**Total**
Registros	26615
Licitações	2701
Empresas	3051
Empresas que já ganharam pelo menos uma licitação	1162
Empresas que já ganharam pelo menos cinco licitações	121

3.1. Primeiro experimento

Dois *datasets* foram preparados, no intuito de aplicar as técnicas de regras de associação para detecção de grupos suspeitos de fazer rodízio de licitações. O algoritmo utilizado nesse experimento foi o *Apriori,* apresentado em and Srikant (1994) e disponível na versão 3.6.1 da ferramenta *Weka*.

O primeiro *dataset* foi construído contemplando todas as licitações da base e todos os fornecedores. Já o segundo contemplou apenas os fornecedores que já tinham participado de pelo menos duas licitações (Tabela 2). Essa escolha se deu pelo fato de estarmos procurando grupos de empresas atuando em cartéis, não fazendo assim sentido procurar entre aquelas que participaram de apenas uma licitação.

Tabela 2: Resultados da execução do *Apriori* para os dois *datasets*

Instâncias	Atributos	Suporte Mín.	Confiança Mín.	Regras Obtidas
2701	3051	1,00%	70,00%	294
2370	1086	1,00%	80,00%	145

A estratégia usada para procurar associação de fornecedores foi organizar os *datasets* de forma que cada fornecedor da base fosse um atributo, e cada instância fosse uma licitação. Assim, para cada licitação, o atributo relativo a um determinado fornecedor era preenchido com o valor 'sim', caso ele tivesse participado da licitação, ou 'não', caso contrário.

A preparação dos *datasets* para regras de associação resumiu-se em construir a matriz A por m linhas e $n+1$ colunas, tal que:

m = número total de licitações da base
n = número total de fornecedores da base de dados

$$a_{i,j} = \begin{cases} \text{sim (se fornecedor } j \text{ da licitação } i) \\ \text{? (caso contrário)} \end{cases}$$

$$a_{i,n+1} = \text{vencedor } (i)$$

$$\forall i \leq m, \forall j \leq n, i, j \in \mathrm{N}^{*}$$

O valor *"?"* é entendido pelos algoritmos do Weka como valor faltante (*missing value*). A inserção desse símbolo na nossa matriz foi proposital, para suprimir regras envolvendo a não participação de fornecedores, o que certamente ocorreria caso substituíssemos o valor "?" pelo valor "não". Como o nosso interesse é encontrar regras indicando a participação de fornecedores em licitações, optamos por usar o símbolo *"?"* na matriz.

A Tabela 2 mostra o resultado da execução do algoritmo *Apriori datasets*.

A escolha de valores altos na configuração do suporte mínimo para execução do algoritmo não nos garante boas regras para identificação de cartéis. Uma regra que associa alguns fornecedores e que tem suporte alto provavelmente indica a presença de grandes fornecedores participando de várias licitações. Dessa forma, a configuração de um suporte mínimo alto para execução do algoritmo pode suprimir a aparição

de diversas regras boas, com reais características de cartéis. Valores altos de confiança, por sua vez, garantem a seleção de regras boas. Assim, foi definida uma função de avaliação de regras, com ajuda do especialista, para poder classificar e selecionar as melhores regras obtidas. Isso porque, com a redução do suporte mínimo, muitas regras foram obtidas na execução do algoritmo, como mostra a Tabela 2.

3.2 Segundo experimento

A tentativa de aplicar a técnica de regras de associação em dados de todo o país deixa o espaço de soluções bastante esparso.

O estudo do negócio possibilitou verificar que, muitas vezes, os fornecedores não se restringem necessariamente a regiões macroeconômicas. Um exemplo típico é a situação de Mato Grosso do Sul, Goiás e Tocantins. Embora Mato Grosso do Sul e Goiás pertençam à mesma região, é mais provável que os fornecedores do estado de Goiás atendam ao estado de Tocantins, pela proximidade geográfica, do que ao estado de Mato Grosso do Sul, embora Tocantins pertença a uma região diferente.

Dessa forma, a necessidade de aplicar técnicas de *clusterização* para mapear os grupos comuns de atuação dos fornecedores se tornou necessária.

Foi aplicado o algoritmo *Expectation-Maximization* (E.M.) na base de dados, com o objetivo de definir as regiões geográficas comuns de participação de empresas em licitações. O algoritmo foi executado nas 26615 instâncias da base, tendo como atributos *Fornecedor* a *UF* participou da licitação. A execução do algoritmo trouxe como resultado 10, conforme apresentado na Tabela 3.

Tabela 3: *Clusters* de Estados

Cluster	Estados
1	SP, MT, MS, AL, CE, PB, PE, PI, RN
2	DF
3	RJ
4	ES, MG
5	AP, MA, PA
6	PR
7	RS, SC
8	GO, TO
9	AC, AM, RR, RO
10	BA, SE

Na Figura 1, pode-se notar que a maioria dos clusters encontrados tem como característica a proximidade geográfica.

FIGURA 1: *Clusters* encontrados com a execução do algoritmo E.M.

3.3 Terceiro experimento

A partir das regiões obtidas no Segundo Experimento, foi aplicada a técnica de regras de associação em cada *cluster*, na tentativa de identificar grupos de empresas associadas atuando especificamente na região.

Os resultados desse experimento podem ser vistos na Tabela 4.

3.4 Avaliação dos resultados obtidos

Com ajuda dos especialistas, definimos também um método de avaliação das regras que seriam obtidas por meio do processo de mineração de dados. A fórmula de avaliação definida foi:

$$M = 100.V(F)/Suporte \quad (1)$$

Sendo que *V(F)* é a função que retorna o número de vezes que algum fornecedor do grupo F de fornecedores ganhou uma licitação da qual todo o grupo participou. As regras foram avaliadas por meio da Equação 1. Para análise dos resultados, foram selecionadas as 10 melhores regras, segundo a função de avaliação. Os valores mínimos de suporte e confiança foram 9 (absoluto) e 80%, respectivamente.

Tabela 4: Execução do APRIORI para *datasets* de *clusters*					
Cluster	**Inst.**	**Atrib.**	**Sup.**	**Conf.**	**Regras**
1	787	614	2,00%	80,00%	851
2	211	164	4,00%	80,00%	1406
3	261	166	3,00%	80,00%	100
4	194	257	5,00%	80,00%	86
5	134	168	6,00%	80,00%	115
6	98	152	9,00%	80,00%	2848
7	270	196	4,00%	80,00%	1679
8	94	118	1,00%	80,00%	3
9	211	204	4,00%	80,00%	22
10	134	259	10,00%	80,00%	5869

Foram selecionadas as 10 melhores regras obtidas nos Experimentos 1 e 3. As melhores regras obtidas no Experimento 1 tiveram na média melhor suporte que as melhores regras obtidas pelos Modelos gerados no Experimento3. No entanto, as regras obtidas no Experimento 3 tiveram um aumento no valor de avaliação de cerca de 100%. O gráfico da Figura 2 mostra a comparação.

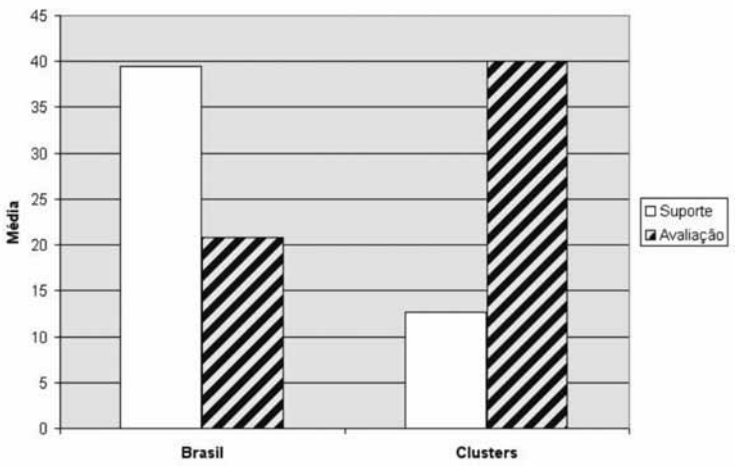

FIGURA 2: Média de suporte e avaliação das 10 melhores regras

Os resultados mostram também que as melhores regras na nossa base, segundo a avaliação adotada, tendem a aparecer quando o suporte é baixo e quando há uma melhor definição do espaço de soluções, nesse caso, definido pelos *clusters* encontrados. Por isso as regras abrangendo o Brasil todo não foram tão boas quanto as encontradas em regiões específicas do País.

Entre os modelos gerados a partir dos clusters (Tabela 4), as melhores regras foram obtidas no *Cluster* 6. A comparação entre as 10 melhores regras obtidas nesses modelos pode ser vista no gráfico da Figura 3.

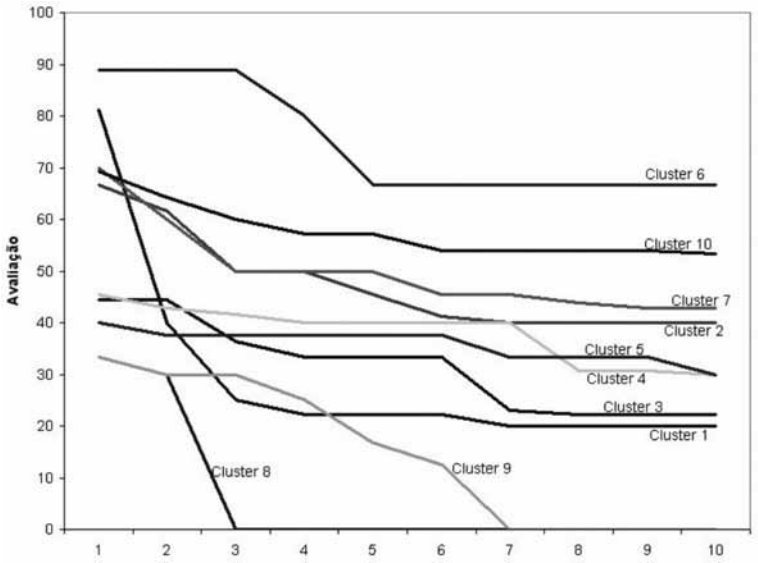

FIGURA 3: Comparação das 10 melhores regras de cada *cluster*

3.5 Conhecimento descoberto

O modelo de *cluster* gerou interesse por parte do especialista, que explicou que as atividades de rodízio de licitações são tipicamente regionais. Isso significa que, mesmo que uma empresa tenha atuação em âmbito nacional e pratique rodízio de licitações com um grupo, é improvável que esse grupo atue em todo o país. Assim, a regra que apresenta uma associação de fornecedores provavelmente em conluio teria maior suporte em apenas uma região, que seria a região de atuação do cartel.

O *Cluster* 1 trouxe um resultado interessante, por fugir do padrão de regionalização geográfica. Os estados de São Paulo, Mato Grosso e Mato Grosso do Sul agruparam-se com os estados de Alagoas, Ceará, Paraíba, Pernambuco, Piauí, Rio Grande do Norte. Esse resultado trouxe outras propostas de pesquisa, no intuito de levantar, dentre as empresas que atuaram nesses estados, quais delas contribuíram para essa distribuição atípica nas participações em licitações. Um rápido levantamento mostrou 76 empresas que atuaram na região formada pelos estados de Alagoas, Ceará, Paraíba, Pernambuco, Piauí e Rio Grande do Norte e na região formada pelos estados de São Paulo, Mato Grosso e Mato Grosso do Sul. Dessas empresas, 8 participaram de mais de 15 licitações, tanto numa região quanto na outra. Dessas 8 empresas, nenhuma é da sub-região composta por São Paulo ou Mato Grosso ou Mato Grosso do Sul.

As próximas atividades serão no sentido de experimentar novas bases, na tentativa de detectar outros *clusters* de interesse nas investigações, como, por exemplo, *clusters* envolvendo órgãos superiores.

Quanto às regras de associação, algumas das melhores regras foram selecionadas pelo especialista para verificação.

Grupos de empresas foram detectados onde a média de participações juntas e as vitórias em licitações levavam a indícios de conluio. Alguns exemplos de regras encontradas são detalhados abaixo:

- Duas empresas de um mesmo estado, sendo que o total de licitações de que cada uma participou individualmente foi de 75 e 78 licitações. Dentre essas, em 68 licitações, participaram juntas ganhando 14 contratos entre os anos de 2005 a 2007.
- Outra regra envolveu 3 empresas que somavam 14 certames de participação conjunta. O grupo celebrou 8 contratos com a Administração. Cada uma delas tinha uma média de participação individual relativamente baixa na base de dados (média de 30 licitações).
- No ano de 2008, uma empresa ganhou 9 licitações num mesmo órgão, concorrendo com outra empresa que não ganhou nenhuma das licitações de que ambas participaram. O detalhe é que as 9 licitações que a segunda empresa perdeu foram as únicas licitações de que ela participou na base de dados. O histórico de vitórias da primeira empresa na base de dados não passa de 12 certames.

Embora o processo utilizado tenha permitido a seleção de boas regras, a fórmula de avaliação ainda pode ser melhorada. Isso porque a fórmula selecionou também regras que mostravam apenas coincidência de participações conjuntas de empresas em processos de licitação, em especial no caso de serem grandes fornecedores.

Conclusão

Foram apresentados neste trabalho aplicações de técnicas de inteligência artificial como suporte às atividades de auditoria e combate à corrupção.

Os resultados da aplicação de técnicas de mineração de dados geraram um conjunto de conhecimentos interessantes, que pode ser utilizado pelos auditores quando em execução de ações de controle. A análise de *clusters* apresentou, no caso estudado, fortes indícios de cartelização, informação reforçada pelas regras de associação selecionadas na sequência. Dessa forma, a técnica demonstra grande potencial para processar os volumes de dados armazenados nos sistemas de informação do Governo na busca de fragilidades ou mesmo de fraudes nos processos controlados por esses sistemas.

O estudo também se mostrou bastante promissor quanto à aplicação das técnicas apresentadas em bases de dados de outras áreas de despesas e serviços contratados, podendo ser largamente utilizada como insumo para o trabalho do auditor.

Em estudos futuros, serão analisadas de forma mais profunda as regras de associação, objetivando a construção de um modelo de seleção mais apropriado, que priorize as regras mais importantes. Além disso, será proposta uma integração das atividades de mineração de dados com sistemas multiagentes. Dessa forma, espera-se automatizar os processos de mineração de dados, além de enriquecer o conhecimento descoberto, fazendo uso de mais de uma técnica, de forma cooperativa, autônoma e independente.

Referências Bibliográficas

AGRAWAL, R. and SRIKANT, R. (1994). *Fast algorithms for mining association rules in large databases.* In '94: Proceedings of the 20th International Conference on Very Large Data Bases, pages 487–499, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

FAYYAD, U., PIATETSKY-SHAPIRO, G., and SMYTH, P. (1996a). *The kdd process for extracting useful knowledge from volumes of data.* Commun. ACM, 39(11):27–34.

FAYYAD, U. M., PIATETSKY-SHAPIRO, G., and SMYTH, P. (1996b). *From data mining to knowledge discovery: an overview.* In Advances in knowledge discovery and data mining, pages 1–34. American Association for Artificial Intelligence, Menlo Park, CA, USA.

FRAWLEY, W. J., SHAPIRO, P. G., and MATHEUS, C. J. (1992). *Knowledge discovery in databases - an overview.* Ai Magazine, 13:57–70.

JAIN, A. K. and DUBESs, R. C. (1988). *Algorithms for clustering data.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

TAN, P.-N., STEINBACH, M., and KUMAR, V. (2005). *Introduction to Data Mining.* Addison Wesley, us ed edition.

WITTEN, I. H. and FRANK, E. (2005). Data Mining: *Practical Machine Learning Tools and Techniques.* Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, second edition.