

DATA MINING PARA AVALIAÇÃO DO CONHECIMENTO SOBRE EFEITOS PATRIMONIAIS NAS FORMAS DE UNIÃO CONJUGAL

Luiz Guilherme Carvalho Guimarães¹

João Vitor Moraes²

Deborah Ribeiro Carvalho³

RESUMO

O presente trabalho tem como objetivo identificar o potencial de uso das técnicas de *Data Mining*, bem como as formas de pós-processar os padrões descobertos para apoio ao processo decisório. Para tanto, foi construído um experimento com vistas a avaliar o quanto a população está ciente das diferenças entre as distintas formas de união conjugal, assim como as suas implicações sobre os efeitos patrimoniais decorrentes de cada uma delas. Dessa forma, realizou-se uma pesquisa de campo a partir da aplicação de um formulário contendo 17 questões, com o propósito de identificar o nível de conhecimento das pessoas sobre tal questão. A fim de melhor explorar os dados, construiu-se um plano tabular para extração de informações (*software* SPSS), identificaram-se as regras de associação (APRIORI) e as regras de exceção. As regras gerais (de senso comum) comumente representam padrões conhecidos pelo usuário/gestor, ao contrário das regras de exceção. Sendo assim, as regras de exceção tendem a ser mais interessantes/surpreendentes, dado o fato de representarem uma contradição em relação à regra geral.

Palavras-chave: *Data Mining*; práticas do direito; descoberta de conhecimento.

¹ UNICENP – Bacharel em Direito

² IPARDES – Instituto Paranaense de Desenvolvimento Econômico e Social, UFPR

³ IPARDES – Instituto Paranaense de Desenvolvimento Econômico e Social, UTP

1 INTRODUÇÃO

Ao longo da história, percebem-se várias diferenças no âmbito familiar, tanto no que tange às relações internas, de proteção recíproca, quanto relativamente à sociedade. Inicialmente, acreditava-se que a família tinha por missão a conservação dos bens, a prática de um ofício comum a todos e a ajuda mútua para enfrentar o dia-a-dia, no qual homens e mulheres não poderiam sobreviver independentemente.

No século XVI, incia-se um processo de nuclearização da família, juntamente com a individualização de seus membros, que começam a valorizar a privacidade, o que modifica fortemente os espaços habitacionais, ocorrendo a diluição das famílias conjugais.

Somente no século XIX, a partir da família conjugal, apresentou-se a marca da "indivisão", agrupada sob a razão das relações de trabalho, que originou a família patriarcal.

Alguns estudiosos, como Freyre, Cândido, Berquó, Goldani e outros, salientam na família brasileira a desagregação patriarcal, que resultou em "novos modelos de família" ou "arranjos familiares" (MEDEIROS, 2002).

Na linha dessas mudanças, há também que se considerar a visão jurídica sobre a família, ou seja, na área dos direitos e responsabilidades dela decorrentes, a igualdade de gênero e de oportunidades também constitui papel decisivo nesses novos "arranjos".

Os levantamentos mais detalhados sobre outras formas de união são bastante recentes, sendo que a coleta de dados pelos Censos Demográficos (Brasil) ocorreu a partir de 1960.

Apesar de essas questões estarem bastante claras sob a perspectiva legal, bem como dados levantados, pouco se tem conhecimento sobre o quanto a população está realmente ciente do assunto, principalmente no que se refere aos efeitos patrimoniais envolvidos.

Este artigo tem como objetivo identificar o potencial de uso das técnicas de *Data Mining* para investigar o quanto as pessoas estão cientes das diferenças entre as

distintas formas de união conjugal e as suas implicações sobre os efeitos patrimoniais decorrentes de cada uma delas.

2 DATA MINING

Nas bases de dados disponíveis, existem informações preciosas, mas muitas são de difícil identificação/utilização, pois muitas vezes a atividade de avaliação ultrapassa a capacidade humana de análise e interpretação. Para facilitar a recuperação e o uso desses dados, uma das alternativas possíveis é a aplicação do processo de *Knowledge Discovery in Database* (KDD). Segundo Fayyad (1996), esse processo é composto das seguintes etapas:

- Seleção de dados - prevê a coleta e seleção dos dados;
- Limpeza - prevê a análise dos dados coletados verificando a existência de ruídos, tratamento de valores ausentes etc.;
- Transformação ou enriquecimento dos dados - trata-se da incorporação de novos dados criados a partir dos já existentes;
- *Data Mining* - consiste em aplicar um algoritmo que efetivamente procura por padrões/relações e regularidades em um determinado conjunto de dados;
- Interpretação e avaliação - verifica a qualidade do conhecimento (padrões) descoberto, procurando identificar se o mesmo auxilia a resolver o problema original que motivou a realização do processo KDD.

Os algoritmos de *Data Mining* podem ser identificados em três tarefas principais: classificação, descoberta de regras de associação e *clustering*.

Neste artigo será aplicada a etapa de *Data Mining*, mais especificamente a tarefa de descoberta de regras de associação e a identificação de regras de exceção a partir desse conjunto descoberto.

A justificativa para a adoção da etapa de descoberta das regras de exceção se deve ao fato de em geral o volume de padrões encontrados ser muito grande, o que dificulta a sua análise e, fundamentalmente, inviabiliza o seu uso no apoio à tomada de decisão. Além disso, nestes conjuntos de padrões é freqüente existirem redundâncias, ou relações irrelevantes (CARVALHO, 2005).

3 DESCOBERTA DE REGRAS DE ASSOCIAÇÃO

Por meio da tarefa de descoberta de regras de associação identificam-se afinidades entre itens de um subconjunto de dados, e essas afinidades são expressas na forma de regras.

Um dos padrões mais comuns encontrados por meio do processo *Data Mining* são os conjuntos de regras de associação que expressam a probabilidade de um item ocorrer em conjunto. Por exemplo, 80% dos clientes que adquiram o produto A também adquiriram o produto B.

A associação resume-se a descobrir afinidades entre os dados de certa natureza mediante um determinado número de transações. Dessa forma, os algoritmos que identificaram regras de associação objetivam encontrar relacionamentos entre os dados (BERRY, 1997). Tendo-se um conjunto de registros, em que cada qual constitui um conjunto de dados, uma regra de associação é uma expressão do tipo $X \rightarrow Y$ (Se X, então Y), em que X e Y são itens (conjunto de itens) na forma $X \cap Y = \emptyset$.

Os algoritmos de associação fornecem ao usuário um conjunto de regras que nem sempre pode ser analisado em função da quantidade. Por essa razão, faz-se necessário um refinamento/redução desse conjunto, viabilizando, assim, a análise/uso das regras.

4 PÓS-PROCESSAMENTO DO CONJUNTO DE REGRAS DESCOBERTAS

Embora as regras de associação sejam padrões valiosos, por permitirem uma percepção útil da dependência entre atributos da base de dados, elas apresentam dois inconvenientes: a quantidade de regras geradas pode ser muito grande, e nem todas são de qualidade e interesse. Uma das alternativas para minimizar esses problemas é a adoção de técnicas de pós-processamento. Desse modo, o número de regras descobertas pode ser reduzido, facilitando a avaliação.

Neste trabalho, além de extrair informações e descobrir regras de associação, foram identificadas regras de exceção. É importante observar que uma regra de exceção é uma especialização de uma regra de senso comum (geral) e nega o conseqüente previsto pela regra de senso comum. Esse método assume que regras de senso comum figuram padrões conhecidos pelo usuário, tendo em vista que elas têm uma grande cobertura, ao contrário das regras de exceção, que em geral são desconhecidas, uma vez que são de baixa cobertura. Assim, as regras de exceção tendem a ser surpreendentes, por representarem uma contradição em relação à regra de senso comum (CARVALHO; MOLIN, 2005).

$Y \leftarrow X$: regra de senso comum (alta cobertura e alta precisão).

$\neg Y \leftarrow X, B$: regra de exceção (baixa cobertura, alta precisão).

$\neg Y \leftarrow B$: regra de referência (baixa cobertura e/ou baixa precisão).

A partir desse conjunto de regras, "se X, então Y" corresponde à regra geral, tendo como exceção a regra "se (X e B), então (não) Y". Essa notação de seta invertida (\leftarrow) é o padrão de saída do APRIORI (BORGELT, 2004).

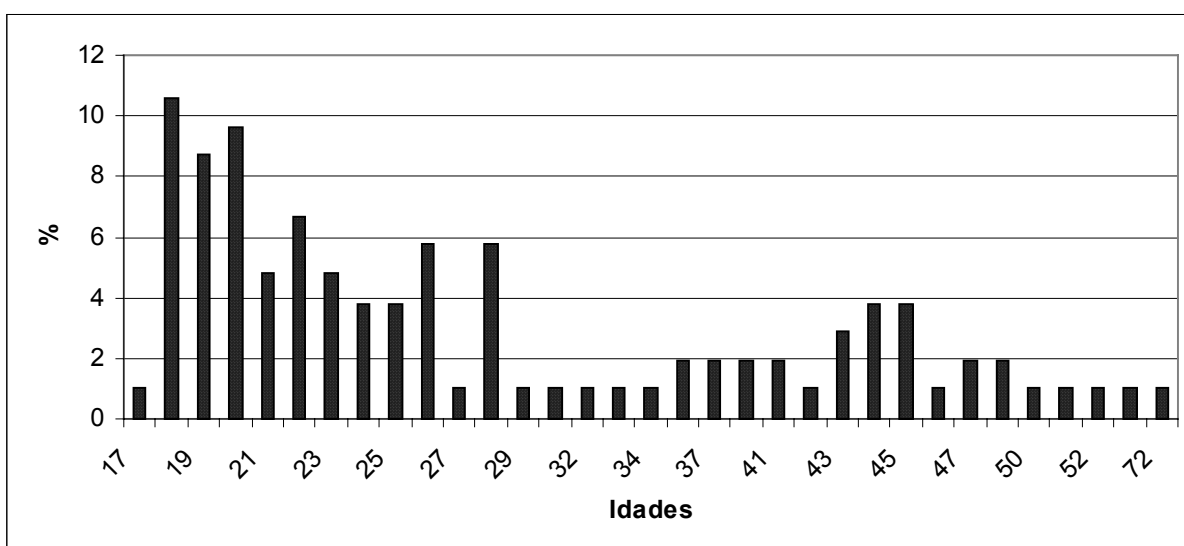
5 BASE DE DADOS ADOTADA

Especificamente para este trabalho foi realizada uma pesquisa de campo por meio da aplicação de um formulário contendo 17 questões, considerando 110 respondentes, com o objetivo de identificar o quanto as pessoas têm ciência de conhecer

e/ou desconhecer as distintas formas de união conjugal, bem como avaliar se aquelas que declararam saber das diferenças realmente demonstram conhecimento sobre o que distingue cada uma delas.

Na figura 1 é possível perceber uma concentração dos entrevistados com idade entre 19 e 29 anos. Essa concentração ocorreu propositadamente, devido ao fato de as pessoas geralmente iniciarem seus relacionamentos afetivos nesse período da vida.

FIGURA 1 - DISTRIBUIÇÃO DE FREQUÊNCIA RELATIVA DA IDADE DOS RESPONDENTES

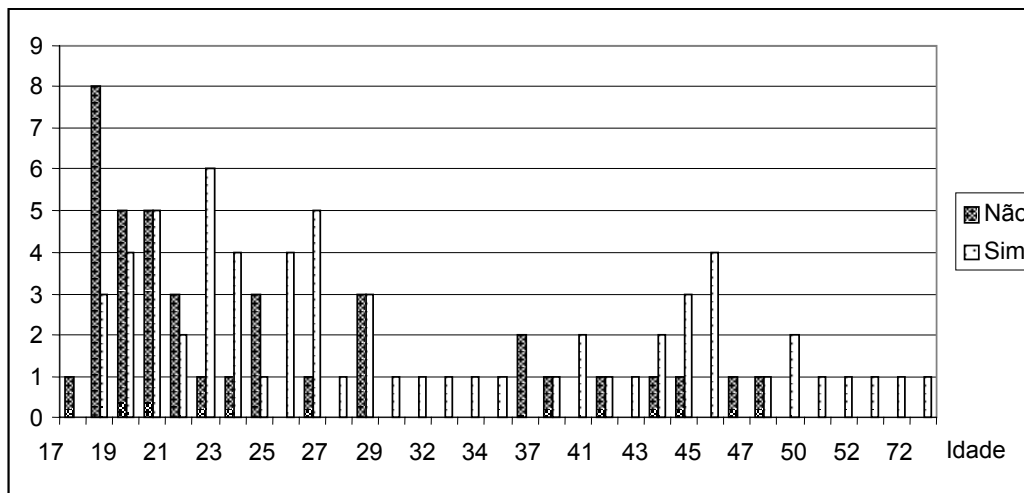


6 RESULTADOS

A fim de melhor explorar os dados, construiu-se um plano tabular para extração de informações, bem como foram descobertas regras de associação. Para a execução do plano tabular, utilizou-se a ferramenta SPSS; e para a identificação de regras de associação, aplicou-se o *software* APRIORI (BORGELT, 2004).

A figura 2 permite avaliar, por estrato etário, a frequência dos entrevistados que se declaram conhecedores ou não das diferenças entre namoro, união estável, concubinato e casamento, sendo possível perceber que não apenas pessoas mais jovens se declaram desconhecedoras, mas também aquelas acima dos 40 anos.

FIGURA 2 - DISTRIBUIÇÃO PERCENTUAL, POR FAIXA ETÁRIA, DOS ENTREVISTADOS E RESPOSTA PARA A QUESTÃO 1 ("VOCÊ SABE A DIFERENÇA ENTRE NAMORO, UNIÃO ESTÁVEL, CONCUBINATO E CASAMENTO?")



As respostas mostram que a maioria das pessoas diz saber quais são as diferenças entre namoro, união estável, concubinato e casamento (figura 3), contudo boa parte não sabe afirmar se tais relações implicam efeitos patrimoniais (figura 4).

FIGURA 3 - RESPOSTA À PERGUNTA "VOCÊ SABE A DIFERENÇA ENTRE NAMORO, UNIÃO ESTÁVEL, CONCUBINATO E CASAMENTO?"

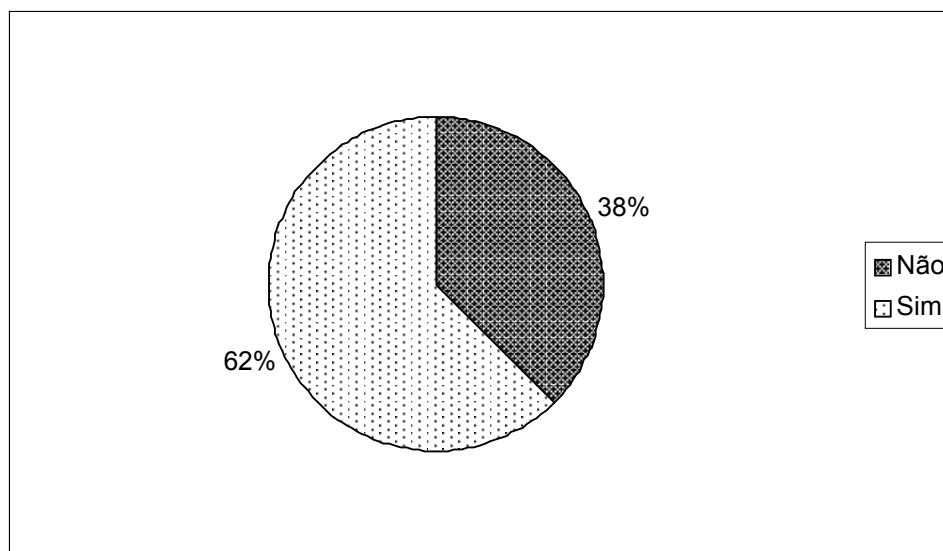
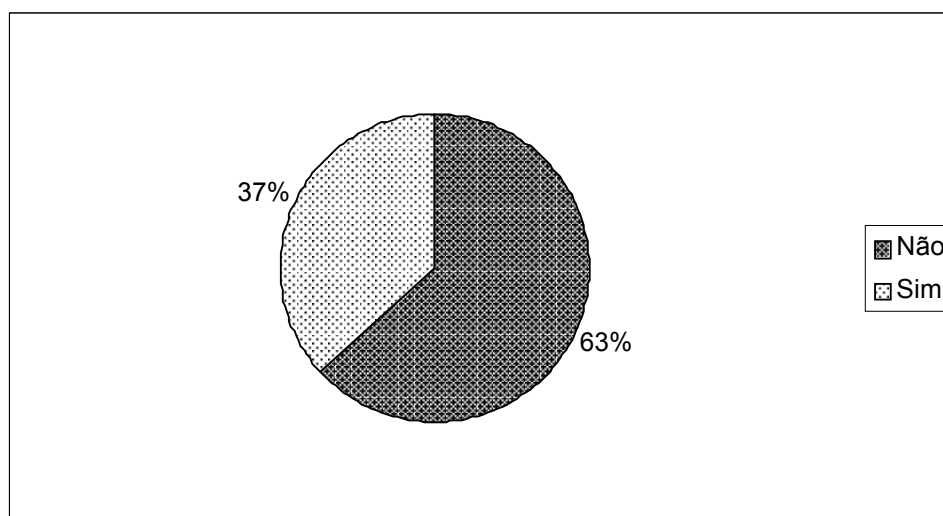


FIGURA 4 - RESPOSTA À PERGUNTA "VOCÊ SABE QUAIS OS EFEITOS PATRIMONIAIS DECORRENTES DE CADA UM DOS REGIMES DE BENS (CASAMENTO) DISPOSTOS NA LEGISLAÇÃO?"



Com relação especificamente ao significado de concubinato, apenas 40% dos entrevistados responderam corretamente (figura 5). Esse desconhecimento talvez se deva à recente alteração da definição legal, que antes da entrada em vigor do novo Código Civil tinha significado sinônimo ao da união estável, e passou a corresponder à relação não eventual entre pessoas impedidas de se casar. Porém, considerando que a grande maioria dos entrevistados esteja cursando Ensino Superior ou já o tenha concluído, era de se esperar que tivessem maior acesso à informação que a maior parte da população.

Sobre as respostas obtidas para a pergunta "Com qual forma de tratamento dos bens o concubinato mais se assemelha?", apenas 15% dos questionados responderam de forma correta, ou seja, sociedade de fato (figura 6). Isso revela um desconhecimento ainda maior sobre o tema, o que é preocupante, dado que no Brasil a prática do concubinato é muito comum.

FIGURA 5 - RESPOSTA À PERGUNTA "COM QUAL FORMA DE TRATAMENTO DOS BENS O CONCUBINATO MAIS SE ASSEMELHA?"

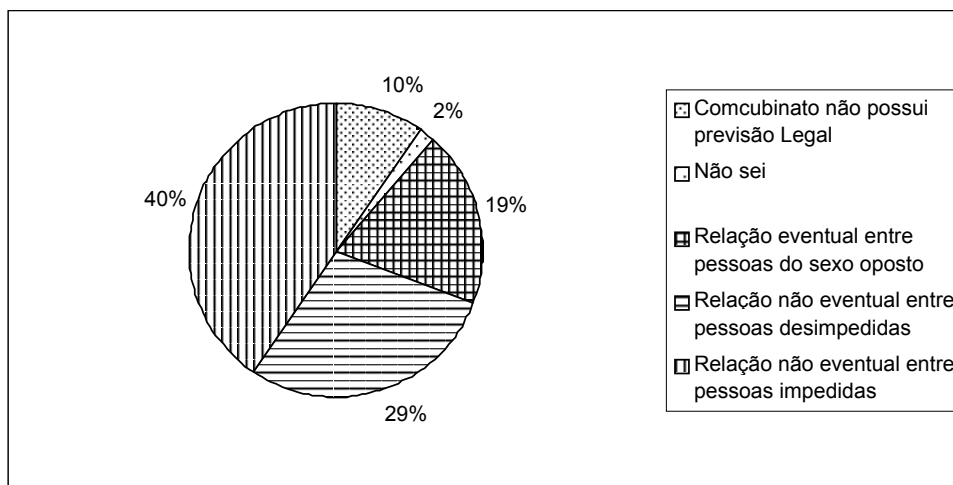
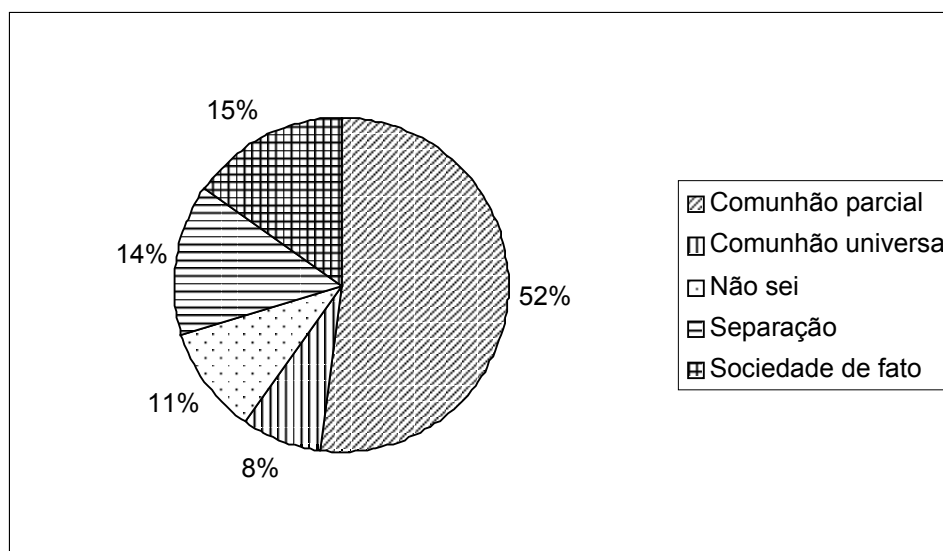


FIGURA 6 - RESPOSTA À PERGUNTA "NO CASO DE CONCUBINATO, A FORMA DE TRATAMENTO DOS BENS SE ASSEMELHA MAIS A QUAL DAS LISTADAS A SEGUIR?"



Com o objetivo de melhor entender o padrão de respostas dos entrevistados, a base de dados correspondendo a todas as respostas foi oferecida como entrada para o algoritmo APRIORI (BORGELT, 2004), o qual descobriu um conjunto de 684.000 regras, como, por exemplo, as duas regras (A e B) a seguir:

Regra A - `questao3_sim <- questao11_comunhao_parcial questao4_sim conf = 83.3`

Regra B - `questao16_nao <- questao11_comunhao_parcial questao4_sim conf = 87.0`

As regras associam as respostas à questão 3 ("No caso de união estável, você acredita ser possível a atribuição de efeitos patrimoniais à relação?"), à

questão 4 ("No caso de casamento, você acredita ser possível a atribuição de efeitos patrimoniais à relação?"), à questão 11 ("No caso de concubinato, a forma de tratamento dos bens se assemelha mais a qual das listadas a seguir?") e à questão 16 ("A legislação brasileira permite que uma pessoa case pela segunda vez sem ter se divorciado da primeira companheira?"). Percebe-se que, apesar de os entrevistados acertarem a resposta à questão 4, desconhecem a situação proposta na questão 11; e em ambos os casos eles acertam as respostas às questões 3 e 16.

Ao comparar-se o potencial de contribuição do conjunto de gráficos *versus* algumas regras de associação na construção da análise/diagnóstico do problema proposto, fica claro que em algumas situações seria absolutamente possível representar graficamente as regras A e B. Porém, para tanto se exigiria que o usuário entendesse e identificasse como uma hipótese interessante a ser investigada a relação entre tais atributos, situação que nem sempre é fácil, tendo em vista a complexidade gerada na combinação de todos os atributos disponíveis no problema.

Contudo, apesar da grande contribuição que as regras de associação oferecem ao gestor, é desnecessário enfatizar a inviabilidade de avaliação e análise adequadas sobre um conjunto de tal magnitude. Sendo assim, esse conjunto de regras foi submetido à etapa de pós-processamento, a qual procurou descobrir regras gerais e suas respectivas regras de exceção. A seguir, são descritos alguns desses subconjuntos.

Regra geral 1:

```
questao1_Sim <- questao6_Nao questao15_Sim questao4_Sim conf = 83.3
```

Exceção 1:

```
questao1_Nao <- questao6_Nao questao17_Nao questao15_Sim questao4_Sim conf = 100.0
```

O par de regras (1) tem como conseqüente a resposta "sim" ou "não" à questão 1 ("Você sabe a diferença entre namoro, união estável, concubinato e casamento?"). A análise desse par de regras permite identificar que, em geral, as pessoas que respondem corretamente às questões 4, 6 e 15 declaram saber essas diferenças. Porém, a exceção a essa regra ocorre quando a pessoa responde

equivocadamente à questão 17 ("A legislação brasileira estipula que os filhos havidos dentro do casamento deverão receber tratamento igual aos havidos fora dele?"). Para esses casos, a pessoa tem ciência de que desconhece tais diferenças.

Na avaliação desse par de regras (1), verifica-se que uma pessoa, por mais experiência que tenha em determinada área do conhecimento, dificilmente perceberia tais sutilezas apenas considerando a análise de tabelas de frequências ou mesmos gráficos relacionando atributos. Dessa maneira, evidencia-se a real importância de disponibilizar ao usuário/gestor padrões descobertos a partir de bases de dados, oferecendo, assim, condições à elaboração de um diagnóstico mais elaborado do problema que originou a construção de uma ferramenta/elemento de apoio à decisão. Essa mesma situação pode ser observada na análise dos pares de regras a seguir.

Regra geral 2:

```
questao13_Nao <- questao10_Relacao_Naoeventualpessoasimpedidas questao15_Sim
questao4_Sim conf = 80.8
```

Exceção 2:

```
questao13_Naosei <- questao10_Relacao_Naoeventualpessoasimpedidas
curso_Administracao questao15_Sim questao4_Sim conf = 100.0
```

O par de regras (2) tem como conseqüente a resposta à questão 13 ("Concubinato e União Estável são sinônimos?"). A análise desse par de regras permite identificar que, em geral, as pessoas que respondem acertadamente às questões 4, 10 e 15 afirmam corretamente que essas duas situações não são sinônimas. Todavia, a exceção a essa regra ocorre quando os respondentes são formados em Administração, pois neste caso declaram não saber se são ou não sinônimos.

Regra geral 3:

```
questao14_Naosei <- questao6_Nao questao13_Naosei questao15_Sim conf =80.0
```

Exceção 3:

```
questao14_Sim <- sexo_Feminino questao6_Nao questao13_Naosei questao15_Sim conf =
100.0
```

O par de regras (3) tem como conseqüente a resposta à questão 14 ("O concubinato é considerado pela legislação instituto capaz de formar família?"). A análise desse par de regras possibilita identificar que as pessoas que respondem corretamente à questão 15 (as respostas às questões 6 e 13 são opiniões) declaram não saber a questão 14. Percebe-se que a exceção a essa regra ocorre com os respondentes do sexo feminino, que declaram saber a resposta à essa questão.

Regra geral 4:

```
questao15_Sim <- questao12_Relacao_Naoeventualpessoasde_Simpedidas questao2_Nao
questao1_Sim conf =81.0
```

Exceção 4:

```
questao15_Naosei <- questao12_Relacao_Naoeventualpessoasde_Simpedidas
questao2_Nao questao1_Sim idade_20 conf =100.0
```

Regra geral 5:

```
questao15_Sim <- questao16_Nao questao4_Sim questao5_Sim conf = 81.1
```

Exceção 5:

```
questao15_Naosei <- questao16_Nao questao4_Sim questao5_Sim idade_20 conf = 100.0
```

Os pares de regras 4 e 5 têm como conseqüente a resposta à questão 15 ("A União Estável é considerada pela legislação instituto capaz de formar família"). A partir desses conjuntos de regras é possível constatar que o fato de as pessoas terem por volta de 20 anos constitui uma situação de exceção entre o saber e o não saber sobre a questão proposta. As respostas às questões 4 e 16 estão corretas, e a questão 5 trata-se de declaração de ciência.

Regra geral 6:

```
questao16_Nao <- sexo_Masculino questao3_Sim questao4_Sim conf = 82.0
```

Exceções 6:

```
questao16_Sim <- sexo_Masculino questao3_Sim questao4_Sim idade_20 conf = 100.0
questao16_Naosei <- curso_Bacharelado sexo_Masculino questao3_Sim questao4_Sim
conf = 100.0
```

O conjunto de regras 6 tem como conseqüente a resposta à questão 16. Por meio desse conjunto de regras é possível novamente inferir que o fato de as pessoas terem por volta dos 20 anos estabelece uma situação de exceção entre o saber e o não saber sobre a questão proposta, bem como o fato de os respondentes serem do curso de Bacharelado em Informática. As respostas às questões 3 e 4 estão corretas.

Regra geral 7:

```
questao1_Sim <- questao8_Comunhaoparcial questao17_Sim questao4_Sim conf = 82.5
```

Exceções 7:

```
questao1_Nao <- questao8_Comunhaoparcial questao17_Sim questao4_Sim idade_24 conf = 100.0
```

```
questao1_Nao <- questao8_Comunhaoparcial curso_Engenharia questao17_Sim questao4_Sim conf = 100.0
```

O conjunto de regras 7 tem como conseqüente questão 1. Também a partir desse conjunto é possível perceber que a idade e a formação da área de Ciências Exatas acabam por determinar o desconhecimento dos entrevistados sobre a questão formulada.

Regra geral 8:

```
questao7_Nao sei <- questao1_Nao questao16_Nao conf = 85.7
```

Exceções 8:

```
questao7_Nao <- questao12_UniaoEstavel_NaopossuiprevisaoLegal sexo_Feminino questao1_Nao questao16_Nao conf = 100.0
```

```
questao7_Nao <- questao12_UniaoEstavel_NaopossuiprevisaoLegal curso_Computacao questao1_Nao questao16_Nao conf = 100.0
```

O conjunto de regras 8 tem como conseqüente a resposta à questão 7 ("Casais compostos de uma ou de duas pessoas com mais de 60 anos de idade podem escolher qual Regime de Bens incidirá sobre seu casamento?"). Mediante esse conjunto verifica-se que, apesar de em outras situações os entrevistados da área de Ciências Exatas estarem desinformados, nesse caso eles sabem da impossibilidade de pessoas de maiores idades escolherem o regime de bens. Também as mulheres demonstram ter clara essa questão.

7 CONCLUSÃO

A partir da pesquisa realizada com 110 entrevistados, bem como das informações e padrões descobertos, é possível perceber que muitas vezes as pessoas têm ciência de que não sabem e realmente demonstram desconhecimento sobre as diferenças entre as diversas formas de união conjugal, assim como a respeito de suas implicações sobre os efeitos patrimoniais decorrentes de cada uma delas. Todavia, a questão mais preocupante não se concentra sobre esse grupo, mas sim sobre as pessoas que dizem que sabem do assunto, porém em verdade o desconhecem. Ou seja, são essas pessoas as mais vulneráveis de envolver-se em situações de risco ou mesmo de conflitos.

Um dos méritos deste trabalho foi não apenas realizar um experimento sobre uma questão relevante, mas sobretudo demonstrar que o uso de técnicas mais elaboradas, como, por exemplo, descoberta de regras de associação e regras de exceção, permite potencializar o apoio oferecido aos especialistas da área/usuários/gestores.

A elevada quantidade de regras de associação comumente descobertas pelos algoritmos de *Data Mining* motivam o desenvolvimento de algoritmos de pós-processamento, com abordagem objetiva, capazes de analisar essas regras, identificando a existência de regras de exceção ou mesmo ranqueando as regras, de certa forma contribuindo com a análise do conhecimento produzido, tornando esse trabalho mais produtivo, ou mesmo viabilizando a sua utilização.

Uma sugestão de trabalho futuro seria a comparação do método testado neste estudo (descoberta de regras de exceção) com outros métodos que possuam como objetivo principal o pós-processamento do conhecimento descoberto na fase de *Data Mining* (CARVALHO, 2005), a fim de identificar as vantagens, desvantagens e possíveis otimizações de código, neste caso, visando a um ganho de performance no trabalho de avaliação dos padrões descobertos.

REFERÊNCIAS

BERRY, M. J. A.; LINOFF, G. **Data Mining Techniques**: for marketing, sales and customer support. John Willey & Sons, 1997.

BORGELT, C. Working Group Neural Networks and Fuzzy Systems, Department of knowledge Processing and Language Engineering. Otto-von-Guericke-University of Magdeburg, Alemanha. Disponível em: <<http://www-ics.cs.uni-magdeburg.de/iws.html>>. Acesso em: 05 jun. 2004

CARVALHO, D. R. **Árvore de decisão**: algoritmo genético para tratar o Problema do Pequeno Disjunto. 2005. Tese (Doutorado) – COPPE/UFRJ, Rio de Janeiro, 2005.

CARVALHO, D. R., MOLIN, V. Data Mining: gestão pedagógica de cursos de graduação Data Mining. **Cadernos de Educação da Unibrasil**, Curitiba, 2005. No prelo.

FAYYAD, U. M; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. **Advances in Knowledge Discovery and Data Mining**. American Association for Artificial Intelligence. 1996.

MEDEIROS, M. G. L. Novos arranjos familiares: inquietações sociológicas e dificuldades jurídicas. In: ENCONTRO DA ASSOCIAÇÃO BRASILEIRA DE ESTUDOS POPULACIONAIS, 13., Ouro Preto, 2002. **Anais**. Ouro Preto: ABEP, 2002.