

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**

Cesar B. Perizzolo

**SISTEMA DE IDENTIFICAÇÃO ESPACIAL EM
AMBIENTES DE CONSUMO UTILIZANDO
MINERAÇÃO DE DADOS**

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação

Professor Orientador: PAULO SERGIO DA SILVA BORGES, Dr.

Florianópolis, fevereiro de 2005

SISTEMA DE IDENTIFICAÇÃO ESPACIAL EM AMBIENTES DE CONSUMO UTILIZANDO MINERAÇÃO DE DADOS

Cesar B. Perizzolo

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação Área de Concentração Sistemas de Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Prof. Raul S. Wazlawick, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Paulo Sérgio da Silva Borges, Dr.
Orientador

Prof. Luiz Fernando Jacintho Maia, Dr.

Prof. Jovelino Falqueto, Dr.

Carlos Eduardo Negrão Bizzoto, Dr.

“Estamos afogados em informação, mas sedentos por conhecimento.”

(John Naisbett)

Dedicatória

Aos funcionários, professores e demais colaboradores do Curso de Pós-Graduação em Ciência da Computação, que viabilizaram meus estudos, e aos muitos colegas que fiz por aqui.

Especialmente ao meu orientador, Dr. Paulo Sérgio da Silva Borges, que me transmitiu conhecimento e segurança fundamentais para a realização deste trabalho.

Agradecimentos

A Deus, pela *vida* maravilhosa que possuo.

Aos meus pais, minha esposa Patrícia, familiares e outros amigos, pela *vida* maravilhosa que compartilhamos.

Ao meu orientador, Dr. Paulo Sérgio da Silva Borges, responsável direto por esta grande conquista em minha *vida*.

SUMÁRIO

1. INTRODUÇÃO	12
1.1 Justificativas	13
1.2 Relevância	19
1.3 Objetivo geral	21
1.4 Objetivos específicos	23
1.5 Limitações da pesquisa	23
1.6 A busca pelo conhecimento	24
1.7 Organização do trabalho	26
2. DADOS, INFORMAÇÕES E CONHECIMENTO	27
2.1 Os dados	30
2.2 A informação	31
2.3 O conhecimento	32
3. MINERAÇÃO DE DADOS	34
3.1 Sistemas de apoio à decisão (SAD)	41
3.2 As Etapas envolvidas na Mineração de Dados	41
3.2.1 Identificação do problema	42
3.2.2 Pré-processamento	44
3.2.3 Extração de padrões	45
3.2.4 Pós-processamento	46
3.3 Classes de Tarefas de Mineração de Dados	47
3.4 Técnicas de apoio à Mineração de Dados	48
3.4.1 Estatística	49
3.4.2 Data Warehouse	49
3.4.3 Aprendizado de máquina	49
3.4.4 Visualização de dados	50
4. O AMBIENTE DE CONSUMO	52
4.1 Os consumidores	55
4.2 Conceitos Elasticidade-Espaço	58
4.3 A fidelização	59
5. PROPOSTA DE TRABALHO	63
5.1 Introdução	63
5.2 Materiais e métodos	67
5.3 Etapas do projeto	68
5.4 O protótipo	69

6. CONCLUSÕES	79
6.1 Dificuldades encontradas	79
6.2 Sugestões para trabalhos futuros	80
GLOSSÁRIO	82
REFERÊNCIAS BIBLIOGRÁFICAS	83

LISTA DE QUADROS

Quadro 1. Diferenças entre Dado, Conhecimento e Informação	28
Quadro 2. Tipos de Conhecimento	33
Quadro 3. Algumas aplicações comuns para a Mineração de Dados	38
Quadro 4. Fontes de Informação	39
Quadro 5. Dados apresentados na forma tabular	50

LISTA DE FIGURAS

Figura 01. Deslocamento Distância Geométrica x Distância de <i>Manhattan</i>	13
Figura 02. Exemplo de leiaute interno de uma loja com auto-atendimento	22
Figura 03. Ilustração da transformação de dados em conhecimento	27
Figura 04. Algumas técnicas para a Mineração de Dados	34
Figura 05. Relações entre bases de dados, DW, MD e SAD	40
Figura 06. Tarefas de Mineração de Dados	43
Figura 07. Classes de tarefas de Mineração de Dados	48
Figura 08. Dados apresentados na forma gráfica	50
Figura 09. Representação da área útil existente no ambiente de consumo	65
Figura 10. Representação da área útil em unidades amostrais	66
Figura 11. Formulário de cadastramento de produtos	71
Figura 12. Formulário de aplicação de filtros para a seleção de vendas	74
Figura 13. Formulário de representação dos trajetos definidos pelo simulador	75
Figura 14. Gráfico de superfície (negativo)	76
Figura 15. Gráfico de superfície (acumulativo)	77
Figura 16. Gráfico de superfície (positivo)	78

RESUMO

A busca pelo conhecimento a partir da Mineração de Dados é um campo de pesquisa que tem crescido muito rapidamente e seu desenvolvimento tem sido dirigido, muitas vezes, em benefício de necessidades práticas.

Em um grande volume de dados podem estar muitas informações, tendências e padrões úteis para melhorar os critérios adotados durante o processo da tomada de decisões.

Neste trabalho propõe-se a transformação dos dados envolvidos durante uma venda, aliada à conversão dos espaços físicos ocupados pelos produtos em pequenas unidades amostrais, na tentativa de descrever o trajeto percorrido pelo potencial consumidor.

O desenvolvimento de processos de análise automática desses grandes volumes de dados é o desafio confiado aos algoritmos da Mineração de Dados.

O acúmulo da representação dos prováveis trajetos descritos por um determinado conjunto de consumidores será atribuído a uma matriz dinâmica de múltiplos atributos, que será apresentada através de gráficos de superfície.

Conclui-se que as técnicas e ferramentas para a Visualização de Dados estimulam naturalmente a percepção e a associação de novos padrões e podem melhorar, em muito, o processo de Mineração de Dados.

ABSTRACT

The seek for knowledge from Data Mining is a research field that has developed quite fast and its development has many times been aimed at practical need benefits.

A great volume of data may contain lots of information, tendencies and useful patterns for improving the adopted criteria during the decision taking process.

The objective of this work is the transformation of the involved data during a sales process, in connection with the conversion of physical spaces occupied by small sample units of the products, in order to describe the trajectory made by the potential customer.

The development of processes for automatically analyze those large data volumes is a challenge for the algorithms of Data Mining.

The accumulation of the representation of probable trajectories described by a determined group of customers is assigned to a dynamic matrix of multiple attributes, which is presented through surface graphics.

It was conclude that techniques and tools for Data Visualization naturally stimulate the perception and association of new patterns and may greatly improve the process of Data Mining.

1. INTRODUÇÃO

O presente trabalho tem por objetivo oferecer aos interessados uma forma de avaliação do tráfego em um ambiente de consumo sob o ponto de vista espacial.

Sua elaboração exigiu a conversão do espaço físico total, do comércio onde se encontram dispostos os produtos, em pequenas unidades amostrais. Essa conversão foi necessária para que cada produto pertencente ao cadastro obtivesse uma identificação espacial. Dependendo do tamanho de cada uma das unidades amostrais e também do espaço ocupado pelo produto, poderão, elas, agregar mais de um produto.

A indicação dos acessos de Entrada e Saída do espaço onde estarão armazenados os produtos permitirá a identificação dos pontos inicial e final, os quais representarão as localizações efetivas, quer do local por onde o proposto consumidor entrou no espaço físico destinado ao ambiente de consumo, quer do local por onde dele saiu.

Destarte, obter-se-á, inicialmente, um conjunto de registros de vendas através da aplicação de filtros de seleção e controle, podendo ter, cada um deles, um ou mais produtos relacionados, assim como em um carrinho de compras. Os critérios de seleção dos registros serão aplicados, de forma direta e interativa, pelo próprio usuário.

A partir do ponto identificado como sendo a entrada do estabelecimento, cada produto contido na venda será representado em sua forma espacial. Após a identificação de todos os pontos físicos obtidos do cadastro de produtos e tendo como ponto inicial a Entrada e, como ponto final, a Saída, será feita a ligação entre esses pontos na tentativa de apresentar um provável deslocamento do cliente no ambiente de consumo, desde a sua entrada até a saída, sem a consideração da existência de obstáculos.

Para encontrar a menor distância entre esses dois pontos, foi utilizada a Distância de *Manhattan* ou *City-Block*.

A distância de *Manhattan* é muito empregada para o cálculo de distâncias entre dois pontos no perímetro urbano das cidade onde, nesses casos, os resultados obtidos diretamente através de cálculos de distâncias geométricas entre as coordenadas (χ_1, γ_1) e (χ_2, γ_2) são ditos de baixa eficiência por não considerar a existência de quarteirões (Figura 01).

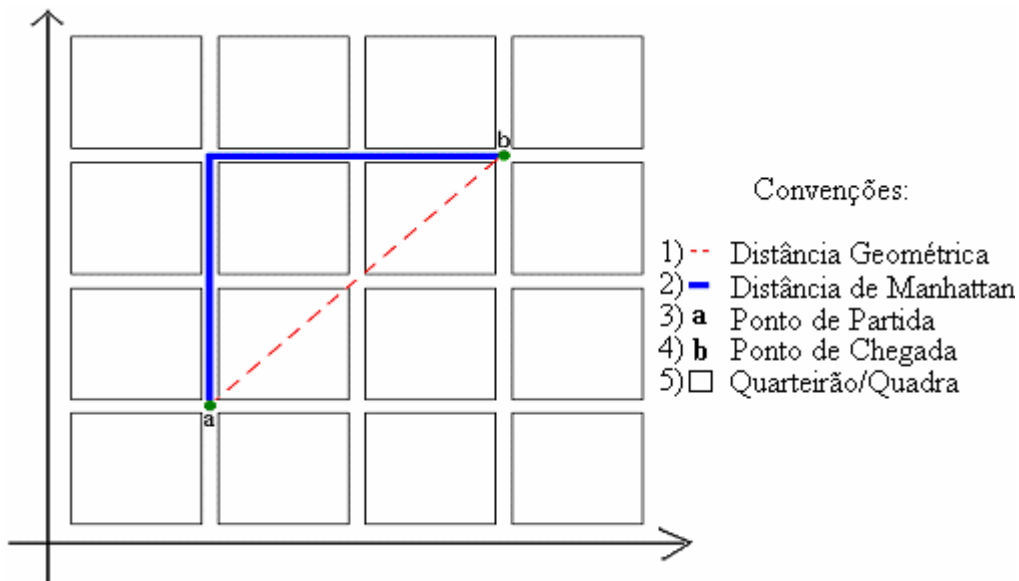


Figura 01 - Deslocamento Distância Geométrica x Distância de *Manhattan*.

A Distância de *Manhattan* mostra-se uma boa alternativa para a distância euclidiana quando os deslocamentos são efetuados em formatos já conhecidos, como nos realizados entre as prateleiras e os expositores existentes no interior de um ambiente de consumo, objeto da pesquisa realizada neste trabalho. Entende-se, por distância, o comprimento do segmento de reta que liga dois pontos.

Após a identificação do provável percurso descrito durante um registro de venda, é ele atribuído a uma matriz dinâmica, que receberá, de forma sobreposta, todos os outros percursos oriundos das demais vendas selecionadas na fase inicial.

Com o término da leitura de todas as vendas selecionadas, a matriz estará completa e poderá ser apresentada através de gráficos de superfície e também exportada para Sistemas de Apoio à Decisão (SAD)

1.1 Justificativas

Com o advento da concorrência globalizada, empresas buscam aumentar sua competitividade para assegurar uma fatia maior do mercado em disputa. A descoberta de informações relevantes, porventura ocultas em bancos de dados, é algo precioso e útil para aumentar essa fatia.

Com o aumento da capacidade de armazenamento e de processamento dos atuais sistemas de computação, um gigantesco número de dados pode ser acumulado durante o processo de aquisição e até mesmo durante a fase de processamento. FAYYAD *et al* (1996a) demonstram que os avanços nas tecnologias de armazenamento de dados, como velocidade no acesso, grande capacidade de armazenamento e barateamento nos dispositivos de armazenamento, têm possibilitado gerar grandes volumes de dados acessíveis eletronicamente. Mesmo assim, muitos desses dados não oferecem respostas claras e distintas para questionamentos considerados triviais no cotidiano das empresas e de seus administradores.

A busca exaustiva pela informação, a partir da manipulação dos dados armazenados, também pode significar um desgaste muito grande para o sistema e gerar um alto custo de produção. Segundo SINGH (2000), a gestão eficiente de informações representa atualmente um dos maiores desafios enfrentados pelas empresas.

Muitas das decisões tomadas pelos administradores de grandes empresas são perfeitamente representadas numa simples planilha de dados, mas sua compilação, organização e padronização nem sempre são automáticas, pois dependem de ajustes por parte dos usuários. Essas informações, contidas numa simples planilha ou, até mesmo, num gráfico, podem ser fundamentais para a sobrevivência de uma empresa.

Face ao fato, já citado, de que as aplicações que utilizam conceitos de Mineração de Dados visam à solução de problemas práticos, WITTEN & FRANK (2000) afirmam que as áreas de *marketing* e vendas são as de desenvolvimento mais ativo nas aplicações que utilizam técnicas de Mineração de Dados.

Segundo BERRY (1999), um especialista em comércio varejista e Diretor do Centro de Estudos sobre Varejo do Texas, A&M University - USA, as empresas varejistas do futuro competirão com base no valor criado para seus clientes e não mais pela concorrência baseada em preços, podendo ser a exposição dos produtos uma grande aliada na criação desses valores.

O comércio varejista ou, simplesmente, o varejo, é caracterizado pela exposição direta do produto ao consumidor final, sem a existência de intermediários.

KOTLER (2000, pág. 540) nos diz que varejo pode ser compreendido como sendo “qualquer atividade relacionada com a oferta de produtos ou serviços diretamente ao consumidor final, realizada através de uma loja de varejo, também conhecida como empreendimento varejista”.

Grandes varejistas já assimilam a idéia de que a concorrência simples e baseada em preços está demonstrando os primeiros sinais de desgaste. Surge assim a necessidade de desenvolver, para os clientes, um modelo de criação de valor que maximize os benefícios, minimize os custos e, mais, que esteja em constante mutação para não ser imitado por concorrentes.

Com o dinamismo exigido pelo varejo, surge a informática para oferecer o tempo de resposta adequado ao exigido pelo mercado. Conforme TAKAOKA (1999), a capacidade para agir rapidamente e decisivamente num mercado cada vez mais competitivo passou a ser um fator crítico de sucesso.

Práticas de *marketing* agressivo e identificação de padrões de comportamento entre os clientes, antes comuns e restritas aos grandes conglomerados varejistas, já estão sendo incorporadas às rotinas de varejistas de menor porte.

O *marketing* sempre foi utilizado como ferramenta indispensável na promoção de valores individuais, sendo essa identificação o maior desafio.

A célebre frase de Henry Ford: “Dêem-no a eles em qualquer cor, desde que seja preto”, em ANSOFF e MCDONELL (1993, pág. 26), reflete a grande mudança entre a preocupação existente com a indústria em relação aos seus clientes em épocas passadas e o esforço das indústrias automobilísticas atuais, que deixam à escolha do cliente até mesmo os acessórios que acompanharão seus veículos. Da mesma forma, o comércio também teve que realizar as mudanças exigidas pelos consumidores, agregando valor aos seus produtos através da prestação de serviços diferenciados. É comum o cliente ser questionado, pelo operador do *checkout*, se conseguiu encontrar todos os produtos que foram procurados. Caso exista algum produto que seja indicado pelo cliente, os comerciantes costumam viabilizar a compra, cientificando-o no momento da chegada do produto.

Essa postura estratégica dos varejistas, com relação às mudanças exigidas pelo consumidor, reflete o perfil adotado pela empresa frente aos seus concorrentes. FAHEY & RANDALL (1999) afirmam que a estratégia cria e alavanca para a mudança na empresa por meio da definição do escopo, da postura e das metas. Enquanto a postura reflete o comportamento adotado face aos concorrentes, a questão escopo diz respeito aos produtos pretendidos pela empresa, às necessidades dos clientes que serão atendidas pelo comerciante e à competência e à tecnologia que a empresa detém para suprir a demanda gerada pela relação produto-cliente.

Entre algumas dessas práticas de caráter estratégico está a do *enxoval*. Trata-se do subsídio, total ou parcial, de um determinado número de produtos do fornecedor para o varejista no momento da inauguração de uma nova loja. O tamanho desse subsídio está diretamente relacionado ao grau de importância do ponto no qual ficarão expostos os produtos daquele fornecedor no interior da loja. Uma gôndola expositora num local privilegiado no interior da loja poderá significar para o varejista várias semanas de abastecimento, de seus produtos, com custo reduzido ou nenhum custo para a sua compra e/ou também para o seu transporte.

A prática de disponibilizar promotores de vendas no interior da loja também é uma estratégia para fortalecer e incentivar a venda de um pré-determinado grupo de produtos. Esse esforço é empreendido para, entre outros fins, tornar pública a qualidade de um produto, melhorar a imagem da marca, revelar comparativos realizados com produtos concorrentes e, principalmente, estimular o desejo de posse ou o de consumo do produto anunciado. Usualmente, a permanência de um funcionário promotor de vendas ou até mesmo de um ponto-de-venda falante (*shelftalker*) deve ocorrer num espaço privilegiado sob a ótica da exposição e do número de visitantes.

O processo pelo qual é definido o nível de importância de um espaço no interior de uma loja ainda é feito de modo empírico, levando-se em consideração experiências anteriores. É difícil determinar o tamanho do sucesso de um ponto de vendas no interior de um estabelecimento comercial que sequer foi inaugurado.

A agilidade pretendida nas grandes redes varejistas para a adequação de seus leiautes exige a aplicação de sistemas computacionais para obter respostas e gerar subsídios capazes de permitir a otimização. As constantes mudanças na busca do melhor leiaute também são exigidas para não permitir que os concorrentes o imitem. Entende-se, por leiaute, o modo de distribuição e de apresentação dos produtos e serviços que estão sendo oferecidos aos clientes.

A identificação de novas oportunidades, assim como o sucesso de vendas de um novo produto ou de um novo estabelecimento comercial, é excessivamente dependente dos resultados obtidos através das estratégias adotadas.

No mercado varejista existem alguns produtos que agem como referenciais para determinar a *nobreza* de um espaço e são conhecidos como produtos *construtores de tráfego*. Esses produtos estão presentes na maioria das compras efetuadas e são utilizados de forma estratégica, ocupando espaços que são mais valorizados quando da prática do *enxoval*. Áreas vizinhas à utilizada pelos produtos construtores de tráfego são disputadas entre os fornecedores para aumentar o volume de vendas ou para o lançamento de um novo produto.

A competitividade entre empresas que atuam no mesmo segmento comercial tende a ser um fator agravante e exige uma identificação mais rápida das oportunidades.

A descoberta de novas oportunidades é tarefa abrangente e requer atitudes em vários níveis da organização. Para HARISON (1998), o processo de criação e implementação de decisões depende muito da informação disponível, da comunicação e da colaboração eficaz.

O empenho necessário para estabelecer novos conceitos é dependente dos recursos existentes, principalmente de recursos humanos. No próximo capítulo deste trabalho é citada a importância de contar com a compreensão dos colaboradores, nos mais diversos níveis hierárquicos, para assegurar o sucesso na implantação de novas medidas estratégicas.

Pesquisas realizadas pelo POPAI-BRASIL (Point of Purchase Advertising International) – organização internacional dedicada ao estudo e conhecimento das atividades varejistas – indicam que 85% das compras realizadas por consumidores brasileiros são decididas no ponto-de-venda, número que justifica os enormes esforços desenvolvidos para estimular o impulso de compra dos consumidores. (BLESSA, 2003).

Considerando-se os esforços realizados para tornar um produto mais competitivo frente aos demais existentes numa mesma loja, cabe observar que o espaço físico destinado à exposição do produto é de grande ou, até mesmo, de fundamental importância para o seu desempenho nas vendas.

Para um experiente profissional ligado ao setor varejista, um simples olhar sobre os produtos existentes no interior de um carrinho de compras já lhe seria suficiente para emitir alguns possíveis conceitos a respeito do cliente e de suas compras. Alguns especialistas definem como complementares os produtos que, habitualmente, seguem juntos na mesma compra. São exemplos de produtos complementares, num supermercado, relações como queijos e vinhos, carnes e carvão, massas e molhos, entre muitas outras. Aplicando-se regras de associação entre os produtos comprados, o próprio funcionário do *checkout* poderia ousar questionamentos com o cliente como: “Onde será o churrasco?”. A proposta deste trabalho é a aquisição de conhecimento obtido através da interpretação dos prováveis deslocamentos efetuados pelo consumidor durante a compra. Os trajetos descritos, no interior da loja, por um grupo de clientes podem esconder padrões de consumo, os quais seriam mais facilmente identificados através da Mineração de Dados (MD).

Segundo RUSSEL & KAMAKURA (1997), o comportamento do consumidor, em se tratando de produtos complementares e substitutos dentro da loja, pode ser monitorado por meio da análise das cestas de produtos adquiridas em cada compra da loja. Principal aliada nos Sistemas de Apoio à Decisão, a Mineração de Dados é o ramo da computação responsável pelo tratamento de questões relacionadas à descoberta de informações valiosas que podem estar inseridas entre grandes volumes de dados, assim como nos prováveis trajetos descritos por um suposto consumidor no interior de uma loja.

Para demonstrar a funcionalidade do modelo ora apresentado, foi desenvolvido um protótipo que utiliza algoritmos mineradores de dados, sem a existência de *Dataware House* (DW). A não exigência de um DW foi possível, dado o reduzido número de amostras e o alto nível de otimização do ambiente de simulação para a apresentação dos resultados, num formato meramente ilustrativo.

Em sua forma, serão exibidos gráficos capazes de representar os prováveis deslocamentos efetuados pelos consumidores no interior da loja.

1.2 Relevância

Os varejistas estão passando por grandes transformações para adequar as suas instalações e seus comportamentos frente aos consumidores. As atuais mudanças, especialmente no segmento lojista, estão causando grandes alterações no comportamento dos vendedores habituados a conduzir a venda do início ao fim.

Cada vez mais, o cliente está praticando o auto-atendimento, fazendo suas compras e dirigindo-se diretamente para o *checkout*, não permitindo que o vendedor exerça influência direta sobre suas compras. Não são poucos os supermercados que oferecem um setor dedicado à comercialização de eletro-eletrônicos, gerando a necessidade de pessoal de apoio com treinamento especializado para sanar dúvidas técnicas. Por contemplar uma grande variedade de produtos, como telefonia móvel, informática, eletrodomésticos e eletrônicos, os vendedores passaram a figurar como pessoal de apoio, sendo que cada um deles tem treinamento específico para um determinado grupo de produtos. Esta característica faz com que a prática do auto-atendimento seja desejável, deixando o cliente livre para escolher os produtos de sua preferência e facultando ao próprio cliente a possibilidade de realizar uma consulta técnica aos vendedores, que agora agem como consultores técnicos habilitados por segmentos de produtos.

Um comportamento mais crítico e seletivo, atualmente observado nos consumidores, está fazendo com que a idéia de oferecer concorrência baseada em preços seja superada por práticas mais envolventes. Segundo RAPHEL (1999), uma pesquisa realizada em todo o território norte-americano perguntou aos consumidores: “Por que você compra onde compra?” A resposta mais freqüente não foi “preço”.

A divulgação de produtos promocionais não tem o objetivo único de atrair novos consumidores. A manutenção da carteira de clientes existente passou a exercer grande importância, até mesmo, para os grandes supermercados.

O grande volume de vendas não permite aos supermercados identificar o cliente nas vendas à vista ou, mesmo, as vendas efetuadas através de cartão de crédito e outros convênios de terceiros. A falta da identificação do cliente no momento da venda não tem sido motivo para minimizar a manutenção dos clientes fiéis. Especialistas estimam que a maioria das empresas gasta cinco vezes mais com cada novo cliente que com os que já tem. A teoria do economista, sociólogo e engenheiro italiano, Vilfredo Pareto (1848-1923), sustenta que 80% dos negócios são realizados com 20% dos clientes.

Conhecendo os prováveis fluxos de deslocamentos, dos consumidores no interior da loja, o varejista poderá aplicar com maior eficiência as técnicas de valorização de seus clientes. A valorização do cliente é de suma importância para a sua fidelização. De acordo com RAPHEL (1999), um de cada cinco clientes deixará de comprar num determinado estabelecimento comercial, por problemas não resolvidos, pelos apelos da concorrência ou pela falta de uma dedicação especial. A identificação de áreas com elevado número de visitantes poderá fomentar a comercialização de outros produtos através de sua exposição em paralelo.

Para a solução dos problemas apresentados neste trabalho, é importante a utilização de um sistema de computação frente ao grande volume de dados envolvidos e à flexibilidade desejada para o simulador. A possibilidade de mudar um determinado produto de lugar é facilmente interpretada pelo sistema, possibilitando, inclusive, a Visualização de Dados históricos anteriores à mudança. De acordo com Piatetsky-SHAPIRO (1991), a leitura desses dados, produzidos e armazenados em larga escala, é inviável, já que é impossível que especialistas os consigam ler e analisar através de métodos manuais tradicionais.

A opção de apresentar os resultados utilizando técnicas de Visualização de Dados, através de gráficos de superfície, deu-se para oferecer uma maneira genérica de obter a compreensão e interpretação dos resultados. (LEE, 1995)

1.3 Objetivo Geral

O objetivo geral é desenvolver um conjunto de ferramentas de Mineração de Dados que possibilite a visualização e a identificação de padrões comportamentais baseados no estudo dos prováveis trajetos percorridos por consumidores em um ambiente de comercialização e de consumo onde os produtos são selecionados diretamente pelo consumidor.

As técnicas utilizadas, aqui descritas, poderiam ser aplicadas a diversos setores e ambientes de consumo, contudo, houve-se por bem priorizar o ambiente do comércio varejista para tornar mais simplificadas a demonstração e a apresentação dos resultados.

As respostas obtidas pelo modelo proposto neste trabalho apresentam muitas funcionalidades no interior de uma organização, não estando limitadas apenas ao *marketing* envolvido na exposição dos produtos. A utilização desses resultados, juntamente com a identificação comportamental de um grupo de clientes poderia oferecer muitos benefícios à organização empresarial.

A descoberta dos prováveis trajetos descritos por um determinado grupo de consumidores é obtida através do cálculo da menor distância entre dois pontos, que aqui são representados pelos produtos que compõem o seu carrinho de compras. Para estabelecer a menor distância, foi utilizada a distância de *Manhattan*, pelo motivo já apresentado e que pode ser mais bem observado nas Figuras 01 e 02.

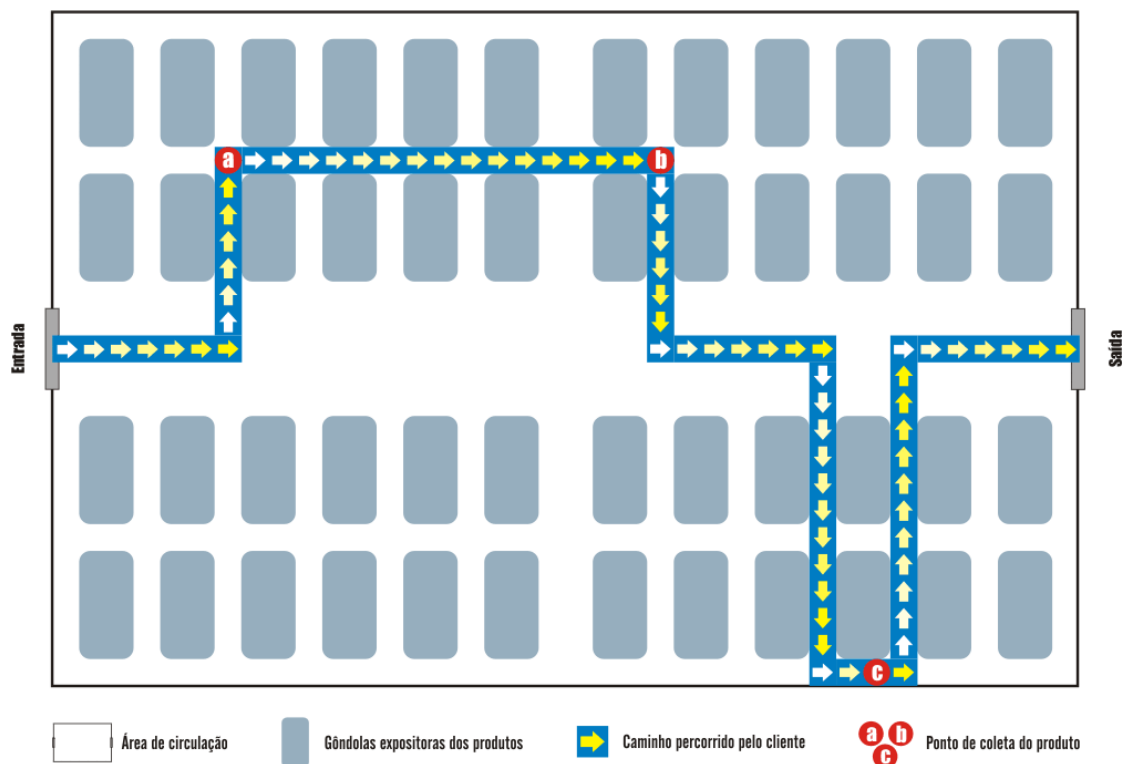


Figura 02 - Exemplo de leiaute interno de uma loja com auto-atendimento.

Sabe-se que os caminhos encontrados pelo simulador computacional, construído neste trabalho para a realização de testes, podem não ser fiéis aos reais caminhos descritos, mas, espera-se que, após a observação de um grande volume de informações, os resultados obtidos representem uma grande parte dos efetivos deslocamentos realizados.

O conhecimento dos trajetos mais praticados e dos pontos mais visitados poderá ser utilizado de várias formas e em vários departamentos da empresa analisada.

Um formato mais eficiente na exposição de produtos poderá tornar mais simples o processo de compra pelos clientes e também possibilitará um aproveitamento melhor das instalações.

Ganhos operacionais provenientes da gestão eficiente de um estabelecimento comercial poderão ser convertidos em vantagens para os clientes e gerar diferenciais competitivos frente aos concorrentes.

Assim o objetivo geral é o desenvolvimento de um protótipo que permita a visualização dos prováveis trajetos descritos por um grupo de clientes no interior de um ambiente de consumo.

1.4 Objetivos específicos

1. Apresentar gráficos de superfície com os resultados alcançados;
2. Permitir a identificação de possíveis padrões de comportamento entre os consumidores;
3. Facilitar o processo de exposição dos produtos no interior da loja;
4. Estimular a venda de produtos através de sua exposição adequada;
5. Gerar subsídios capazes de elevar a competitividade da empresa frente aos concorrentes.

1.5 Limitações da Pesquisa

O presente trabalho buscou identificar pontos de visitação dos clientes no interior de um ambiente de consumo. Para tal, exigiu-se a conversão do espaço físico total, em unidades amostrais. A transformação em células amostrais poderá indicar divergências de precisão em relação ao espaço físico instalado, sendo esta imprecisão diretamente proporcional ao tamanho e à unidade utilizados na célula.

Com relação aos prováveis deslocamentos, é válido salientar que o presente modelo não observou a existência de eventuais obstáculos que poderiam impedir o deslocamento assim como no proposto pelo simulador. O modelo proposto neste trabalho considerou a menor distância obtida pela *Distância de Manhattan* como sendo o único estímulo para definir o próximo ponto a ser visitado.

1.6 A busca pelo Conhecimento

O conhecimento e a informação sempre estiveram muito próximos nos trabalhos de pesquisa, porém, separados em suas definições. Existem, ainda, as várias áreas de pesquisa e desenvolvimento, onde cada uma delas tenta adequar-se às definições que mais refletem suas realidades.

Tratada por abordagens matemáticas quando objeto das ciências exatas, a informação já havia recebido interpretações filosóficas quando trabalhada pelas ciências humanas e sociais e pela própria filosofia.

SIRIHAL & LOURENÇO (2003) explicam que, assim como os conceitos são relacionados aos dados, informação e conhecimento também recebem atribuições de acordo com a linha de pesquisa dos trabalhos, criando diferentes definições entre o “conhecimento matemático” e o “conhecimento”, como se o próprio conceito da palavra “conhecimento” pudesse ser alterado em função da sua contextualização.

Para CAPURRO (1992):

O ponto de vista positivista no estudo da informação é sustentado por três paradigmas, a saber: o paradigma da representação, o paradigma emissor-canal-receptor e o paradigma platônico.

De acordo com o paradigma da representação, os seres humanos são conhecedores ou observadores de uma realidade externa. O processo de conhecimento consiste na assimilação das coisas através de suas representações na mente/cérebro do sujeito. Essas representações, uma vez processadas ou codificadas em nossos cérebros, podem então ser comunicadas a outras mentes e/ou armazenadas e processadas em máquinas (computadores). Seres humanos são processadores de informação biológicos. Informação é a réplica codificada da realidade. Seres humanos podem utilizar informação para propósitos específicos e racionais, mas nada se fala contra a hipótese de que as máquinas também possam alcançar esse nível de processamento e uso da informação. De acordo com esse paradigma, a Ciência da Informação preocupa-se com o estudo da representação, codificação e uso racional da informação.

O paradigma emissor-canal-receptor trata o fenômeno da comunicação humana como uma metáfora a ser aplicada em diferentes níveis da realidade. Ao comunicar-se, emissores e receptores trocam informação. Para que o receptor compreenda o significado da mensagem enviada pelo emissor, é necessário que exista um “estoque” de signos. De acordo com este paradigma, a Ciência da Informação preocupa-se principalmente com o impacto da informação do receptor. Ao mesmo tempo, receptores são “buscadores” ou usuários de informação com o objetivo de resolver seus problemas.

O paradigma Platônico apresenta um ponto de vista oposto aos dois anteriores: ao invés de considerar a existência de um sujeito detentor do conhecimento, considera a *informação por si mesma*. Pode ser materialista (o conhecimento não é um processo biológico, sociológico ou psicológico, mas é objetivado em “transportadores” não-humanos) ou idealista (o conhecimento é algo objetivo em si mesmo, independentemente de qualquer “transportador” material).

A citação de algumas diferentes convenções a respeito do conhecimento pode tornar mais simples a tarefa de compreender a transformação dos *dados* em *informação* e desta em *conhecimento*. Pode-se, também, identificar a preocupação existente com a portabilidade de informações através de máquinas, não restringindo o seu uso aos seres humanos apenas.

Em sistemas computacionais, os *dados* têm sido a forma mais primitiva e o fato gerador do *conhecimento*. Desconsiderada a forma binária utilizada pelos computadores, o *dado* pode ser interpretado como o menor fragmento de *informação* junto a um sistema. Essa realidade computacional é o fator que realiza e fortalece a relação entre o fato gerador (*dado*) e sua forma de armazenamento e manipulação, tarefas últimas destinadas aos SGBD – Sistemas Gerenciadores de Bancos de Dados.

Muito populares, os SGBD's são conhecidos pela eficiência em armazenar e recuperar dados. Mesmo com a evolução da computação, que possibilitou um grande incremento nas capacidades de processamento e armazenamento, a interpretação e a busca do conhecimento a partir da análise destes *dados* ainda geram demanda de tempos de resposta superiores aos desejáveis pelas organizações empresariais.

Atualmente, os analistas de negócios necessitam de respostas para questionamentos mais complexos como: “identifique o grupo de consumidores potenciais para determinado produto e que tenha filhos em idade escolar”. Respostas para este tipo de questionamento são essenciais para a manutenção e para a sobrevivência da empresa num mercado competitivo e dinâmico como o atual. Para suprir esta demanda, gerada pela busca de informações potenciais existentes nos Bancos de Dados, surgiram os algoritmos mineradores.

A Mineração de Dados pode ocorrer, ainda, fora do ambiente de banco de dados relacionais, como mineração de textos, *data warehouses*, arquivos de *logs* e outras fontes.

Para que este trabalho possa estabelecer referências espaciais e utilizar-se das técnicas de Mineração de Dados, propõe-se a geração de identidade espacial para cada um dos produtos pertencentes aos estoques gerais. Além desta característica, a busca pelo provável trajeto, o acúmulo de trajetos e a observação de uma matriz dinâmica são partes integrantes do conhecimento adquirido.

A forma de apresentação final, através de gráficos de superfície, é uma das tecnologias de suporte à Mineração de Dados, conhecida como Visualização de Dados, e que será mais bem apresentada no capítulo 3.

1.7 Organização do Trabalho

Esta dissertação está disposta em 6 capítulos, a seguir:

No Capítulo 1 faz-se a introdução do trabalho, demonstrando suas justificativas, objetivos e limitações. Ainda nesse capítulo é citada a busca pelo conhecimento e também é realizada uma breve abordagem dos assuntos envolvidos na realização deste trabalho.

No Capítulo 2, faz-se breve descrição das diferenças existentes entre dados, informação e conhecimento (Capítulo 2). Mostram-se, ainda, algumas diferenças conceituais quando o assunto é abordado em ramos distintos da ciência.

No Capítulo 3, apresenta-se uma revisão teórica das principais etapas envolvidas no processo de Mineração de Dados e na utilização dos recursos por outros Sistemas de Apoio à Decisão.

No Capítulo 4, realiza-se a avaliação do Ambiente de Consumo e se inserem algumas características e particularidades envolvidas na ampla atividade do comércio varejista.

O Capítulo 5, além de tratar da proposta principal apresentada neste trabalho, aborda também as técnicas utilizadas, suas justificativas e a funcionalidade do protótipo desenvolvido.

No Capítulo 6 serão apresentadas as conclusões obtidas, as dificuldades encontradas e também as sugestões para trabalhos futuros.

Ao final, estão o glossário e as referências bibliográficas utilizadas no desenvolvimento do trabalho.

2. DADOS, INFORMAÇÃO E CONHECIMENTO

A contextualização da *informação*, assunto já abordado na introdução deste trabalho, pode conduzir à interpretação de que *dados*, quando tratados pela Ciência da Computação, podem ser tão somente um fragmento da informação. SETZER (2001, p.242) diz que não é possível processar informação diretamente em um computador. Para isso é necessário reduzi-la a dados.

Considerando as divergências existentes entre os conceitos específicos de cada uma das atividades no tratamento das informações, DAVENPORT (1998) nos diz que o fascínio pela tecnologia fez esquecer o principal objetivo da informação: informar. No quadro 1, o mesmo autor estabelece relações entre os termos *dado*, *informação* e *conhecimento*. A figura 03 ilustra a transformação de dados em conhecimento.

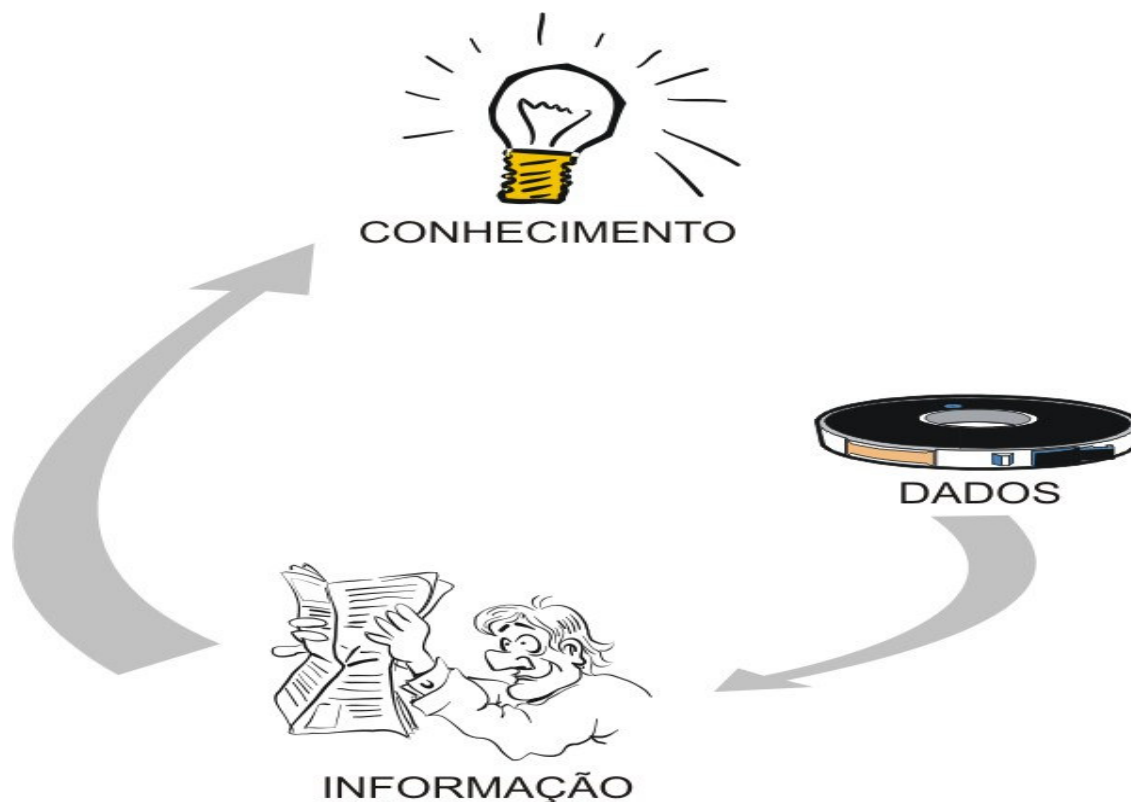


Figura 03 - Ilustração da transformação de dados em conhecimento.

Quadro 1 - Diferenças entre Dado, Informação e Conhecimento. Fonte: Davenport (1998)

Dado	Informação	Conhecimento
Simples observações sobre o estado do mundo. -Facilmente estruturado -Facilmente obtido por máquinas -Freqüentemente quantificado -Facilmente transferível	Dados dotados de relevância e de propósito. -Requer unidade de análise -Exige consenso em relação ao significado -Exige, necessariamente, a mediação humana	Informação valiosa da mente humana. -De difícil estruturação -De difícil captura em máquinas -Freqüentemente tácito -De difícil transferência

O motivo para o estabelecimento das diferentes interpretações apresentadas entre as noções de *dados*, *informação* e *conhecimento* foi o fato de que a base de consulta utilizada na construção desta dissertação é oriunda dos estudos realizados, em sua maioria, em ciências exatas.

As definições encontradas ao longo desta pesquisa poderão não exprimir, de maneira universal, os conceitos existentes na tríade, *dados*, *informação* e *conhecimento*, porém, saber-se-á que o imperativo tecnológico deste trabalho utilizar-se-á da Ciência da Computação e de suas vertentes para descrever os principais conceitos envolvidos e abordados no discorrer desta dissertação.

Os avanços tecnológicos no campo da eletrônica digital estão permitindo, cada vez mais, incrementos na capacidade de armazenamento de grandes volumes de dados. Este acréscimo oferece aos analistas e programadores recursos para salvar dados que antes eram considerados descabidos frente às limitações e aos altos custos impostos pela escassez de recursos de armazenamento.

Os dados produzidos por grandes organizações, freqüentemente, precisam ser analisados por um especialista para propiciar a obtenção de respostas satisfatórias. Ainda assim, a velocidade com que os dados são armazenados é muito superior à velocidade de seleção e descoberta de padrões pelos analistas de negócios.

A Mineração de Dados aparece justamente como uma ferramenta capaz de suprir a demanda imposta pela descoberta de informações *escondidas* em grandes volumes de dados e facilitar a obtenção do conhecimento de maneira rápida e segura a partir de uma fonte de dados.

Com a flexibilidade oferecida pelos sistemas que fornecem suporte à Mineração de Dados, cada um dos usuários poderá realizar sua consulta de forma rápida e personalizada, sem o condicionamento restrito ao uso de um único relatório ou formulário utilizado por todos os demais colaboradores.

A grande velocidade de processamento e a identificação da importância existente no conhecimento estão oferecendo à informática um novo rumo, o de uma verdadeira Tecnologia da Informação.

De acordo com as definições encontradas em alguns dicionários da Língua Portuguesa, poucas são as diferenças encontradas entre *dados*, *informação* e *conhecimento*.

De maneira generalizada, pode-se citar que *dados* são puramente sintáticos e sem um cenário contextual, tornam-se algo muito pouco ou quase nada expressivo.

A *informação*, por sua vez, é algo que requer um emissor e um receptor. Pode-se dizer que a *informação* é a representação da relação entre os *dados*.

O *conhecimento* é algo mais difícil de ser descrito por dizer respeito a algo interior e também necessitar de algum tipo de abstração pessoal. Em alguns campos pesquisados em sistemas de computação, esse ato ou efeito de conhecer é obtido através da aplicação de um conjunto de regras.

De acordo com OLIVEIRA (1992), informação é o dado trabalhado que permite ao executivo tomar decisões, enquanto dado é qualquer elemento identificado em sua forma bruta que por si só não conduz a uma compreensão de determinado fato ou situação.

O aumento na relevância existente durante a interpretação dos dados está contribuindo para o desenvolvimento de novos conceitos relacionados à Tecnologia da Informação (TI), trazendo muitas mudanças nas grandes corporações. Cada vez mais o profissional de TI está evoluindo para entender melhor o negócio da empresa. A busca pela compreensão de todas as etapas envolvidas na administração dos negócios de uma empresa está permitindo algo muito incomum em tempos passados: o profissional de TI transformado em presidente da empresa.

Essa transformação, com relação ao profissional de Tecnologia da Informação, demonstra claramente a preocupação das grandes corporações empresarias, com a busca do conhecimento através da Mineração de Dados e, também, a importância do conteúdo encontrado para toda a cadeia produtiva empenhada na obtenção de resultados.

Empresas investem para oferecer a seus colaboradores uma compreensão da instituição como um todo, assim como aqui é demonstrado, através da correta disposição de um produto na prateleira. Esse trabalho, efetuado pelo funcionário responsável pela reposição das mercadorias, também pode ser considerado um dado importante na avaliação do desempenho da empresa.

Ao adquirir um produto, o consumidor está determinando o fim de um ciclo de trabalhos que teve início no planejamento da compra junto ao fornecedor até a chegada do produto ao carrinho desse consumidor. Os estágios envolvidos entre estes dois ciclos são representados por um esforço realizado por toda uma equipe de profissionais que planejou desde a compra até a efetiva venda do produto.

2.1 Os Dados

Um conjunto de dados pode ser entendido como uma seqüência de símbolos, um som, figuras e animações, entre outros sentidos. Desta maneira pode-se observar uma seqüência de letras, que representam uma palavra, como sendo um dado. A simples observação de um estado também pode ser considerado um dado.

Os dados podem ser obtidos, com facilidade, a partir de um universo de informações, pois é fácil transferi-los para os sistemas de computação. Um sensor de temperatura, por exemplo, pode obter *dados* a respeito de uma caldeira a cada instante. Os números acusados pelo sensor são facilmente armazenáveis e, com a mesma facilidade, podem ser transferidos para um banco de dados.

A estruturação dos dados também é muito simples, pois o universo dos possíveis valores já deve ser conhecido no momento do projeto do banco responsável pelo armazenamento e os resultados obtidos devem estar dentro da faixa suportada e implementada pelo projetista.

Uma seqüência de leituras de um sensor, por si só, não representaria uma grande informação por não trazer outros dados, nem as ligações entre eles. Esta contextualização dos dados será o fato gerador da informação.

Várias são as classificações existentes quanto aos tipos de dados existentes. Estas classificações podem trazer referências relativas à fonte, origem, estado, qualidade, armazenamento e outras.

2.2 A Informação

A *informação* deve representar a associação entre um conjunto de dados. Esta associação deve exprimir a real relação entre eles, sua relevância e finalidade. A caracterização desta relação entre os dados é, para o sistema, algo muito mais difícil do que a sua aquisição, pois esta poderia ser automática, como no exemplo anterior, através de um sensor de temperatura. Esta dificuldade requer uma análise e, conseqüentemente, um considerável esforço por parte do responsável pela criação de um cenário em que os dados passam a manter uma relação entre si.

A representação da informação também exige outros cuidados, como a integridade entre as relações estabelecidas. A ausência de um dado num conjunto pode torná-lo descartável para o sistema.

Imagine-se a seguinte seqüência: *equipamento, temperatura, pressão, data e hora*. A perda de qualquer um desses dados, mesmo sendo eles de fácil aquisição por sensores automatizados, poderá significar o comprometimento da informação como um todo. A informação obtida através da representação desse conjunto de dados pode ser o nível de perigo oferecido por um equipamento a um grupo de trabalhadores no interior de uma fábrica.

Informações quantificadas, como as propostas no parágrafo anterior, possuem um tratamento mais simples, já que representam diretamente um dado dentro de uma variação escalar.

Uma frase que possua dados qualitativos pode representar melhor a complexidade da extração da informação. A frase, “*Minha Brasília é muito econômica*”, somente poderá transmitir informação ao leitor caso este saiba que a Brasília, à qual a frase se refere, é um automóvel e também se o leitor tiver algum tipo de conhecimento e referência capaz de traduzir o termo “*econômica*”. A frase anterior poderia ser facilmente armazenada no formato de texto em um banco de dados, mas dificilmente poderia ser considerada uma informação. O processamento da frase por alguém que tivesse o conhecimento necessário poderia transformar esse *dado* em *informação*.

2.3 Conhecimento

A transformação da *informação* obtida através dos *dados* em *conhecimento* é algo que exige do interpretador um grau de abstração ainda maior que o necessário para transformar *dados* em *informação*.

Essa transformação depende de inúmeros fatores, neles incluído até mesmo o estado emocional do avaliador. As características individuais podem fazer com que o conhecimento tome, por exemplo, dois rumos distintos, quando as mesmas informações são avaliadas por pessoas diferentes.

O conhecimento depende, ainda, do próprio nível de conhecimento do avaliador. Alguns autores estabelecem a condição de *informação* ao seu autor, admitindo que o conhecimento seja algo exclusivo daquele que apurou as informações necessárias para torná-las *conhecimento*.

A hipótese acima é perfeitamente cabível se admitida, por exemplo, a leitura de um jornal por duas pessoas. Certamente os conhecimentos adquiridos durante as leituras serão igualmente distintos, porque a *bagagem* que cada um dos indivíduos possuía antes da leitura deverá fazer com que os conteúdos sejam absorvidos de maneiras diferentes. Os próprios conceitos existentes antes da leitura poderão ser alterados após o seu término.

Citando o conhecimento transmitido através da comunicação verbal, a informação recebida pelo ouvinte poderá ser influenciada pelo seu humor, pela sua disponibilidade de tempo, suas experiências anteriores e, dependendo do assunto, até mesmo de sua condição emocional. Além das condições do ouvinte, as condições do responsável pela fala também podem exercer algum tipo de influência sobre o conteúdo do que está sendo transmitido.

MIRANDA (1999) diz que existem 3 diferentes tipos de conhecimento, conforme se lê, a seguir, no quadro 2.

Quadro 2 - Tipos de conhecimento. Fonte: MIRANDA (1999)

Conhecimento explícito	É o conjunto de informações já elucidadas em algum suporte que caracteriza o saber disponível sobre um tema específico.
Conhecimento tácito	É o acúmulo do saber prático sobre um determinado assunto, que agrega convicções, crenças, sentimentos, emoções e outros fatores ligados à experiência e à personalidade de quem o detém.
Conhecimento estratégico	É a combinação dos conhecimentos explícito e tácito, agregada ao conhecimento de especialistas.

As características implícitas na transformação dos dados brutos em conhecimento tornam-se grandes desafios aos sistemas de informação. As dificuldades para esta transformação ainda permanecem como imensas barreiras para os sistemas computacionais.

Mesmo a própria comunicação através da fala, entre pessoas de mesma língua materna, é capaz de provocar equívocos e problemas de interpretação. Estes equívocos podem ser ocasionados, entre outros fatores, por diferenças existentes entre os conhecimentos específicos de cada um dos integrantes da comunicação. Além das diferenças existentes entre os conhecimentos específicos do transmissor e do receptor, é prudente considerar que os próprios ânimos envolvidos durante a comunicação poderiam alterar, ainda mais, o conhecimento envolvido em uma simples conversa.

Referindo-se a tríade, *informação, dado e conhecimento*, DAVENPORT & PRUSAK (1999, pág. 17) dizem que “informação, além do mais, é um termo que envolve todos os três, além de servir como conexão entre os dados brutos e o conhecimento que se pode eventualmente obter.”

A seguir, serão apresentadas as principais etapas envolvidas ao longo da Mineração de Dados (Figura 04).

3. MINERAÇÃO DE DADOS

GROTH (1998) define a Mineração de Dados como sendo “o processo que representa a Automação aplicada na aquisição de conhecimento” e descreve que uma analogia com o processo de Mineração de Dados poderia ser a busca de um cão de estimação perdido às duas horas da manhã: pode-se usar uma lanterna, mas seria muito mais fácil caso existissem sensores capazes de descrever todos os movimentos efetuados pelo cão dentro de uma superfície na área de busca. A Figura 04 apresenta algumas técnicas utilizadas em Mineração de Dados.



Figura 04 - Algumas técnicas para a Mineração de Dados. Fonte: VIANA (2004).

Segundo FAYYAD *et al.* (1996a), Mineração de Dados ou “Extração de Conhecimento de Base de Dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados”.

O processo de busca de padrões úteis escondidos nos dados recebe algumas denominações, como *extração de conhecimento*, *descoberta de informações* e *arqueologia dos dados*.

A busca do conhecimento extraído de grandes bases de dados tem crescido rapidamente em função das necessidades práticas.

Em TURBAN *et al.* (2003, pág. 153), lê-se que “o acesso a informações exatas e atualizadas, necessárias ao gerenciamento de operações diárias e ao planejamento estratégico de longo prazo, tem-se tornado cada vez mais importante para o atual mercado de trabalho global”.

A popularização de Sistemas Gerenciadores de Banco de Dados, juntamente com a evolução do *hardware* e dos aplicativos, permitiu a manutenção de uma infinidade de dados isolados que, através da Mineração de Dados, podem ser submetidos a uma busca de valiosas informações ali contidas e de vital importância para a tomada de decisões.

De posse desses grandes volumes de dados, sendo muitos adquiridos através de transações operacionais exercidas por aplicativos diversos, as empresas percebem as limitações impostas pela simples utilização de relatórios e outras telas de consultas desenvolvidas na forma tradicional, sem a utilização de ferramentas capazes de conduzir e auxiliar o processo decisório.

Para CHIAVENATO (1999, pág.288), o processo decisório está compreendido em seis etapas:

- identificar a situação,
- obter informação,
- gerar soluções alternativas,
- avaliar e escolher a melhor alternativa,
- transformação da solução em ação,
- avaliar os resultados obtidos.

A elevação na capacidade de armazenamento dos dados, resultante do desenvolvimento das tecnologias do hardware, também elevou o grau de complexidade no momento de extrair as informações destes depósitos de dados. O aumento desta capacidade está gerando uma sobrecarga de dados para os sistemas, tornando ainda mais difícil a sua interpretação pelo usuário. “Dados armazenados não valem nada, mas sim o conhecimento que pode ser extraído deles. Os métodos tradicionais de análise estão tornando-se lentos e ineficazes diante de grandes volumes” (FAYYAD & UTHURUSAMY, 1996).

Linguagens estruturadas, como o SQL, ditas de fácil compreensão até mesmo para profissionais não relacionados às áreas de TI, ainda são obstáculos para executivos que não dispõem do tempo necessário para entender o funcionamento dos comandos utilizados na sua manipulação.

Para aproximar os dados armazenados de um formato mais adequado e produtivo para os vários interessados no interior de uma organização, surgiu um ambiente chamado *On Line Analytical Processing* (OLAP) capaz de exercer as transformações necessárias e armazenar os dados de uma maneira mais eficiente e dirigida para customizar e acelerar as análises previstas.

Com o processamento efetuado pelas ferramentas OLAP, o depósito de dados passará a atender a todas as demandas previstas na fase de planejamento e os dados estarão compilados nos formatos mais adequados para a sua recuperação, a qual poderá ser feita de forma direta e sustentável pelo próprio usuário.

Com o tratamento de um OLAP, as informações podem ser analisadas mais facilmente, num formato sintético ou analítico, cabendo ao usuário final a decisão.

A busca e a identificação de padrões e tendências obtidas na compilação de dados são características desejáveis de uma boa ferramenta que incorpore os conceitos de Mineração de Dados e esta busca poderá ser alcançada, com maior eficiência, quando os dados já estiverem armazenados num formato otimizado para a sua interpretação.

No interior das empresas, tomadores de decisões seguem em busca de respostas para situações ainda não contempladas, mas compatíveis com padrões existentes em dados históricos que podem ser obtidos na forma de conhecimento quando submetidos a um bom algoritmo Minerador de Dados.

Para que os dados estejam centralizados numa única base, é necessário que haja a adoção de critérios de organização no armazenamento e na classificação. Este tipo de classificação é necessária durante a realização do *Data Warehousing* responsável pela seleção e limpeza dos dados que, após a sua classificação, podem ser apresentados de maneira mais clara e objetiva, facilitando a obtenção e a extração do conhecimento. Para tal, um repositório de dados é criado, o *Data Warehouse*, que é direcionado para as consultas.

Esse campo de pesquisa, responsável pela interpretação de grandes volumes de dados, é chamado de Extração de Conhecimento de Base de Dados, largamente descrito como *Knowledge Discovery In Database* (KDD).

O termo “*Data Mining*”, aqui tratado como Mineração de Dados (MD), é mais comumente utilizado entre os profissionais da estatística, enquanto o termo KDD foi utilizado em 1989 para descrever o amplo processo da aquisição de conhecimento nos dados e também para enfatizar o alto nível das aplicações utilizadas na Mineração de Dados. Atualmente, o termo KDD é mais utilizado em pesquisas que incorporem técnicas de *inteligência artificial e aprendizado de máquina*.

A Mineração de Dados também oferece opções de busca do conhecimento sem a necessidade de implementação do *Data Warehouse* (DW), sendo, o próprio algoritmo de mineração, responsável pela classificação e limpeza dos dados obtidos diretamente na base de dados. Esta técnica, que suprime a utilização do DW, é indicada somente nos casos em que não haja falta de recursos em momento de execução e onde o tempo de resposta não seja um fator crítico, pois a classificação em tempo de execução exige maior processamento e, conseqüentemente, maior tempo de resposta. Segundo INMON (1996), para o processo de MD não é necessário ter implementado um DW.

A seguir, no quadro 3, revelam-se alguns segmentos considerados comuns, mas que se prestam para um grande número de aplicações que incorporam conceitos de Mineração de Dados.

Quadro 3 – Algumas aplicações comuns para a Mineração de Dados

Aplicações Comuns da Mineração de Dados	
Segmentação do mercado	Identifica as características comuns dos clientes que compram os mesmos produtos em uma empresa.
Desistência de clientes	Faz uma previsão dos prováveis clientes que sairão de uma empresa e passarão para um concorrente.
Detecção de fraudes	Identifica as transações com maior probabilidade de serem fraudadas.
Marketing direto	Identifica as pessoas interessadas que devem ser incluídas em uma lista de correspondência, para se obter o mais alto índice de respostas.
Análise de tendências	Revela a diferença entre o comportamento de um cliente comum no mês atual e no mês anterior
Ciência	Simula explosões nucleares e visualiza a física quântica.
Entretenimento	Modela fluxos de clientes em parques temáticos e analisa a segurança em parques de diversões.
Seguro e sistema de saúde	Faz uma previsão de quais clientes comprarão novas apólices, identifica padrões de comportamento que aumentam o risco da seguradora e detecta reclamações fraudulentas.
Produção	Otimiza o projeto de produtos, equilibrando a possibilidade de produção e a segurança, melhora o escalonamento da produção e a utilização das máquinas.
Medicina	Indica terapias adequadas para diferentes doenças, prevê a eficácia de drogas e descobre novas drogas e tratamentos.
Óleo e Gás	Analisa dados sísmicos, em busca de sinais de depósitos subterrâneos, prioriza os locais de perfuração e simula fluxos subterrâneos para melhorar a recuperação.
Varejo	Distingue padrões de comportamento de compras e prevê como os clientes reagirão a campanhas de marketing.
Transporte	Otimiza as escalas de distribuição e uso de veículos e analisa os padrões de carregamento de caminhões e vagões ferroviários.

Muitas são as aplicações desenvolvidas que incorporam protótipos de ferramentas de Mineração de Dados assim como o protótipo proposto durante a realização deste trabalho. Quer seja no ambiente acadêmico ou mesmo através de ferramentas comerciais, a MD está encontrando o espaço pretendido pela sua proposta que busca resolver problemas comuns em muitas atividades humanas, conforme consta no Quadro 4, a seguir.

Quadro 4 - Fontes de informação. Fonte: REZENDE (2003)

Nome	Técnicas Disponíveis	Fabricante	Site
PolyAnalyst	Classificação, regressão, regras de associação, clustering, sumarização e modelagem de dependência.	Megaputer Intelligence	www.megaputer.com
Magnum Opus	Regras de associação.	Rule Quest	www.rulequest.com
XpertRule Miner	Classificação, regras de associação e clustering.	Attar Software Ltd.	www.attar.com
DataMite	Regras de associação.	Dr. Phillip Vasey	www.lpa.co.uk/dtm.htm
MS Data Analyzer 2002	Classificação e clustering.	Microsoft Corp.	www.microsoft.com
Oracle 9i Data Mining	Classificação e regras de associação.	Oracle Corp.	www.oracle.com
Darwin	Classificação, regressão e clustering.	Oracle Corp.	www.oracle.com
MineSet	Classificação, regressão, regras de associação e clustering.	Silicon Graphics Inc.	www.sgi.com
WEKA	Classificação, regressão e regras de associação.	University of Waikato	www.cs.waikato.ac.nz
Intelligent Miner	Regras de associação, padrões sequenciais, classificação, clustering, sumarização e modelagem de dependência.	IBM Corp.	www.ibm.com
MLC++	Classificação, regressão e clustering.	Silicon Graphics Inc.	www.sgi.com/tech/mlc
See5 Cubist	Classificação. Regressão.	Rule Quest	www.rulequest.com
Clementine	Classificação, regras de associação, clustering e padrões sequenciais.	SPSS Inc.	www.spss.com
Data-Miner Software Kit	Classificação e regressão.	Data-Miner PTy Ltd.	www.data-miner.com

THURAISINGHAM (1999) afirma que esses protótipos de ferramentas MD podem ser agrupados, de acordo com a sua funcionalidade, em: Novos Modelos Funcionais, Tratamento de Novos Tipos de Dados, Escalabilidade e Compreensibilidade dos Resultados.

Novos Modelos Funcionais: O objetivo desses projetos é integrar métodos de MD com o gerenciamento do BD. As principais características para projetos que se enquadram nesta funcionalidade estão relacionadas à otimização das consultas, na tentativa de oferecer um suporte mais adequado para os métodos de Mineração de Dados.

Tratamento de Novos Tipos de Dados: Esses projetos contemplam o tratamento de dados não tradicionais, como os dados multimídia. Assim como na Visualização de Dados, técnica pertencente à MD, alguns projetos buscam oferecer aos usuários os resultados obtidos e que devem estar armazenados em seu formato de apresentação.

Escalabilidade: Com o crescimento exponencial no armazenamento de dados, alguns projetos buscam tratar grandes volumes. A priorização, na formatação dos dados, para a MD, poderá colocar em colapso os aplicativos, caso os depósitos de dados não estejam adequados.

Compreensibilidade dos Resultados: Os resultados obtidos durante a MD podem ser complexos e tornar difícil a sua interpretação. A solução para este problema é a Visualização de Dados. A proposta deste trabalho de dissertação vem a confirmar a funcionalidade deste modelo. A apresentação de matrizes que contêm células com dados quantitativos, relativos ao número de visitas para um determinado ponto da loja, tornam a sua interpretação muito mais difícil quando comparada com a apresentação de um gráfico de superfície.

Após a apresentação das informações e a interpretação por parte dos especialistas, os sistemas que incorporam algoritmos de Mineração de Dados são capazes de fornecer suporte a outros sistemas, como os ditos Sistemas de Apoio à Decisão (Figura 05)

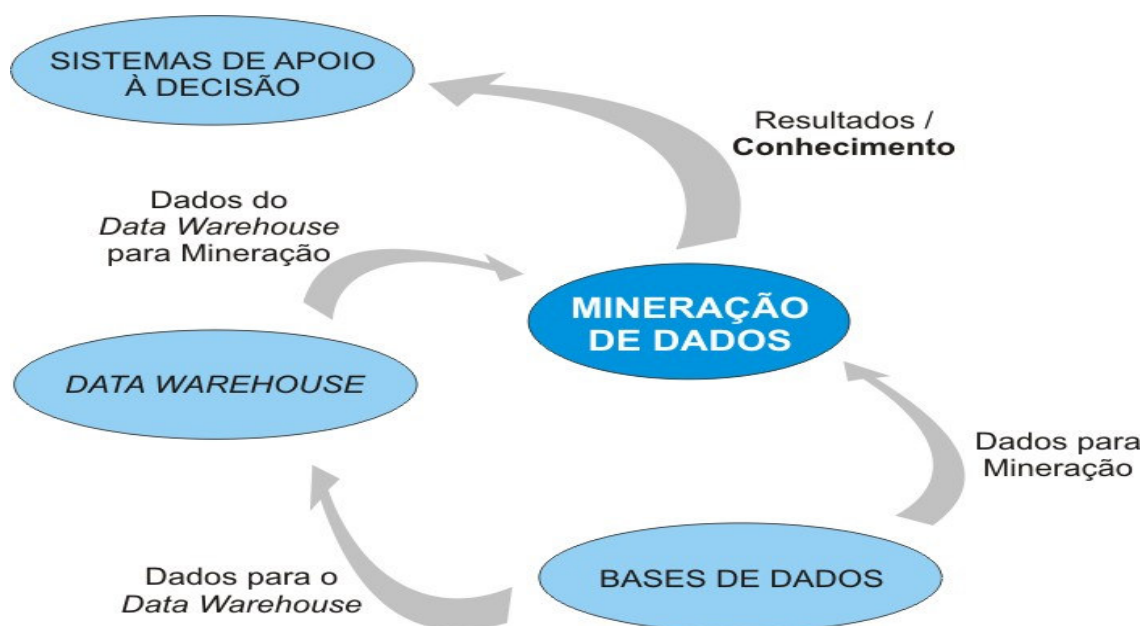


Figura 05 - Relações entre Base de Dados, DW, MD e SAD. Fonte: Rezende (2003)

3.1 Sistemas de Apoio à Decisão (SAD)

Também conhecidos como DSS (“Decision Support Systems”), os SAD são considerados por alguns autores como sendo uma evolução dos Sistemas de Informações Gerenciais (SIG).

Surgidos entre os anos 1960 e 1970, os sistemas desenvolvidos sob esta sigla são caracterizados pela proposta de interação direta com o usuário e são conceituados como “ferramentas eficientes” no auxílio à tomada de decisões para problemas não estruturados ou de difícil estruturação.

De acordo com GHODDOSI (2003), *apud* BINDER (1994), os SAD são sistemas mais complexos, que permitem total acesso à base de dados corporativa, à modelagem de problemas, à simulações, e possuem uma interface amigável. Além disso, auxiliam o executivo em todas as fases de tomadas de decisão, principalmente nas etapas de desenvolvimento, comparação e classificação dos riscos, além de fornecer subsídios para a escolha de uma boa alternativa.

3.2 As Etapas envolvidas na Mineração de Dados

Existem algumas divergências quanto às etapas envolvidas no processo de Mineração de Dados. Para FAYYAD *et al* (1996b), o processo foi dividido em nove etapas, enquanto WEISS & INDURKHIA (1998) apresentaram apenas quatro. Alguns autores prevêm, ainda, a existência de classes de tarefas em MD, as quais serão apresentadas no item 3.3 deste trabalho.

Durante a implementação do protótipo desenvolvido para ilustrar a funcionalidade da metodologia empregada durante este trabalho, foram seguidas as etapas descritas por WEISS & INDURKHIA (1998), por tratar-se de uma versão reduzida, sem a necessidade de implementação do *Data Warehouse*.

3.2.1 Identificação do problema

Nesta fase é requerida a participação da equipe responsável pelo desenvolvimento do projeto e, em muitos casos, inclusive a do usuário final.

Enquanto aos responsáveis técnicos pelo processo de Mineração de Dados é atribuído o domínio da aplicação, em suas numerosas etapas, ao usuário final é atribuída a tarefa de representar a classe de usuários que, efetivamente, faz uso do conhecimento extraído na sua forma de apresentação pelo sistema. Em determinados casos, o usuário final também poderá ser um especialista do domínio e exercer fundamental importância na fase de planejamento e identificação do problema.

Durante essa fase são discutidas e apresentadas questões como a precisão desejada nos dados, critérios de desempenho e outras características de segurança e controle na formatação dos resultados.

O planejamento deve abranger o conceito de formatação, sigilo das informações e perfil do usuário final. Algumas aplicações não contemplam a utilização dos dados por um usuário final, ficando os resultados disponíveis apenas para os próprios especialistas do domínio.

Em outros casos, a atribuição hierárquica vai assegurar que a informação chegue somente aos níveis desejáveis, não permitindo que usuários não autorizados tenham acesso às informações de interpretação e de caráter criterioso e restrito.

A escolha do conjunto de dados também acontece nesta etapa e deve suprir a demanda pelas informações em todas as suas formas de apresentação e recuperação. Os conjuntos de dados, por sua vez, devem oferecer formas distintas de apresentação, garantindo a exclusividade das informações e minimizando o re-processamento dos dados já classificados (atividades preditivas) ou segmentados (atividades descritivas). A figura 06 traz referências para algumas tarefas compreendidas na Mineração de Dados quanto ao modelo das atividades e as tarefas descritas em cada um deles.

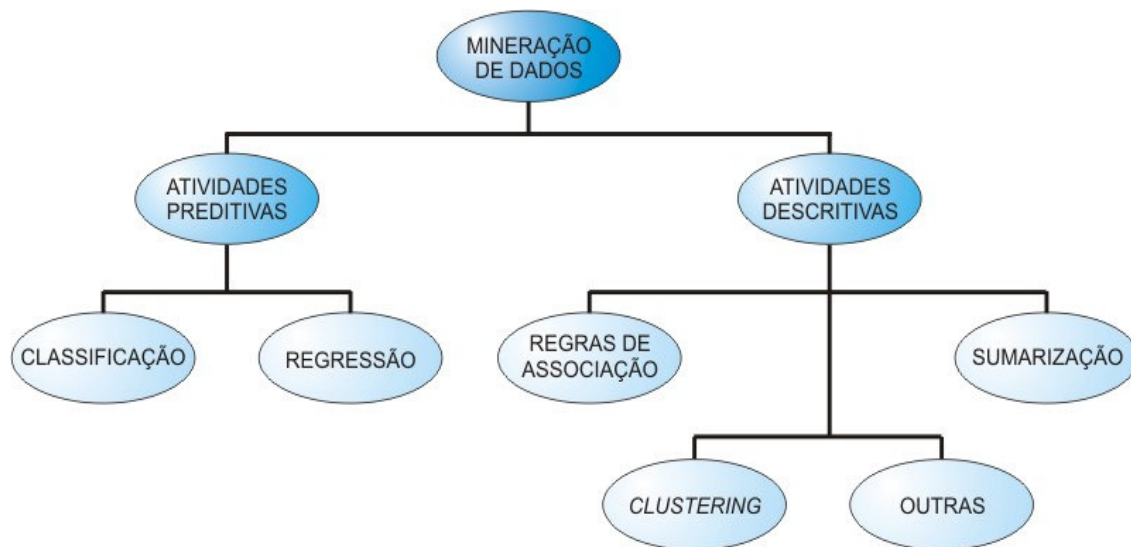


Figura 06 - Tarefas de Mineração de Dados. Fonte: Rezende (2003)

Estabelecendo-se um paralelo com o objeto desse trabalho, BRONDMO (2001) diz que a única certeza que se tem a respeito dos perfis de clientes é que eles se modificarão ao longo do tempo. Por isso, uma das exigências mais importantes do projeto de um banco de dados de perfil de clientes é que ele seja flexível e expansível.

As atividades pertinentes à área de atuação também são fundamentais na escolha correta do conjunto de dados e dos processos envolvidos na Mineração de Dados. Atividades nas quais exista uma forte relação entre as entidades geradoras dos dados, tornam mais simples a sua interpretação, como no caso deste trabalho, no qual se pode perceber uma relação muito forte entre o consumidor e o produto. O processo da compra, pelo consumidor que se desloca no interior da loja, tende a ser repetitivo, e o modelo adotado no conjunto de dados pode ser mais simplificado, evitando até a necessidade da criação do repositório de dados.

É prudente salientar que a existência de um repositório de dados é altamente recomendado para a otimização das rotinas computacionais que exijam a recuperação do conteúdo armazenado, principalmente nos casos em que existe fraca relação entre as entidades geradoras.

3.2.2 Pré-Processamento

A necessidade da implementação dessa etapa depende do nível de personalização do banco de dados existente, em relação aos dados desejados nas fases seguintes.

Em determinados casos, as formas de apresentação dos dados são incompatíveis, tornando-se necessária uma integração entre estes distintos formatos. O tratamento consiste em padronizar a recuperação dos dados a fim de minimizar o processamento na fase interativa.

Existem situações em que a recuperação dos dados é mais rápida e a forma de apresentação já se encontra equalizada com a forma exigida pelo algoritmo minerador. Para estes casos, o processo pode ser minimizado.

Questões relacionadas à capacidade de processamento e ao tempo de resposta também são definitivas para a opção da realização do pré-processamento, objetivando diminuir o tempo de resposta nas fases seguintes.

Com a utilização de bancos distribuídos em mais de uma plataforma operacional, muitas poderão ser as etapas descritas nesta fase de pré-processamento, podendo-se citar dentre as principais, extração e integração, transformação, limpeza, seleção e redução.

Extração e Integração: Os dados que farão parte dos processos de mineração poderão estar em formatos distintos quanto a sua tabulação, formatação e forma de apresentação. Nesta fase os dados deverão ser unificados e formatados com a mesma estrutura.

Transformação: É uma etapa caracterizada por dois eventos principais, o agrupamento e a transformação propriamente ditos. O agrupamento visa a agilizar e reduzir o processamento disposto no algoritmo de extração de padrões, podendo levar à criação de um resumo num formato mais próximo do utilizado pelo algoritmo.

Quanto à transformação, ela é necessária para unificar os formatos. Algumas respostas para as mesmas perguntas podem estar em formatos diferentes por motivos diversos. Pode-se citar, por exemplo, uma data gravada no formato numérico, contando-se o número de dias corridos a partir de uma data de referência (ex.: 1367), com a finalidade de compactação sendo comparada a uma outra data no seu formato original (ex.: 10/01/2005). Esta equalização reduzirá o processamento durante a fase de extração de padrões.

Limpeza: Aqui, os dados são validados de forma a evitar informações incompletas que poderiam inviabilizar a sua utilização. Os dados inválidos podem ter origem no processo de aquisição, através de erros de digitação ou, até mesmo, em falhas de *hardware* para os dados com entradas automatizadas. Falhas nos formulários de entrada de dados e a falta de testes de consistência e validação podem permitir a gravação de dados inválidos no banco.

Seleção e Redução de Dados: Restrições de ordem técnica durante o processamento exigem a redução do número de amostras utilizadas durante o processo de extração de padrões. Assim como em qualquer processo estatístico, a redução e a seleção dos dados deve ser criteriosa, para evitar a ineficácia da amostra. Há a necessidade da manutenção de dados capazes de representar as tendências existentes anteriormente à realização do processo.

3.2.3 Extração de Padrões

A busca pela identificação de padrões pode exigir mais de um algoritmo de busca, bem como a sua execução na forma recursiva. Durante a Extração de Padrões, o algoritmo deve buscar respostas no formato mais próximo possível aos objetivos pré-estabelecidos na fase inicial. BARRETO (2001) diz que, em determinados casos, a facilidade de reconhecer padrões depende dos padrões a que o indivíduo foi exposto anteriormente, afirmação esta que justifica a execução na forma recursiva.

Esta etapa prevê a escolha do algoritmo minerador, do processo de mineração e a tarefa da própria extração dos padrões.

Um algoritmo de extração de padrões pode ser classificado em duas formas, a preditiva e a descritiva que foram brevemente apresentadas através da Figura 06, na página 41 deste trabalho.

Na forma preditiva, há uma generalização das experiências passadas, e o algoritmo é capaz de atribuir à classe um novo exemplo.

Já na forma descritiva, os dados não possuem uma classe específica, e o que há é uma sumarização dos dados através de regras de associação.

3.2.4 Pós-Processamento

A *Extração de Padrões* não põe fim ao processo de *Extração de Conhecimento de Base de Dados*. É necessário que o especialista do domínio faça uma validação dos padrões encontrados. Esses padrões podem ser muito numerosos e necessitar de refinamento. Podem, ainda, apresentar uma alta complexidade, ao ponto de não fornecer as respostas nas formatações desejadas.

A validação desse processo pode ser obtida seguindo critérios de desempenho e de qualidade. Seja qual for o critério utilizado e a complexidade envolvida para a avaliação, o objetivo é oferecer, ao usuário da informação, uma medida de caráter prático e eficiente na avaliação do quesito desejado.

A resposta obtida deve ser comparável ao conhecimento prévio do avaliador. Caso a resposta obtida não ofereça a seu usuário um critério inequívoco de avaliação, o processo poderá ser repetido com o intuito de obter outro resultado mais adequado.

3.3 Classes de Tarefas em Mineração de Dados

O processo para a descoberta de padrões e tendências escondidos em grandes volumes de dados exige a utilização de diversas tarefas e técnicas. Segundo VIANA (2004), tarefas são classes de problemas enquanto técnicas são grupos de soluções para os problemas apresentados nas tarefas. O mesmo autor apresenta algumas classes de tarefas básicas:

- **Classificação:** Objetiva a criação de modelos de classes de dados, nas quais estes poderão ser agrupados. Esta tarefa é preditiva por ocorrer de forma pré-determinada ao evento de classificação.
- **Associação:** Busca estabelecer relações entre dois ou mais dados. Esta tarefa é considerada descritiva, por ocorrer sobre dados históricos.
- **Segmentação:** Para estes casos, o próprio sistema poderá definir as classes, considerando, para isso, as semelhanças encontradas. Por exemplo, uma base de dados contendo as principais características de doenças pode ser dividida em grupos, considerando-se as semelhanças existentes nos efeitos colaterais produzidos. Esta tarefa é considerada descritiva.
- **Regressão:** Representa o esforço realizado para atribuir um valor numérico a uma variável ainda desconhecida. Por exemplo, estimar a probabilidade de uma *tsunami* atingir o litoral brasileiro, considerando a existência de falhas geológicas. Esta tarefa é considerada preditiva.

AMARAL (2001), *apud* VIANA (2004), sugere a utilização do exemplo ilustrado na Figura 07 para identificar as diferenças entre classes de tarefas utilizadas em Mineração de Dados. Para tal, cada ponto no gráfico representa um mutuário a quem foi concedido um empréstimo. Num total de 14 amostras, o eixo horizontal representa o salário do tomador do empréstimo, enquanto o eixo da vertical representa o valor de sua prestação mensal tomada por conta do financiamento. Os clientes adimplentes são representados pelos círculos, enquanto os inadimplentes são representados pelas formas quadradas. O objetivo é definir padrões em que os clientes adimplentes estejam separados dos inadimplentes.

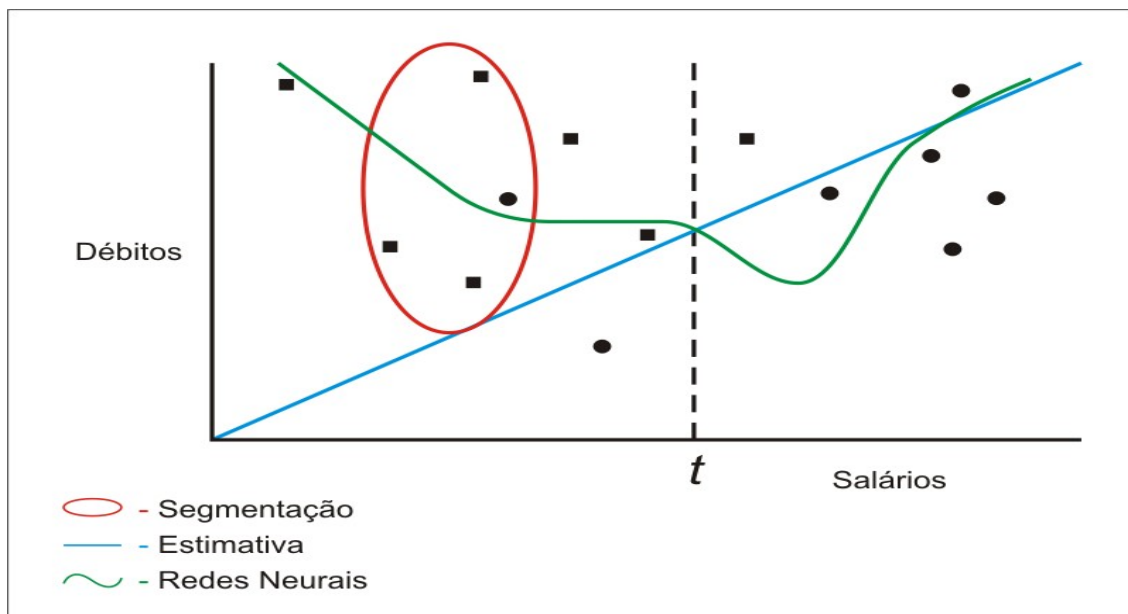


Figura 07 - Classes de Tarefas de Mineração de Dados. Fonte: VIANA(2004)

3.4 Técnicas de Apoio à Mineração de Dados

Existem várias técnicas que podem estar envolvidas no processo de Mineração de Dados, como Aprendizado de Máquina, *Data Warehousing*, Banco de Dados, Estatística e Visualização de Dados.

As técnicas são utilizadas em vários domínios da aplicação, onde são escolhidas de acordo com o problema a ser resolvido. As etapas envolvidas no processo de Mineração de Dados poderão utilizar uma ou mais técnicas, sendo imprescindível o conhecimento de todo o ambiente da aplicação para a seleção da técnica mais adequada.

Os resultados podem ser extraídos diretamente da base de dados ou obtidos através do *Data Warehouse*, quando este existir.

Assim como descrito na etapa chamada Pré-Processamento, os dados devem ter sua apresentação num formato compatível e equivalente para que sejam interpretados pelo algoritmo minerador.

3.4.1 Estatística

A relação existente entre a Mineração de Dados e a estatística é muito forte, principalmente na fase inicial de recuperação e validação dos dados. Os objetivos entre a estatística e o próprio processo de Mineração de Dados são muito próximos e orientados para a busca de padrões de comportamento e informações escondidas em grandes volumes de dados.

3.4.2 *Data Warehouse*

Apesar de a Mineração de Dados estar fortemente ligada a utilização do *Data Warehouse* (DW), esta técnica pode ser um processo opcional para o algoritmo minerador. Como já citado, a utilização do DW visa diminuir o tempo de resposta e do processamento envolvido para chegar ao dado.

3.4.3 Aprendizado de Máquina

O emprego de algoritmos do Aprendizado de Máquina serve para agrupar os dados numa forma associativa, derivada da observação de regras que tornam o dado mais facilmente compreensível ao ser humano.

Esse tipo de algoritmo é caracterizado pelo melhoramento do desempenho obtido através da avaliação de experiências anteriores.

Para criar a experiência necessária, é implementado um treinamento que pode ser direto ou indireto.

No treinamento direto, a experiência é obtida através da repetição exaustiva de uma atividade, enquanto no treinamento indireto o algoritmo obtém sua experiência através de tarefas já realizadas.

3.4.4 Visualização de Dados

A Visualização de Dados é uma importante técnica empregada nos processos de Mineração de Dados. Seu objetivo é facilitar a compreensão e a interpretação dos resultados obtidos.

A Visualização dos Dados age naturalmente na percepção humana, facilitando a observação de novos padrões. Os seres humanos tendem a exercer uma interpretação mais rápida e funcional quando da apresentação de resultados no formato gráfico frente ao formato tabular. (Quadro 5 e Figura 08)

Quadro 5 - Dados apresentados na forma tabular. Fonte: ALMEIDA (2004)

X	12,000	13,000	14,000	15,000	16,000	17,000	18,000	19,000	20,000	21,000
Y	0,000	0,259	0,500	0,707	0,866	0,966	1,000	0,966	0,866	0,707
X	22,000	23,000	24,000	25,000	26,000	27,000	28,000	29,000	30,000	31,000
Y	0,500	0,259	0,000	-0,259	-0,500	-0,707	-0,866	-0,966	-1,000	-0,966
X	32,000	33,000	34,000	35,000	36,000	37,000	38,000	39,000	40,000	41,000
Y	-0,866	-0,707	-0,500	-0,259	0,000	0,259	0,500	0,707	0,866	0,966

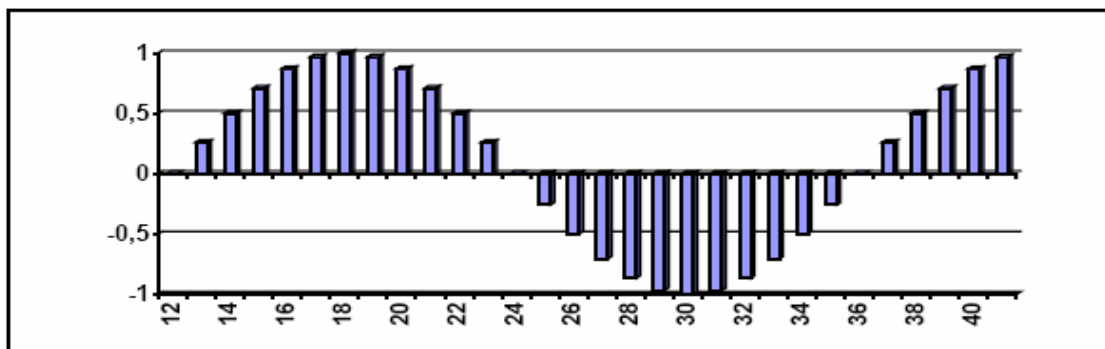


Figura 08 - Dados apresentados na forma gráfica. Fonte: ALMEIDA (2004)

Segundo HAN & KAMBER (2000), de maneira geral, Visualização de dados e *Data Mining* são processos que podem ser integrados das seguintes maneiras:

- **Visualização dos dados armazenados em banco de dados:** os dados em banco de dados podem ser vistos sob diferentes níveis de granularidade ou abstração, ou como diferentes combinações de atributos ou dimensões 3D curvas, superfícies, grafos ligados e outras.

- **Visualização dos resultados do processo de Mineração de Dados:** é a apresentação dos resultados ou conhecimentos, obtidos pela Mineração de Dados, através de formas visuais como árvores de decisão, regras de associação, agrupamentos (*clusters*) e outras.

- **Visualização do processo de descoberta de conhecimento em bases de dados:** esse tipo de visualização apresenta ao usuário as várias etapas do processo de descoberta de conhecimento em bases de dados através de formas visuais, de maneira que ele possa acompanhar todas as etapas desde a extração dos dados da base de dados, passando pela limpeza, integração, pré-processamento e mineração até o armazenamento e apresentação dos resultados.

- **Mineração visual de dados interativa (*Visual Data Mining – VDM*):** nesse tipo de visualização, ferramentas de visualização podem ser utilizadas no processo de *Data Mining* para ajudar o usuário a tomar decisões durante o processo.

São exemplos de técnicas de Visualização de Dados o emprego de gráficos, espectros, árvores, cubos 3D, curvas, grafos e outras formas de apresentação gráfica.

A forma de apresentação dos resultados obtidos, neste trabalho, utiliza a Visualização de Dados, por tornar mais simples a identificação dos pontos visitados, os prováveis deslocamentos e a compreensão de possíveis relacionamentos existentes entre eles.

4. O AMBIENTE DE CONSUMO

Alguns especialistas em computação admitem que prever as mudanças necessárias aos sistemas de computação é tão importante quanto ter a capacidade de reação às mudanças de forma rápida e segura.

Estabelecendo-se uma analogia com as mudanças propostas pela Natureza, vale lembrar que, há um século, não havia como detectar grandes furacões, e foram muitas as grandes catástrofes registradas naquele período. Atualmente, graças aos complexos sistemas de monitoramento, que incluem bóias marítimas, aviões guarnecidos de radar e, até mesmo, satélites, essas calamidades causam efeito menor à vida dos seres humanos.

Contudo, mesmo tendo a previsão de uma grande catástrofe, não é possível anular-lhe os efeitos. O furacão vai passar e caberá aos interessados, minimizar seus efeitos já que a causa não pode ser evitada.

Assim, são as grandes mudanças nos mercados financeiro, social, econômico e outros mais, que cercam as grandes populações de todos os países. DAWKINS (2001, pág. 81) acrescenta que os computadores têm assumido, recentemente, grande parte da função de simulação, não apenas em estratégia militar, mas em todos os campos em que a previsão do futuro é necessária, como Economia, Ecologia, Sociologia e muitos outros.

Um atentado pode dar início a uma guerra e esta a outras grandes transformações mundiais. A tecnologia e a velocidade oferecidas pelo desenvolvimento global também mudaram a velocidade com que o mercado deve reagir aos fatos. O resultado de prévias eleitorais em grandes países ou mesmo a descoberta de um novo campo petrolífero pode imprimir algum tipo de mudança no comportamento mundial.

Atualmente, muitas empresas tentam prever os grandes acontecimentos que possam alterar o rumo planejado para seus negócios. Mas, além de prever, é necessário responder com a devida velocidade.

Gerentes de estabelecimentos comerciais agem como sensores do mercado. A reação deve ser rápida a fim de evitar estoques desnecessários e compras errôneas, além de oferecer ao cliente aquilo que ele está procurando.

A variedade de produtos também é de fundamental importância para que seja oferecido ao cliente o produto em todas as suas modalidades e formas de apresentação.

A disponibilidade de um mesmo produto em embalagens de tamanhos variados é uma prova dos esforços realizados pela indústria para adequar suas ofertas às necessidades do cliente.

Técnicas de abastecimento e reposição de produtos nos expositores estão aliadas aos esforços realizados pelas equipes de logística e administração de estoques. É necessário não apenas a identificação da necessidade de reposição, mas, também, a efetiva reposição em tempo e forma adequados.

A aplicação de técnicas como o cálculo do giro de produtos, curvas ABC e listagem de produtos não-movimentados são exemplos de funções implementadas em *softwares* aplicativos utilizados na gerência de estoques.

É comum nas grandes redes varejistas e em outras centrais de compras, com origem na associação de pequenos varejistas, a manutenção de um canal direto de comunicação com seus fornecedores. Através da *EDI (Electronic Data Interchange)*, os sistemas computacionais existentes nas lojas afiliadas comunicam seus estoques para uma central de compras e esta faz a consulta aos seus fornecedores. Caso a consulta seja aprovada, são estabelecidos critérios pragmáticos de entrega.

A principal função dos sistemas de *EDI* é a integração dos dados e das transações entre diferentes organizações com o objetivo de harmonizar o processo de troca de informações entre seus usuários.

As atribuições destinadas ao sistema *EDI* diferem das aplicações com suporte ao sistema *ECR (Efficient Consumer Response)*. Nesse último, o varejista é monitorado diretamente pelo seu fornecedor, enquanto no *EDI* existe apenas a padronização das informações durante as suas trocas. O sistema *ECR* também é conhecido como *kanban* eletrônico.

A entrega de produtos pelo fornecedor deve ser feita com precisão de horário, inclusive com a imposição de punições comerciais aos fornecedores, caso estes extrapolem as margens permitidas. A fixação da data prevista de entrega é importante para o controle do fluxo de entrada de mercadorias, por gerar uma rotina mais eficiente, evitando sobrecargas nos depósitos e para as equipes de recepção e desembarque.

Segundo CHRISTOPHER (1997), os avanços realizados na introdução de sistemas logísticos de resposta rápida são todos baseados no fluxo de informações, partindo do ponto de demanda real diretamente para os sistemas logísticos e de reabastecimento pelos fornecedores.

O rigor exigido pelo comprador para com o seu fornecedor, aliado à precisão disponível pelos sistemas computacionais de ambos, são fatores de grande relevância para o sucesso do planejamento na reposição dos estoques. Contudo, é concreto que muitos fatores externos possam interromper um planejamento e comprometer todo um ciclo apurado em torno da reposição das mercadorias. A reposição na quantidade e no tempo corretos é de suma importância para assegurar o atendimento pretendido ao cliente. Uma greve ou mesmo um acidente automobilístico poderiam comprometer toda essa etapa de planejamento.

Para McGEE e PRUSAK, (1993, p.4), a gerência de informações estratégicas,

...Não é a tecnologia, mas sim o seu uso que cria valor adicional. O valor da tecnologia da informação depende da informação e do papel desempenhado por ela nas organizações. A informação é capaz de criar valor significativo para as organizações, possibilitando a criação de novos produtos e serviços, aperfeiçoando a qualidade do processo decisório em toda a organização. Esse, no entanto, não é um resultado obrigatório.

Enquanto algumas equipes do comerciante varejista monitoram a qualidade das compras e a gerência dos estoques, outras acompanham o grau de satisfação dos clientes. Todas têm como objetivo fidelizar e expandir a sua carteira de consumidores. Seja um crescimento horizontal ou vertical, o objetivo é a maximização dos serviços e, também, dos dividendos.

Qual seria o impacto causado na liquidação de um determinado produto? Quais são os potenciais consumidores? Onde expor o produto? Esta promoção poderia influenciar na venda de outros produtos? Respostas para estas perguntas é o que esperam os administradores dos grandes conglomerados varejistas.

No presente trabalho, existe a necessidade explícita do entendimento e da identificação da informação como critério decisivo na aquisição do conhecimento, para, assim poder oferecer respostas que serão dadas às questões apresentadas no parágrafo anterior.

A visualização dos resultados obtidos poderá oferecer aos gerentes um quadro mais detalhado dos acontecimentos no interior dos ambientes de consumo.

Para a realização deste trabalho, a Ciência da Computação foi usada face à sua grande adaptabilidade em oferecer soluções para problemas que exijam um elevado grau de processamento e interpretação dos dados.

A velocidade de resposta oferecida pelos meios computacionais é de fundamental importância para adequar os sistemas físicos à real demanda dos consumidores.

4.1 Os Consumidores

Quer estejamos qualificados em camadas sociais, faixas etárias, estados civis ou em outros grupos de classificação, todos somos consumidores diante dos comerciantes varejistas.

A disposição dos produtos, no interior dos supermercados, busca atingir todos os consumidores de maneira direta e objetiva. Enquanto os pais estão em busca dos produtos de primeira necessidade, os filhos estão cercados por outros produtos feitos e dispostos especialmente para eles.

Aliadas aos esforços empreendidos no interior desses ambientes, estão as propagandas na TV, a utilização de personagens, a disposição dos produtos em embalagens de cores e formas atraentes e as campanhas de distribuição de brindes e similares.

Atualmente, as mudanças causadas pela elevação dos índices de desenvolvimento social estão permitindo um aumento no número de consumidores mais receptivos às variações de marcas e modelos de produtos concorrentes.

A manutenção de clientes, invariavelmente fiéis a determinadas marcas, é cada vez mais difícil. A indústria responde aumentando o número de formatos de apresentação de um mesmo produto com o intuito de não deixar o cliente sem a opção de compra e, deste modo, evita a formação de um potencial candidato à compra do produto similar oferecido pela marca concorrente. CHRISTOPHER (1997) acrescenta que nos mercados de consumo, os clientes decidem sua escolha entre as marcas que estiverem disponíveis naquele momento.

O impacto gerado pela alta variabilidade de marcas, a que os novos consumidores estão expostos e, também, dispostos, pelo poder de sedução que elas, sobre eles, exercem, torna o percurso no interior das lojas uma grande incógnita, visto que os produtos tradicionais, de marcas líderes e antes obrigatórios nos carrinhos de compra, já não exercem a mesma atração de tempos passados. No passado, é fato, esses produtos estavam dispostos de modo a garantir a vinda do cliente até determinado ponto da loja.

Essas características são de difícil aceitação pelas grandes redes varejistas. A falta de caminhos obrigatórios a percorrer pode tornar o consumidor um fugitivo de alguns pontos considerados tradicionais para a venda.

Sem essas referências de espaço, a venda de outros produtos a eles associados torna-se mais difícil e dispendiosa, exigindo um esforço maior por parte dos anunciantes e do próprio estabelecimento comercial.

Alguns varejistas utilizam-se de estratégias que *pulverizam* os produtos em promoção, distribuindo-os por todos os espaços existentes no supermercado.

Pensando nisso, muitas das grandes redes varejistas apostam num leiaute único para todas as suas lojas. As lojas são dispostas num formato já certificado como sendo funcional e eficiente. Desse modo, o caminho a ser percorrido por seus consumidores será sempre o mesmo, cabendo aos comerciantes apenas distribuir os novos produtos durante o provável trajeto. Segundo FERNANDES (1998), mesmo que o ambiente da loja seja composto por elementos complexos, deve ele ser organizado de tal forma que permita ao consumidor associar a imagem da rede a todas as lojas dela.

Pequenos varejistas também estão em busca de grandes nichos de consumidores. Para ganhar força e poder de administração e controle, reúnem-se através de centrais de negócios que repassam as informações obtidas aos comerciantes, que também estão comprando seus produtos de forma unificada, aumentando o poder de barganha com seus fornecedores de produtos e serviços.

A boa hospitalidade de novos ambientes, sempre mais atualizados, inclui ar condicionado, área de lazer para as crianças, degustação de produtos, empresas terceirizadas para a limpeza de veículos e outras comodidades que podem aumentar o tempo de permanência dos consumidores no interior das lojas. É com frequência que consumidores, desatentos, são o alvo principal dessas campanhas.

Campanhas promocionais buscam atrair consumidores pela oferta de produtos a preços realmente vantajosos e, após o ingresso do consumidor no ambiente da loja, é uma estratégia eficiente estimular o consumo através da elevação do tempo de permanência do cliente nas dependências do estabelecimento comercial.

A existência de um grande número de marcas competindo no mesmo tipo de produtos tende a tornar pequenas as diferenças entre estes. Esta aproximação é causada pelo excesso de concorrência e pelo desejo de apresentar um aspecto de equivalência para os consumidores que é dado, até mesmo, pela semelhança no formato de apresentação de suas embalagens. Alguns fabricantes utilizam estratégias multimarcas para poder enfrentar os concorrentes e assegurar uma presença *pulverizada* entre as prateleiras e gôndolas disputadas por produtos concorrentes. FLEURY (2000) afirma que, considerando marca, embalagem, tamanho e sabor, é possível encontrar cerca de 200 diferentes alternativas de dentifrícios no interior de um único supermercado de grande porte nos Estados Unidos.

O desenvolvimento de marcas próprias entre os supermercadistas também faz parte de uma estratégia que busca atrair o cliente e promover o próprio supermercado. A localização desses produtos, que comumente estão associados com preços promocionais, é privilegiada, e sua forma de apresentação segue os padrões utilizados na logotipia de concorrentes já consagrados em seus segmentos.

Um estudo de caso publicado na *Progressive Grocer*, por MATHEWS (1995), discorre sobre a crença de que marcas próprias propiciam uma vantagem comparativa sobre a concorrência e aumentam a lealdade dos clientes às lojas. Além disso, o sucesso de uma linha de produtos comercializados com marcas próprias, seja pela sua qualidade, seja pelos preços, traz benefícios diretos à imagem do supermercado.

Segundo CORSTJENS & CORSTJENS (1995), as percepções da marca pelo consumidor na loja são diferentes daquelas mensuradas de forma abstrata em pesquisas de mercado. Um varejista pode influenciar as posições do mapa de percepção das suas marcas e das de seu concorrente através do seu posicionamento de preço, leiaute, promoções, localização das mercadorias nas gôndolas e todos os tipos de comunicação no ponto de venda.

4.2 O Conceito Elasticidade-Espaço

A concepção de que produtos que ocupam maiores e melhores espaços no interior de um ambiente de consumo exercem vantagens sobre outros com localização menos privilegiada é aceita entre muitos autores e especialistas no assunto. Algumas pesquisas indicam que produtos que ocupam maior espaço nas gôndolas possuem maior probabilidade:

- de ser percebidos pelo consumidor,
- de ser por ele analisados e/ou manuseados;
- de ter sua imagem gravada na mente do consumidor e
- de ser efetivamente colocados nos carrinhos de compra.

A relação existente entre a área ocupada e o volume de vendas propiciou vários estudos, pois ela apresenta considerável facilidade para estabelecer métricas de desempenho e comprovação de resultados. De acordo com CURHAN (1972, p. 406), “elasticidade-espaço é a razão entre a mudança relativa em unidades de venda, pela mudança relativa no espaço ocupado pelo produto na gôndola” e pode ser representada por:

$$\Sigma e = \frac{ds}{d\lambda} \frac{\lambda}{s}$$

onde: Σe = elasticidade espaço

s = vendas em unidades monetárias

λ = espaço em unidade de comprimento

4.3 A Fidelização

Muitas são as empresas varejistas que realizam grandes esforços para buscar manter os clientes fiéis aos seus estabelecimentos de venda. A adoção de cartões de relacionamento e programas de pontuação com recuperação de brindes de bonificações fazem parte do cotidiano de muitos consumidores.

Por trás desses programas existem empresas vorazes, ávidas por descobrir padrões de comportamento e consumo de seus afiliados.

Com o monitoramento das compras realizadas pelo consumidor é possível identificar a concorrência, os horários preferidos de compras e até mesmo as categorias de produtos consumidos. Atividades de lazer e diversão complementam as informações a respeito dos consumidores. Os ramos de atividades são fornecidos diretamente pelas empresas administradoras dos cartões.

Até mesmo parceiros com algum grau de concorrência, mas de caráter estratégico, são convidados a compartilhar as instalações físicas dos grandes centros de consumo para evitar que o cliente tenha que se dirigir para outro estabelecimento. Esse conjunto de atividades, que não representam a atividade comercial principal do estabelecimento, é conhecido como *serviço opcional*. Quando o varejista oferece um conjunto mais abrangente de serviços, incluídos aí os serviços opcionais, é ele tido como *full-service*.

Segundo LEVY & WEITZ (1995), estes são alguns dos serviços disponíveis nos estabelecimentos varejistas:

- aceitação de cartão de crédito,
- alteração e montagem de mercadorias,
- manutenção de listas de presentes para noivos,
- reembolso de cheques,
- cuidados e atividades de recreação para crianças,
- financeira para concessão de crédito,
- entrega em domicílio,
- demonstração e exposição de mercadorias,
- manutenção de vestiários para prova de mercadorias,
- horários convenientes de funcionamento,
- sinalização extensiva para identificação de mercadorias,
- embalagem para presentes,
- equipamentos/instalações para deficientes e portadores de necessidades especiais,
- reserva de mercadorias,
- estacionamento,
- assistência pessoal na seleção de mercadorias,
- equipamentos,
- serviços de reparos,
- toaletes,
- setor de devolução, troca e garantia de produtos comercializados,
- guarda-volumes.

É, ainda, comum encontrar sapatarias, vídeo-locadoras, lotéricas, farmácias, laboratórios fotográficos, quiosques de auto-atendimento bancário, *pet-shop* e outros serviços que podem aumentar a conveniência existente nas instalações do varejista e, com isso, elevar o tempo de permanência e a exclusividade de compras naquele ambiente.

Segundo BORGES (1997), a teoria comportamental é de grande valia para o entendimento de como os fatores de ambiente da loja têm influência no comportamento do consumidor.

Ao tornar-se fiel, o consumidor evita buscar produtos similares em outro comerciante, principalmente porque o concorrente não pode oferecer um plano de bonificações com as mesmas vantagens do próprio emissor do cartão de relacionamento.

No processo de construção de um relacionamento fiel com o varejista, o consumidor mostra a mesma exigência que tem em relação a um produto ou serviço. Mas essa relação segue outras variáveis, as quais revelam como o consumidor percebe o varejista. Por isso, a construção de uma proposição de valor é mais complexa em empresas do setor do que em companhias industriais e, em um mercado hipercompetitivo, é condição para a sobrevivência (BERRY, 1999)

A fidelização existente no consumidor faz com que ele não compare apenas os preços oferecidos por outros concorrentes para um produto similar. O valor deve refletir todos os benefícios que ele recebe no esforço representado pelo ato da compra. A qualidade do produto, o ambiente, o ganho obtido em programas de milhagens, a atenção dos funcionários e a comodidade são alguns desses benefícios avaliados pelo cliente fiel.

Até mesmo a falta de disponibilidade imediata de recursos tem sido alvo de campanhas de fidelização. Grandes redes varejistas fazem uso de cartões de crédito com vantagens adicionais quando utilizados em compras efetuadas exclusivamente em suas lojas. Extensão dos prazos de pagamento e elevação de pontuação nos planos de recompensas são partes constantes destas práticas e podem refletir grandes diferenciais na manutenção e ampliação de uma carteira de clientes.

Tendo, a grande maioria dos consumidores, época pré-determinada para receber proventos, é natural que o retorno ocorra somente no próximo ciclo. Ao oferecer cartões de crédito, com vencimento pré-datado, o varejista estará permitindo e estimulando o consumo mesmo nos intervalos existentes entre as remunerações.

Esse estímulo ao consumo também é perceptível nos comerciantes que fazem promoções de grupos específicos de produtos vinculados ao dia da semana. Ao oferecer, para o cliente, um determinado grupo de produtos a preços vantajosos, o comerciante incentiva o retorno do cliente à loja num intervalo menor. Além disso, estas datas podem ter caráter estratégico no reabastecimento das lojas e aumentar o poder de barganha junto aos fornecedores.

5. PROPOSTA DE TRABALHO

A proposta defendida neste trabalho é a aplicação de métodos capazes de fornecer uma avaliação visual do provável tráfego descrito por um conjunto de clientes no interior de um ambiente de consumo com auto-atendimento.

Para o desenvolvimento desta proposta, foram adotadas algumas medidas que serviram para estabelecer uma referência espacial nos locais destinados à exposição dos produtos.

Para ilustrar alguns dos propósitos aqui apresentados, um protótipo foi desenvolvido.

Os procedimentos adotados e também algumas características do simulador serão apresentados ao longo desse capítulo.

5.1 Introdução

A descoberta de padrões de comportamento no ambiente espacial de consumo poderá gerar informações relevantes para a manutenção da carteira de clientes e para a otimização dos recursos.

Conhecer os prováveis trajetos descritos por clientes é algo capaz de fornecer um diferencial exclusivo para um varejista. O administrador, conhecedor desses trajetos, poderá dispor os produtos numa forma mais adequada e também promover a venda de novos produtos entre os espaços mais visitados.

A disposição dos produtos no interior de um supermercado deve seguir alguns critérios para assegurar a confiança do cliente. Produtos que exalam cheiros desagradáveis, como materiais de limpeza, colas e resinados, pneus e produtos do *petshop* devem ficar bem distantes de produtos que necessitam apelo de higiene e limpeza. Salienta-se que o cheiro de um produto pode ser um critério decisivo na escolha do produto.

A ausência de relógios distribuídos no interior das lojas é um fator para aumentar a permanência do cliente sem que se preocupe com um possível atraso.

Produtos associados, como massas, vinhos e queijos devem estar próximos para despertar o interesse do cliente e para evitar possíveis esquecimentos.

Dentro desses ambientes de consumo, as áreas mais valorizadas são as que estão ao alcance dos olhos e nos pontos de maior destaque. Segundo especialistas, a maximização do tempo de permanência do cliente no ambiente da loja é de fundamental importância para estimular e aumentar o consumo. Para aumentar o deslocamento do cliente na loja é habitual entre lojistas expor os produtos considerados básicos em prateleiras existentes no fundo da loja, aumentando a distância e, conseqüentemente, o tempo de permanência do cliente.

Produtos similares com embalagens distintas devem ser expostos lado a lado para que o cliente encontre o produto no tamanho ideal sem abandonar a sua marca de preferência.

As características de exposição dos produtos são de fácil compreensão, porém seguem padrões que inviabilizam alterações frente às evidências apresentadas. A análise do impacto causado por uma mudança nos expositores pode consumir muito dos lucros até a compreensão de uma possível mudança no fluxo existente nos corredores da loja.

A proposta apresentada neste trabalho é a representação dos prováveis deslocamentos efetuados pelos clientes durante as suas compras, tomando-se por referência a localização dos produtos consumidos.

Para representar a área útil oferecida ao consumidor junto ao simulador, esta foi mapeada nos eixos χ e γ , permitindo obter uma identidade espacial para cada um dos produtos arrolados no cadastro. A altura na qual o produto está exposto em relação ao solo não pode ser representada porque o gráfico de superfície está utilizando o terceiro eixo para a apresentação dos resultados obtidos. A representação deste mapeamento pode ser observada com a comparação realizada entre as Figuras 09 e 10, onde a segunda já apresenta a divisão em unidades amostrais.

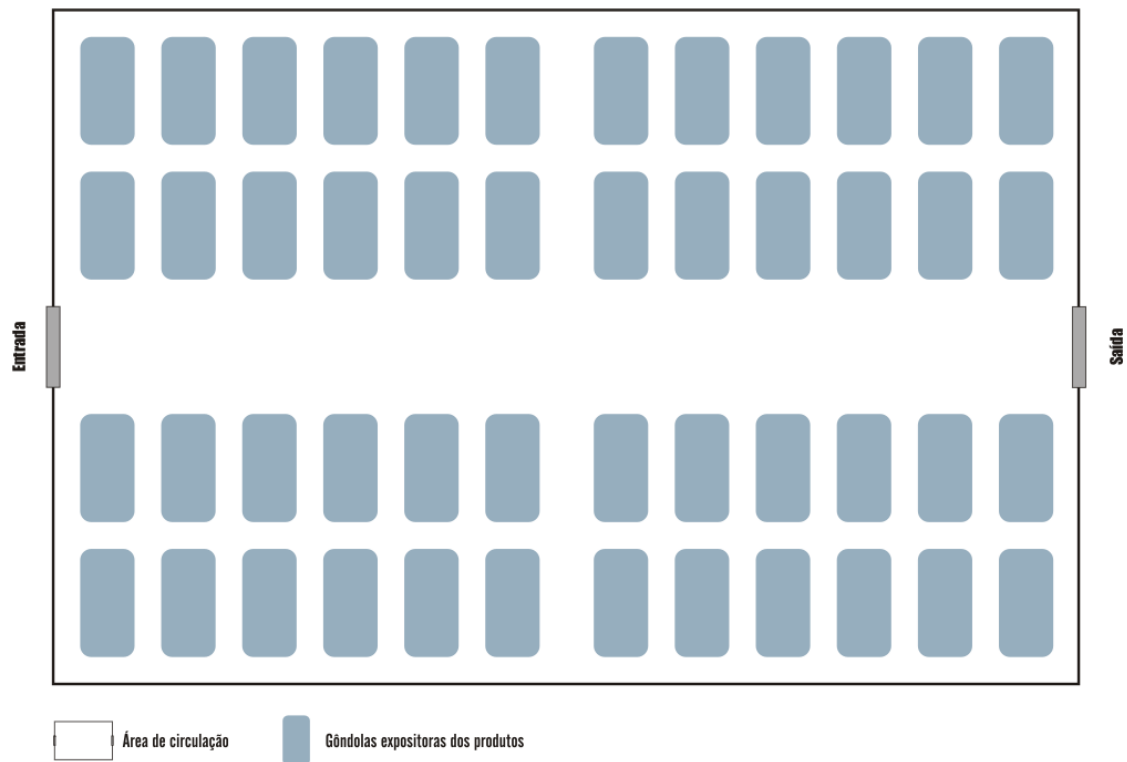


Figura 09 - Representação da área útil existente no ambiente de consumo.

Na figura 09 tem-se um modelo do interior de um ambiente de consumo. Essa figura representa a disposição das estantes (áreas sombreadas), da entrada (à esquerda), da saída (à direita), do corredor central e dos corredores secundários. Os *checkouts* não estão representados em nenhuma das duas figuras, a 09 e a 10.

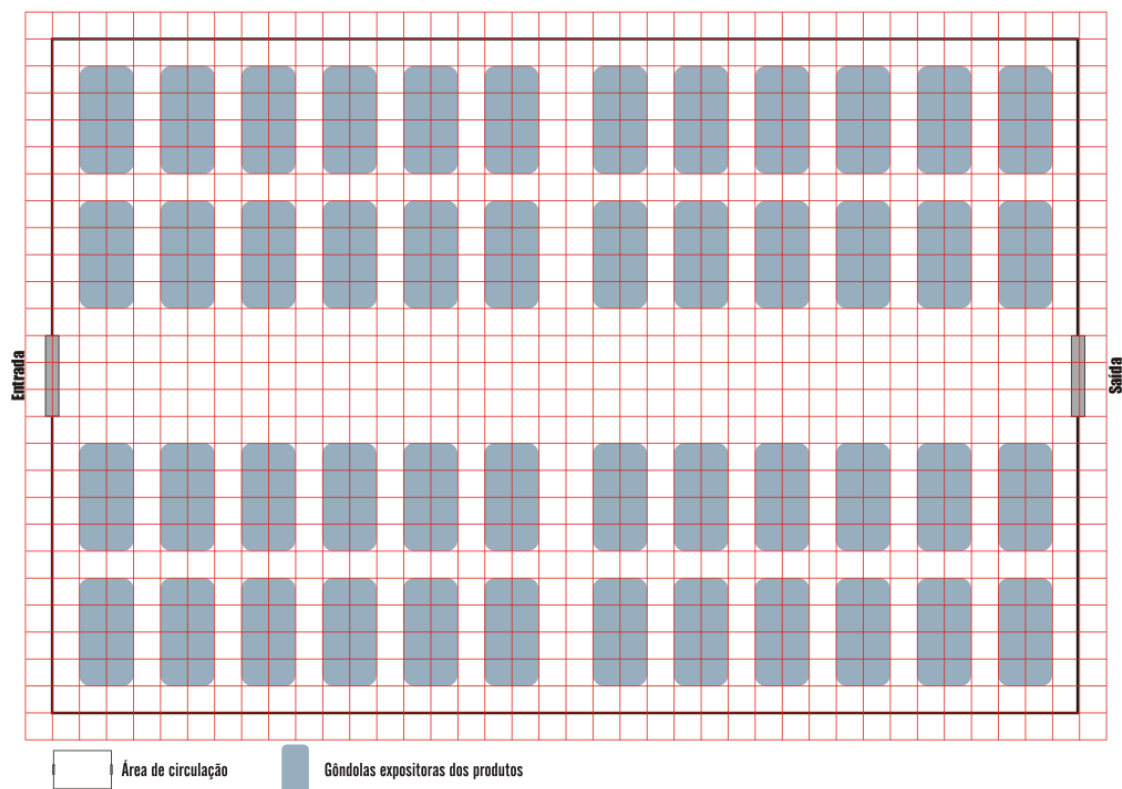


Figura 10 - Representação da área útil em unidades amostrais

O tamanho da área amostral também deve ser adequada aos resultados desejados. Dependendo do tamanho de uma unidade amostral, poderá, ela, compreender mais de um produto.

A granularidade representada pelo tamanho da unidade amostral é fator decisivo para o nível de detalhamento desejado. Segundo INMON (1998), a granularidade se refere ao nível de detalhe ou de resumo contido nas unidades de dados existentes. Para o simulador, a granularidade refere-se ao tamanho físico da unidade amostral, mas a interpretação pode ser a mesma existente quando o assunto abordado for o Banco de Dados.

Para a representação de produtos que são oferecidos em mais de uma localização, é necessária a identificação exclusiva de cada um deles para evitar, além de uma falsa representação do consumo localizado, a fragilização dos resultados obtidos.

Com base nos produtos adquiridos por um único consumidor, serão efetuadas as ligações entre os pontos representados pelas identidades espaciais de cada um dos produtos constantes da compra por ele realizada.

Como ponto inicial, será tomada a entrada no estabelecimento, a qual, como a saída, deverá possuir uma identidade espacial mapeada na área útil de consumo. Tanto a entrada quanto a saída poderão ser informadas pelo próprio usuário do sistema, opção esta que já se encontra implementada no protótipo.

Para calcular o deslocamento entre esses dois pontos, foi utilizada a distância de *Manhattan* por apresentar melhor desempenho nesse tipo de ambiente de consumo onde os produtos estão expostos em prateleiras e seus formatos já são conhecidos e comuns para toda a área.

Os resultados obtidos, após a análise de cada uma das compras totais efetuadas pelo consumidor, serão armazenados numa matriz dinâmica. A matriz dinâmica é representada por um vetor multidimensional, devendo existir uma célula para cada unidade amostral do espaço físico para o ambiente de consumo em questão. Por exemplo, a Fig. 10 representa uma loja com tamanho, em metros, de 38 x 25 (largura x profundidade) e unidade amostral de 1 m². O tamanho da unidade amostral representa a granularidade desejada na compilação dos dados.

A matriz é dita dinâmica porque os dados vão sendo acumulados durante o processamento das informações obtidas pelo simulador, sendo que uma nova matriz é criada a cada novo trecho encontrado. A sobreposição dos resultados armazenados na matriz poderá oferecer os subsídios necessários para a representação gráfica dos prováveis trajetos desenvolvidos.

A visualização dos resultados será disponibilizada através da representação da matriz em gráficos de superfície, podendo, estes, ser exportados para Sistemas de Apoio à Decisão.

5.2 Materiais e Métodos

O protótipo desenvolvido para ilustrar o trabalho não exigiu um hardware específico.

Para o desenvolvimento do protótipo, foi utilizado o compilador *Borland Delphi 7.0*, juntamente com a biblioteca gráfica *TeeChart Pro v6.0 TRIAL* disponível para avaliação.

Uma base de dados para treinamento foi criada utilizando um algoritmo gerador escrito em *Microsoft Visual FoxPro 8.0*. Além da base de treinamento já criada, está disponível a opção de gerar uma nova base aleatória através da passagem de parâmetros para o algoritmo gerador.

O protótipo, em sua versão inicial, já permitiu simular possíveis comportamentos de um grupo de consumidores. A técnica de simulação e seus conceitos básicos são, em geral, facilmente compreensíveis e justificáveis (FREITAS, 2001).

5.3 Etapas do projeto

A fase inicial para a implantação deste modelo exige a conversão dos espaços físicos em unidades amostrais. O tamanho de cada unidade amostral é decidido em função da precisão desejada pelo administrador do sistema. A fim de exemplificar, poder-se-ia converter uma loja, onde os produtos ocupam um espaço de 20 x 10 m, em 200 unidades amostrais de 1 m².

Após a conversão do espaço físico em unidades amostrais, é necessário indicar, dentro de cada um dos cadastros de produtos, qual a sua identificação espacial em relação aos eixos χ e γ . Os pontos de *Entrada* e de *Saída*, no ambiente de consumo, também deverão receber suas identificações espaciais encontradas durante o processo descrito no parágrafo anterior. É necessário, portanto, que os pontos *Entrada* e *Saída* estejam mapeados sobre o mesmo plano utilizado para gerar as identificações espaciais dos produtos.

A interpretação dada a um cliente é o reflexo da análise de um registro de venda obtido no *checkout* da loja. Para casos em que sejam selecionados mais de um registro de venda para o mesmo cliente, devem, os clientes, ser tratados como distintos. Esta característica deve-se ao fato de que o comércio não costuma realizar vendas identificadas, salvo os casos em que exista a necessidade de emitir uma Nota Fiscal por questões burocráticas, administrativas ou até mesmo operacionais.

Ao contrário das vendas realizadas “ao consumidor”, as vendas efetuadas com a utilização de cartões de relacionamento podem fornecer as informações necessárias para a identificação individual do cliente.

O protótipo permite selecionar um grupo de registros de venda e, a partir desse grupo, todos os registros de produtos comercializados serão obtidos através dos registros de vendas.

Com a seleção das vendas e de seus produtos, a interpretação do grupo de vendas ocorre de maneira concomitante, iniciando pela *Entrada*, passando por todos os produtos adquiridos durante a compra, até chegar ao ponto de *Saída*. Todos os pontos são referenciados na matriz temporária, fazendo-se, depois, a ligação entre eles, com a utilização da menor distância obtida pela distância de *Manhattan*.

O resultado é obtido na forma de uma nova matriz temporária a qual é adicionada à matriz dinâmica, que comportará, ao final do processamento, a representação de todos os prováveis trajetos obtidos nas matrizes temporárias e, supostamente, descritos por cada um dos consumidores analisados.

Com a obtenção da matriz em sua versão final, que já contempla a representação da união de todos os prováveis deslocamentos, o protótipo poderá apresentar, utilizando a técnica de Mineração de Dados, chamada de Visualização de Dados, alguns gráficos de superfícies que representem os prováveis percursos descritos pelo conjunto de consumidores analisados. Outras variáveis, como o volume total de vendas em unidades e valor, também poderiam ser acumuladas em matrizes secundárias.

Ao final da apresentação, a matriz dinâmica é descartada, enquanto o seu formato e os resultados obtidos poderiam ser armazenados ou entregues para outros Sistemas de Apoio à Decisão.

5.4 O Protótipo

Inicialmente, foi utilizado somente o compilador *Microsoft Visual Foxpro 8.0*, que apresentou grande velocidade e portabilidade, mas dispunha de um número muito reduzido de gráficos de superfície.

A adoção do compilador *Borland Delphi 7.0* permitiu a utilização de componentes gráficos externos capazes de representar com maior clareza os gráficos de superfície.

Após o término da versão do simulador, optou-se por deixar o gerador de amostras no compilador anterior e manter a versão final com o último compilador adotado. As chamadas para o gerador de amostras são feitas, diretamente, através do ambiente de simulação (Figura 12).

A manutenção da opção de utilizar a força bruta para encontrar o melhor trajeto, inclusive com itens de animação, enfatizou o ganho oferecido pelo modelo adotado.

A natureza deste trabalho não possibilitou a realização de testes comprobatórios para demonstrar a eficácia dos resultados obtidos. Para a descoberta de um índice de eficácia, faz-se necessária a realização de testes práticos, reportando o real trajeto descrito por um grupo de consumidores para comparar os resultados obtidos através do simulador. A grande demanda de recursos humanos para descrever um grupo relevante de consumidores não permitiu que testes desta natureza fossem aplicados.

No formulário de cadastramento do produto, apresentado através da figura 11, é possível estabelecer relações entre o produto, seu fornecedor e o público preferencial. Os campos existentes nesta tela e seus significados são:

- Código: Chave primária junto à tabela: produtos;
- Nome: Descrição do produto a qual será impressa no cupom de venda;
- Fornecedor: Código do fornecedor do produto. Este campo poderá ser utilizado como filtro durante a consulta (Fig. 12);
- Grupo: Código do grupo ao qual o produto pertence. Este campo poderá ser utilizado como filtro durante a consulta (Fig. 12);
- Classe: Código da classe ao qual o produto pertence. Este campo poderá ser utilizado como filtro durante a consulta (Fig. 12);
- G. Consumo: Código do grupo de consumo ao qual o produto pertence. Este campo poderá ser utilizado como filtro durante a consulta (Fig. 12);
- Eixo X: Posição ocupada pelo produto em relação à largura da área de consumo destinada à visitação de clientes;
- Eixo Y: Posição ocupada pelo produto em relação à profundidade da área de consumo destinada à visitação de clientes;

The image shows a software interface for product registration. The window title is 'Cadastro de Produtos' and the current view is 'Listagem de Registros'. The 'Detalhes' tab is selected, displaying a form with the following data:

Codigo.:	00001
Nome.:	CAMARÃO ROSA
Fornecedor.:	00001 ... PESCADOS SANTA CATARINA LTDA.
Grupo.:	00020 ... PEIXARIA
Classe.:	00001 ... CLASSE "A"
G. Consumo.:	00001 ... ADULTO
Eixo X.:	7
Eixo Y.:	20

At the bottom of the window, there is a toolbar with the following actions: Novo, Editar, Gravar, Cancelar, Excluir, and Imprimir.

Figura 11 - Formulário de cadastramento de produtos.

As relações estabelecidas no cadastro do produto são oferecidas como opções de filtragem durante a busca dos dados que serão avaliados e representados pelo simulador.

Além das características pertinentes ao cadastro de produtos, outras variáveis também encontram-se disponíveis para que sejam utilizadas como ferramentas de filtragem durante a seleção dos dados que serão enviados para o simulador.

A inclusão de filtros de caráter temporal no simulador permite uma simulação fracionada para demonstrar o comportamento dos clientes num determinado intervalo de horários, de dias e de ambos.

Estudos apontam que determinados horários são favoráveis para a comercialização de produtos específicos. Características ligadas ao horário são determinantes para estabelecer um perfil de comportamento dos clientes no interior da loja.

Com a utilização de filtros aplicáveis ao grupo de consumo, é possível visualizar apenas parte dos produtos daquele carrinho de compras. Esta característica permite avaliar trajetos específicos para cada um dos integrantes da compra.

A Figura 12 apresenta a tela onde é feita a seleção dos registros que serão utilizados na tentativa de encontrar os caminhos descritos pelos clientes. Os campos existentes nesta tela e seus significados são:

- **Fornecedor:** Caso este item esteja marcado, o usuário deverá informar o código do fornecedor desejado no campo imediatamente à direita. Se marcado, apenas os produtos entregues pelo fornecedor informado serão incluídos na busca. Caso contrário, não haverá distinção dos produtos quanto ao seu fornecedor;
- **Grupo:** Caso este item esteja marcado, o usuário deverá informar o código do grupo de produtos desejados no campo imediatamente à direita. Se marcado, apenas os produtos pertencentes ao grupo informado serão incluídos na busca. Caso contrário, não haverá distinção dos produtos quanto ao seu grupo;
- **Classe:** Caso este item esteja marcado, o usuário deverá informar o código da classe de produtos desejada no campo imediatamente à direita. Se marcado, apenas os produtos pertencentes à classe informada serão incluídos na busca. Caso contrário, não haverá distinção dos produtos quanto à sua classe;
- **G. Consumo:** Caso este item esteja marcado, o usuário deverá informar o código do grupo de consumo desejado no campo imediatamente à direita. Se marcado, apenas os produtos pertencentes ao grupo de consumo informado serão incluídos na busca. Caso contrário, não haverá distinção dos produtos quanto ao seu grupo de consumo;
- **Posição dos pontos inicial e final:** Para que o simulador possa representar com maior clareza o fluxo de deslocamentos encontrado, devem ser informados os pontos de entrada e de saída no ambiente de consumo. Ambos os pontos devem ser referenciados através de sua identidade espacial sob os eixos x e y, respectivamente, onde o eixo x representa a posição em relação à largura e o eixo y em relação a profundidade. Os quatro pontos que representam as posições de entrada e de saída são obrigatórios para a execução do simulador.

- Num: Caso este item esteja marcado, o usuário deverá informar o intervalo desejado entre os códigos das vendas que deverão fazer parte da busca. Caso contrário, não haverá distinção quanto aos códigos de vendas que estão sendo investigadas;
- Data: Caso este item esteja marcado, o usuário deverá informar o intervalo desejado entre as datas das vendas que deverão fazer parte da busca. Caso contrário, não haverá distinção quanto ao período das vendas que estão sendo pesquisadas;
- Hora: Caso este item esteja marcado, o usuário deverá informar o intervalo desejado entre o horário das vendas que deverão fazer parte da busca. Caso contrário, não haverá distinção quanto ao horário das vendas que estão sendo pesquisadas;
- Gerar Consulta: Ao clicar este botão, o usuário dará início ao processo de filtragem dos dados, considerando os filtros informados. Após a seleção dos registros, estes poderão ser consultados através do *grid* existente logo abaixo do campo: Produtos x Vendas Seleccionadas;
- Produtos x Vendas Seleccionadas: Este é um campo de saída de dados e indicará o número de registros que foram submetidos ao simulador após a aplicação dos filtros. Ainda na Figura 12, abaixo deste campo estarão indicados o número da nota, o código do produto e sua identidade espacial para ilustrar o resultado da pesquisa;
- Qtd. Vendas: Este campo é utilizado pelo gerador de amostras e indica o número de registros de vendas que devem ser criados;
- Qtd. MIN: Este campo é utilizado pelo gerador de amostras e indica a quantidade mínima de produtos adquiridos pelo cliente durante uma compra;
- Qtd. MAX: Este campo é utilizado pelo gerador de amostras e indica a quantidade máxima de produtos adquiridos pelo cliente durante uma compra;
- Hora MIN: Este campo é utilizado pelo gerador de amostras e indica a hora mínima para uma compra efetuada pelo cliente;
- Hora MAX: Este campo é utilizado pelo gerador de amostras e indica a hora máxima para uma compra efetuada pelo cliente;

- Data Min: Este campo é utilizado pelo gerador de amostras e indica a data mínima para uma compra efetuada pelo cliente;
- Data MAX: Este campo é utilizado pelo gerador de amostras e indica a data máxima para uma compra efetuada pelo cliente;
- Botão Salva: Serve para armazenar os critérios de geração de amostras no banco de dados;
- Botão Desfaz: Serve para recuperar os critérios de geração de amostras do banco de dados;
- Botão Gerar População: Ao clicar esse botão, o usuário dará início ao processo de geração das amostras, considerando suas preferências indicadas nos campos específicos localizados à esquerda desse item.

Produtos

Fornecedor: ... PESCADOS SANTA CATARINA I

Grupo.: 00008 ... PET SHOP

Classe.: ... CLASSE "A"

G. Consumo. ... ADULTO

XOR

Posição do ponto inicial e final:

X Inicial 00 X Final 37

Y Inicial 13 Y Final 13

Notas:

Num.: ... à ...

Data: / / à / /

Hora: ... à ...

Gerar Consulta

Produtos x Vendas Selecionadas: 347

NOTA	PRODUTO	EIXO_X	EIXO_Y
00001	00093	1	11
00001	00058	1	11
00001	00007	20	4
00001	00014	1	11
00002	00082	1	11

Qtd. Vendas 1000 Qtd. MIN 3 Qtd. MAX 5 Hora MIN 1 Hora MAX 26 Data Min 1/1/2004 Data MAX 31/12/2004

Salva Desfaz Gerar População

Figura 12 - Formulário de aplicação de filtros para a seleção de vendas.

Após a aplicação dos filtros desejados é feita a seleção dos registros de venda que serão utilizados para gerar os prováveis deslocamentos.

No momento da simulação é possível optar entre o uso da distância de *Manhattan* ou da força bruta na busca dos prováveis trajetos. A força bruta efetua a busca como num radar em sentido horário e adotará o primeiro vizinho encontrado como sendo o próximo. É oferecida a animação durante a escolha do próximo ponto a ser visitado, de forma a evidenciar o processo de escolha do trajeto.

A Figura 13 representa os prováveis trajetos descritos pelo grupo de vendas que aqui são tratados como sendo consumidores. O processo de descoberta de trajetos recebe a animação em tempo real, indicando o número do registro que está sendo processado e o número total de registros que serão interpretados. Os pontos de entrada e saída também são apresentados com base nas informações descritas pelo usuário no formulário de aplicação de filtros. (Ver Figura 12).

É permitido ao usuário observar os resultados obtidos através de sua representação em gráficos de superfície (Figuras 14, 15 e 16)

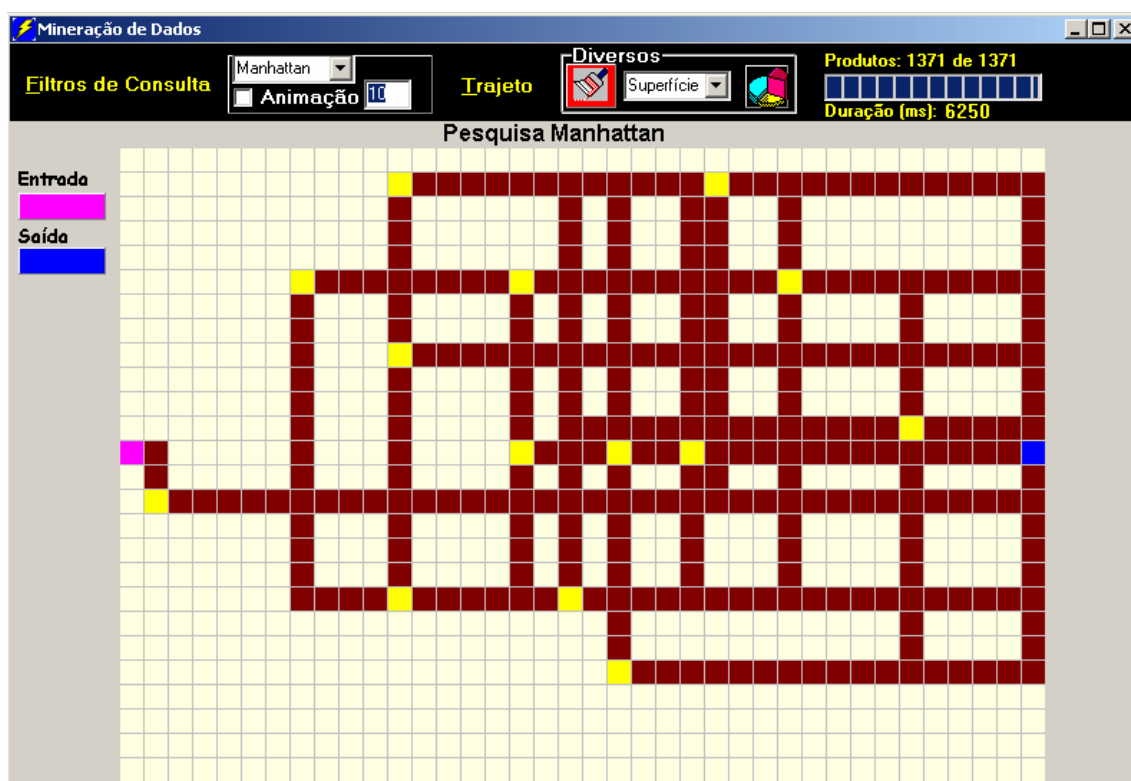


Figura 13 - Formulário de representação dos trajetos definidos pelo simulador.

A definição dos trajetos encontrados pelo simulador permite ao usuário escolher entre três tipos de gráficos de superfície.

O atributo representado no gráfico é o somatório do número de visitas recebidas para cada uma das unidades amostrais existentes e representam o espaço físico da loja destinado ao consumo.

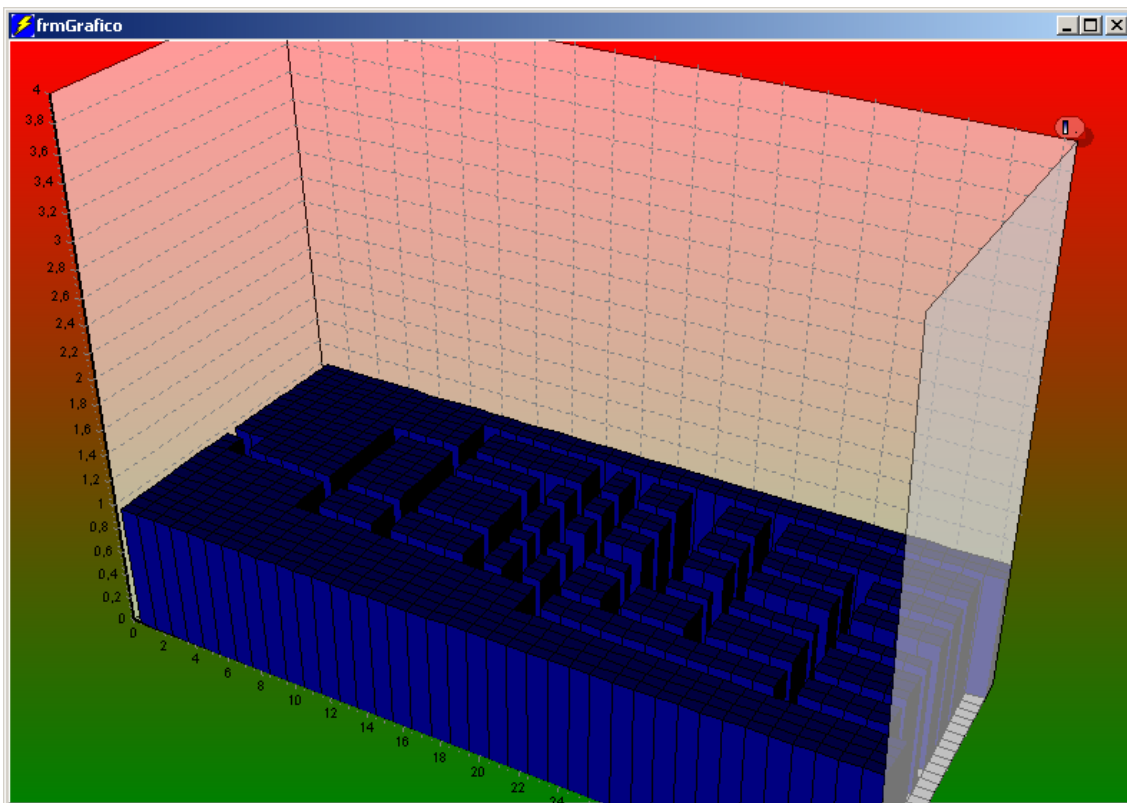


Figura 14 - Gráfico de superfície (negativo).

O gráfico da figura 14 representa a circulação dos clientes no interior do ambiente de consumo, mostrando apenas as áreas visitadas, sem indicar o número de ocorrências. As áreas em baixo-relevo indicam o acúmulo de todos os prováveis trajetos descobertos durante a mineração. Por exemplo, o gráfico da Figura 14 indica que o ponto 1,1 (x,y) não foi visitado durante a compilação dos dados coletados a partir dos filtros indicados pelo usuário.

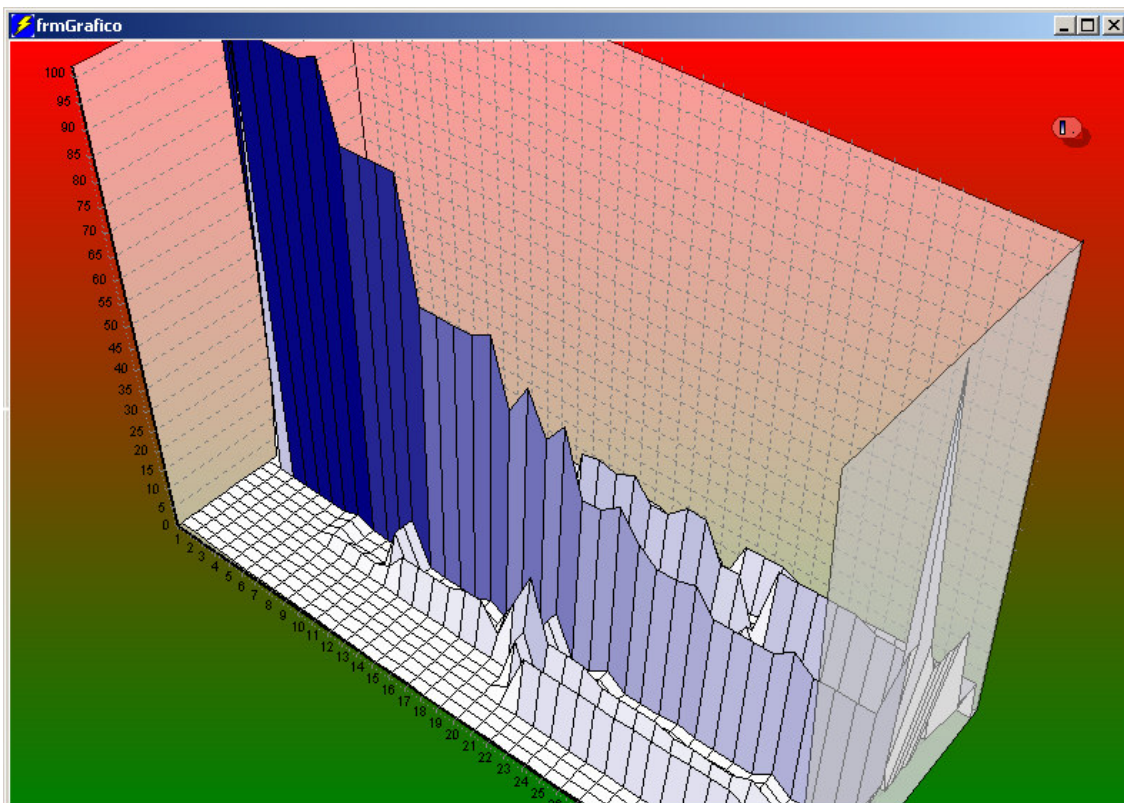


Figura 15 - Gráfico de superfície (acumulativo).

O gráfico da figura 15 representa a circulação, hipotética, de clientes no interior do ambiente de consumo, indicando as áreas visitadas e o número de ocorrências.

A altura do gráfico representa o número de visitas numa determinada unidade amostral. Além da área de consumo é possível visualizar que as unidades amostrais que representam a entrada e a saída no ambiente do consumo terão sempre altura igual ao número de registros processados. Para tanto, o simulador utiliza, como ponto de partida, a entrada e, como ponto de parada, a saída. Os pontos de entrada e saída também são apresentados com base nas informações descritas pelo usuário no formulário de aplicação de filtros. (Ver Figura 12).

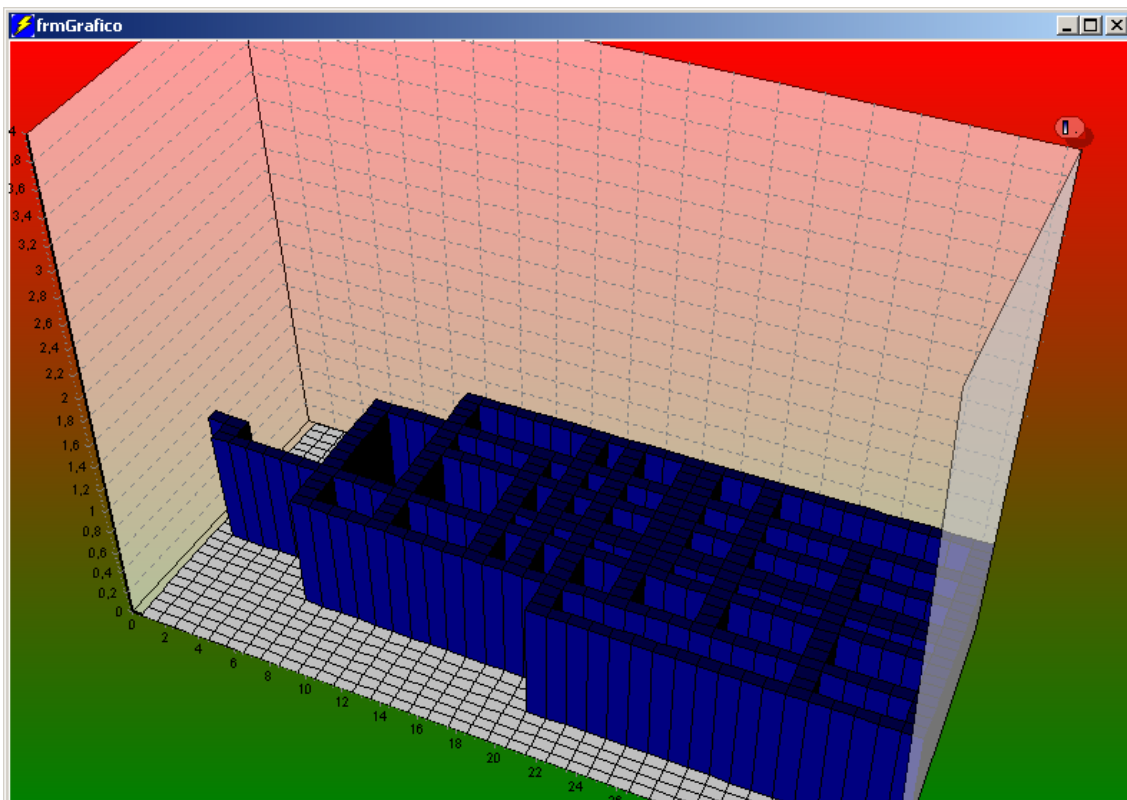


Figura 16 - Gráfico de superfície (positivo).

O gráfico da figura 16 representa a circulação dos clientes no interior do ambiente de consumo, mostrando apenas as áreas visitadas, sem indicar o número de ocorrências. As áreas em alto-relevo indicam o acúmulo de todos os prováveis trajetos descobertos durante a mineração.

Este gráfico exibe os mesmos resultados demonstrados no Gráfico de superfície (negativo), apresentado através da Figura 14, porém com a superfície invertida.

6. CONCLUSÕES

O presente trabalho propôs o desenvolvimento de um conjunto de ferramentas de Mineração de Dados que permitisse a visualização e a identificação de padrões comportamentais baseados no estudo dos prováveis trajetos descritos por clientes em um ambiente de comercialização de produtos selecionados diretamente pelo consumidor (auto-atendimento).

O elevado nível de competição empresarial no segmento varejista impõe a necessidade do desenvolvimento de sistemas que auxiliem nos processos decisórios.

A Mineração de Dados vem ocupando, entre os desenvolvedores de *software*, um espaço cada vez maior. Alguns autores ainda discordam quanto às classificações e processos para cada uma das técnicas, mas é evidente que os chamados sistemas de *Business Intelligence* estão cada vez mais presentes nas empresas do mercado varejista.

Quanto ao cumprimento dos objetivos iniciais, a observação dos gráficos de saída ofereceram a interpretação esperada. Contudo, a validação dos resultados ainda exigirá testes comprobatórios em ambientes reais de consumo.

6.1 Dificuldades encontradas

As discordâncias entre autores e outros especialistas não devem ser tratadas como dificuldades, porém aumentam a complexidade no desenvolvimento do texto. Pode-se citar, como referência, o presente trabalho, por mostrar que alguns autores não admitem a Visualização de Dados como modelo de suporte à Mineração de Dados, enquanto outros a definem, mas como sendo uma técnica.

Discrepâncias como essas tornam árdua a tarefa de oferecer aos leitores um trabalho completo e abrangente. Reafirmações foram freqüentes ao longo deste trabalho, causando a redundância na tentativa de evitar contradições maiores ao longo de seu desenvolvimento.

Os limites, naturalmente impostos para trabalhos desta natureza, impedem que os assuntos sejam abordados por completo em todas as variações apresentadas pelos principais autores.

No campo prático, da implementação dos modelos propostos neste trabalho, podem ser definidas como sendo dificuldades encontradas:

- O fato de a exposição de um mesmo produto ocorrer em várias gôndolas no interior de um supermercado. Com o uso do código de barras, impresso diretamente pelo fornecedor, torna-se muito difícil realizar a diferenciação quanto à origem de sua unidade amostral. Essa dificuldade ocorre quando o mesmo código de barras é utilizado para produtos iguais em localizações distintas no interior da loja.

- A falta de referências bibliográficas, ante o ineditismo deste trabalho. Em grande parte, os estudos mais próximos abordaram questões logísticas, de cadeias de suprimentos (*supply chain*) e de regras de associação (*market basket analysis*).

- A validação dos resultados obtidos frente ao elevado número de recursos humanos necessários para obter dados reais numa amostra representativa.

6.2 Sugestões para Trabalhos Futuros

Uma seqüência natural, presente trabalho, seria a sua implementação em um cenário real, ou seja, o mapeamento de um ambiente de consumo instalado. A validação dos resultados poderia contribuir, em muito, com o prosseguimento e com a evolução dos métodos utilizados.

A presente dissertação também não contemplou a exploração do *checkout* utilizado pelo consumidor. É instigante o desejo de procurar saber se existe relação entre o provável percurso desenvolvido e o *checkout* selecionado. Através do mapeamento e da identificação do *checkout*, seria possível aumentar a precisão do exato ponto de saída do cliente.

Entre as perspectivas mais desejáveis está a busca em profundidade para os casos em que a distância a ser percorrida, partindo do produto A, seja igual, tanto para o produto B quanto para o produto C. Através da utilização da distância de *Manhattan*, o estímulo recebido é igual e a escolha do melhor caminho tende a ser arbitrária. A busca em profundidade poderia suprir, parcialmente, esta arbitrariedade imposta pelo presente modelo. Diz-se “parcialmente”, pois quando de uma compra de muitos produtos, o problema poderá ver crescer o seu grau de dificuldade (espaço de busca), aproximando-se do “Problema do Caixeiro Viajante”, também conhecido, em inglês, como TSP (Traveling Salesman Problem). A pesquisa em profundidade refere-se à seqüência de produtos adquiridos pelo consumidor.

Ainda sobre o tema aqui pesquisado, sugere-se a busca pela descoberta de possíveis relações existentes entre os deslocamentos efetuados e os resultados obtidos através da aplicação de regras de associação (*market basket analysis*).

GLOSSÁRIO

CHECKOUT: Local onde são contabilizados os produtos adquiridos pelo consumidor.

DATA MINING: Mineração de Dados.

DW: *Data Warehouse.*

DSS: *Decision Support Systems.*

ECR: *Efficient Consumer Response.*

EDI: *Electronic Data Interchange.*

KDD: *Knowledge Discovery in Database.*

LOG: Tabela responsável por armazenar os registros de eventos efetuados no sistema.

MD: Mineração de Dados.

OLAP: *On Line Analytical Processing.*

POPAI BRASIL: *Point of Purchase Advertising International.*

SAD: Sistemas de Apoio a Decisão.

SGBD: Sistemas Gerenciadores de Bancos de Dados.

SIG: Sistemas de Informações Gerenciais.

SQL: *Structured Query Language.*

TI: Tecnologia da Informação.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, Márcio Oliveira. **TreeMiner: Uma Ferramenta para Exploração Visual de Dados**. In SQL MAGAZINE, v.10, n.1. Rio de Janeiro: Neoficio Editora, 2004.

AMARAL, Fernanda C. N. **Data Mining: Técnicas e Aplicações para o Marketing Direto**. São Paulo: Berkeley, 2001.

ANSOFF, I. A., & MCDONELL, E.J. **Implantando a Administração Estratégica**. São Paulo: Atlas, 1993.

BARRETO, Jorge Muniz. **Inteligência Artificial no Limiar do Século XXI: Abordagem Híbrida, Simbólica, Conexionista e Evolutiva**. Florianópolis: Duplic, 3. ed., 2001

BERRY, Leonard. **O Modelo Emergente**. HSM Management. Barueri, SP.1999. Disponível em: <http://www.hsmmanagement.com.br/>

BLESSA, Regina. **Merchandising no ponto-de-venda**. São Paulo: Atlas, 2. ed., 2003.

BORGES JR, A.A. **Lay-out e Disposição de Produtos no Varejo: a Influência da Teoria Comportamental**. In: DE ANGELO, C.F.; DA SILVEIRA, J.A.G. **Varejo Competitivo. Programa de Administração de varejo**. São Paulo: Atlas, v.2, cap. 12, p.238 – 251. 1997

BRONDMO, Hans Peter. **Fidelização: como conquistar e manter clientes na era da Internet**. São Paulo: Futura. 2001

CAPURRO, R.1992. **What is Information Science For? A Philosophical Reflection**. In: Vokkary, P. & Cropin, B. Eds: **Conceptions of Library and Information Science**. London. Taylor Graham. p.82-86. Disponível em <http://v.hbistuttgart.de/~capurro/tampere91.htm> Acessado no dia 27/12/2004.

CHIAVENETO, Idalberto. **Administração nos Novos Tempos**. Rio de Janeiro: Campus. 1999

CHRISTOPHER, M. **Logística e gerenciamento da Cadeia de Suprimentos – estratégias para a redução de custos e melhoria dos serviços**. São Paulo: Ed. Pioneira, 1997.

CORSTJENS, J. & CORSTJENS, M. **Store wars, the battle for mind space and shelf space**. England: John Wiley & Sons. 1995

CURHAN, R.C. **The relationship between shelf space and unit sales in supermarkets**. *Journal of Marketing Research*. Vol.9, nº 3. 1972

DAVENPORT, T. H. **Ecologia da informação: porque só a tecnologia não basta para o sucesso da informação.** São Paulo: Futura. 1998.

DAVENPORT, T. & PRUSAK, L. **Conhecimento empresarial.** Rio de Janeiro: Campus. 1999

DAWKINS, Richard. **O gene egoísta.** Rio de Janeiro: Itatiaia, 2001.

FAHEY, Lian, RANDALL, Robert M. **MBA Curso Prático Estratégia.** Rio de Janeiro: Ed. Campus, 1999.

FAYYAD, U., GRINSTEIN, G. G. & WIERSE A. **Information Visualization in Data Mining and Knowledge Discovery.** Morgan Kaufmann Publishers 2002.

FAYYAD, U., PIATETSKY-SHAPIRO, G. & SMYTH, P. **From Data Mining to Knowledge Discovery: An Overview In: Advances in Knowledge Discovery and Data Mining**, pp. 11-34 Menlo Park, CA: AAAI Press / The MIT Press. 1996a.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; & SMYTH, P. **The KDD Process for Extracting Useful Knowledge from Volumes of Data.** Communications of the ACM – Data Mining, 39(11). 1996b.

FAYYAD, U. & UTHURUSAMY, R. **The Data Mining and Knowledge Discovery in Databases.** Communications of the ACM, Vol. 39, n. 11. 1996, p. 24-26

FERNANDES, S. **Comunicação Visual em Supermercados e Hipermercados de São Paulo: um Estudo de Caso.** Dissertação (mestrado) Escola de Comunicação e Artes, Universidade de São Paulo. São Paulo, 1998.

FLEURY, Paulo Fernando; WANKE, Peter; FIGUEIREDO, Kleber Fossatti. **Logística empresarial: a perspectiva brasileira.** São Paulo: Atlas, 2000.

FREITAS, Paulo José Filho. **Introdução à Modelagem e Simulação de Sistemas – com Aplicações em Arena.** Florianópolis: Visual Books Ltda., 2001

GHODDOSI, Nader. **Sistema de Informação Estratégico para o Gerenciamento Operacional (SIEGO) Como Uma Ferramenta de Apoio de Decisão Utilizando Data Warehouse.** Florianópolis. Tese de Mestrado em Ciência da Computação. Universidade Federal de Santa Catarina, 2003.

GROTH, Robert A. **Hands-On Approach for Business Professionals.** New Jersey: Prantice Hall PTR, 1998.

HAN, J, & KAMBER, M. **Data Mining – Concepts and Techniques.** New York: Morgan Kaufmann, 2000.

- HARISON, Thomas H. **Intranet Data Warehouse**. São Paulo: Berkeley Brasil, 1998.
- INMON, William H. **The Data Mining and Knowledge Discovery in Databases**. Communications of the ACM, Vol. 39, n. 11. 1996, p. 49-50
- INMON, William H. **Como construir o Data Warehouse**. Rio de Janeiro: Campus, 1998.
- KOTLER, Philip. **Administração de Marketing**. São Paulo: Prentice Hall.2000.
- LEVY, Michel; WEITZ, Barton. **Retailing Management**. 2nd edition, Chicago (USA), Irwin, 1995.
- MATHEWS, R. **An European Case Study: A Competitive Edge**. Progressive Grocer, New York, 1995
- MCGEE, James; PRUSAK, Laurence. **Managing information strategically**. New York-USA: John Wiley & Sons.1993
- MIRANDA, R. C. da R. **O uso da informação na formulação de ações estratégicas pelas empresas**. Brasília, v.28, n.3, 1999.
- NETO, M.G.M. *et al.* **Aplicação de Técnicas de Mineração Visual de Dados na regulação da Indústria de Energia: Um Estudo de Casos**. Congresso Brasileiro de Regulação de Serviços Públicos Concedidos. Anais. Salvador (Bahia), 2000.
- OLIVEIRA, Djalma de Pinho Rebouças de. **Sistemas de Informações Gerenciais**: 9. ed. São Paulo: Atlas, 2004.
- PIATETSKY-SHAPIRO, G. **Knowledge Discovery in real databases: A report on the IJCAI-89**. AI Magazine, Vol. 11, N. 5, Jan. 1991.
- RAPHEL, Murray. **A Escala da Lealdade**. HSM Management. Barueri(São Paulo): 1999. Disponível em: <http://www.hsmmanagement.com.br> Acessado no dia: 06/12/2004
- REZENDE, S. O (Coordenadora). **Sistemas Inteligentes – Fundamentos e Aplicações**. Barueri: Manole, 2003.
- RUSSEL, G. J.; KAMAKURA W. A. **Modeling multiple category brand preference with household basket data – Journal of Retailing**, v.73, n.4, p.439-61, 1997.
- SETZER, V. W. **Meios Eletrônicos e Educação: uma visão alternativa**. São Paulo: Escrituras, 2001.
- SINGH, Harry S. **Conceitos, Tecnologias, Implementação e Gerenciamento**. São Paulo: Makron Books, 2000.

SIRIHAL, A. B. LOURENÇO, C. A. **Informação e Conhecimento: aspectos filosóficos e informacionais.** 2003. Disponível em: <http://www.informacaoesociedade.ufpb.br/1210203.html> Acessado no dia: 06/12/2004

TAKAOKA, Hiroo. **Marketing de Relacionamento no Varejo.** São Paulo: Atlas, 1999.

THURASINGHAM, B. **Data Mining: Technologies, Techniques, Tools and Trends.** CRC Press, 1999.

TURBAN, E, RAINER, R. K., POTTER, E. R. **Administração de Tecnologia da Informação: Teoria e prática.** Rio de Janeiro: Campus, 2003.

VIANA, Reinaldo. **Mineração de Dados: Introdução e Aplicações.** In SQL MAGAZINE, v.10, n.1. Rio de Janeiro: Neoficio Editora, 2004.

WEISS, S.M. & INDURKHYA, N. **Predictive Data Mining: A Practical Guide.** Morgan Kaufmann Publishers. San Francisco, CA. In: REZENDE, Solange 1998.

WITTEN, I H. & FRANK, E. **Data mining: Pratical Machine Learning Tools and Techniques With Java Implementations.** London: Academic Press, 2000.