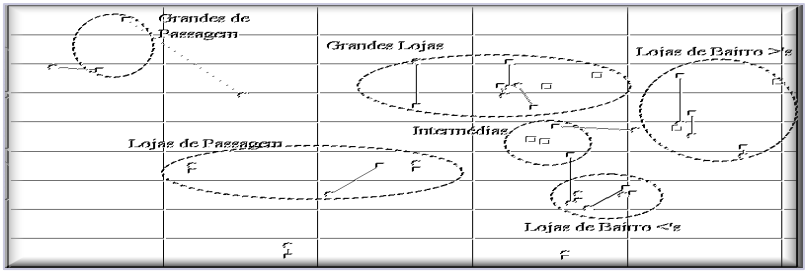
UNIVERSIDADE TÉCNICA DE LISBOA
INSTITUTO SUPERIOR TÉCNICO



Modelação de Vendas de Novas Superfícies Comerciais

Armando Brito Mendes
(Mestre)

Dissertação para obtenção do Grau de Doutor em Engenharia de Sistemas

Orientador: Doutor Rui Manuel Moura de Carvalho Oliveira
Co-Orientadora: Doutora Maria Margarida Guerreiro Martins dos Santos Cardoso

Júri

Presidente: Reitor da Universidade Técnica de Lisboa
Vogais: Doutor Luiz Abel Magro Moutinho
Doutor Luís António de Castro Valadares Tavares
Doutor Rui Manuel Moura de Carvalho Oliveira
Doutor João Agostinho de Oliveira Soares
Doutora Maria Margarida Guerreiro Martins dos Santos Cardoso

Lisboa, Outubro de 2005

Modelação de Vendas de Novas Superfícies Comerciais

RESUMO:

Os retalhistas sempre entenderam a localização como um factor crítico do sucesso de uma nova loja. No entanto, tentar perceber todos os aspectos da localização, potencial da área de influência e comportamentos do consumidor pode revelar-se uma tarefa de elevada complexidade. Nesta dissertação apresenta-se uma metodologia de apoio à decisão na avaliação de localizações potenciais de lojas de retalho alimentar de pequena a média dimensão, com base em modelos de previsão de vendas.

A recolha de dados necessários à modelação inclui dois inquéritos a clientes e um programa de *mystery shopping*. Utilizam-se diagramas de Voronoi multiplicativos ponderados no tratamento espacial de dados demográficos do censo 2001. Descreve-se o estado da arte relativamente a modelos e métodos utilizados em problemas semelhantes e sugere-se uma classificação com três classes.

É definida uma tipologia de lojas com base na comparação de três métodos de integração de conhecimento de especialistas: *a priori*, *a posteriori* e interactivo. Induzem-se regras proposicionais para classificar uma nova localização num dos grupos de lojas análogas. Após a classificação de uma localização potencial utiliza-se um modelo de regressão linear para prever vendas. Os modelos são implementados numa folha de cálculo segundo uma filosofia *loosely coupled*.

A integração de conhecimento de domínio por parte de especialistas, nos modelos construídos, e a geração de novo conhecimento sobre o problema são elementos estruturantes de todo o trabalho apresentado.

PALAVRAS-CHAVE:

Avaliação de Localizações Potenciais em Cadeias de Retalho; Diagramas de Voronoi Multiplicativos Ponderados; Segmentação de Lojas; Integração de Conhecimento de Especialistas; Árvores de Classificação; Geração de Conhecimento

New Food Store Turnover Modelling

ABSTRACT:

The retailers always understood the location as a critical success factor for a new store. However, recognizing all the aspects of location, influence area potential, and consumer's behaviour presents high complexity. In this dissertation, it is described an approach for site selection and evaluation of potential locations of food stores of small to medium size by sales turnover forecast.

Data gathering included two customers' surveys and a mystery shopping program. Multiplicative Weighted Voronoi Diagrams were used in spatial demographic data analysis. The state of the art of models and methods used in similar problems are described and a classification is suggested in three classes.

A typology of stores is defined by comparing three methods of expert knowledge integration: a priori, a posteriori and the interactive method. Decision rules are induced to classify a new location in one of the previous groups of analogue stores. After this classification, a linear regression model is used to forecast store turnover. The models are implemented in a spreadsheet using a loosely coupled philosophy.

The integration of domain knowledge in the models by expert's and the creation of new knowledge about the problem, were the main guiding principles in all the work presented.

KEYWORDS:

Site Selection for Food Retail Outlet; Multiplicative Weighted Voronoi Diagrams; Supermarket Segmentation; Expert Knowledge Integration; Classification Trees; Knowledge Creation

Agradecimentos

O autor não pode começar sem relembrar a Professora Isabel Hall Themido, responsável pela ideia inicial, verdadeira força por detrás de todo o trabalho desenvolvido e inspiradora de todos os que com ela colaboraram. Este trabalho é uma pequena homenagem à memória da Professora Isabel Hall Themido.

Ao Professor Rui Oliveira por ter aceite um trabalho a meio numa área não completamente coincidente com os seus interesses científicos, por todo o apoio e amizade demonstrados ao longo do trabalho. Agradeço ainda o indispensável incentivo e as leituras atentas do texto apresentado.

À Professora Margarida Cardoso por uma colaboração atenciosa e muito próxima e pela permanente disponibilidade. A Professora Margarida Cardoso colaborou neste trabalho além de orientar, sendo nomeadamente responsável pela segmentação de clientes de ambos os inquéritos efectuados, além de outras contribuições.

O autor agradece a colaboração dos especialistas da cadeia de lojas que foram incansáveis na satisfação dos pedidos sucessivos de obtenção de dados e na avaliação dos resultados. Ainda que o interesse por este trabalho dentro do grupo de distribuição não tenha sido sempre o mesmo, a verdade é que a amabilidade e atenção dispensada por estes profissionais foi sempre muito activa. Esta dissertação teria sido impossível sem a sua atenciosa e amiga colaboração e é em grande parte resultado de um trabalho conjunto.

Ao CESUR \ IST por ter aceite e apoiado este projecto e ao ICIST \ IST, na pessoa do Dr. Alexandre Gonçalves e do Professor João Matos, pela amizade e colaboração prestada. Nomeadamente o ICIST foi responsável pelo levantamento das coordenadas das lojas por GPS e pela programação dos algoritmos utilizados na delimitação de áreas de influência por diagramas de Voronoi multiplicativos.

Ao Professor Luís Cavique pelo apoio e incentivo, à Dra. Ana Amorim pela colaboração no tratamento dos inquéritos e dos dados do programa *mystery shopping*. À Dra. Paula Cunha e Dra. Patrícia por toda a atenção dispensada e indispensável apoio logístico.

À Universidade dos Açores e em especial aos colegas do Departamento de Matemática pela compreensão demonstrada e por terem criado as condições para que este trabalho fosse possível.

Por fim, a todos os amigos que leram a presente dissertação e a criticaram.

Índice Temático

Índice de Figuras

Índice de Tabelas

Esta dissertação é dedicada à Sandra e à Inês

«Eu sou o resultado consciente da minha própria experiência»

José Almada Negreiros

"*Ultimatum* Futurista", publicado em Lisboa, Dezembro 1917

Nota Prévia

Este trabalho foi realizado em estreita colaboração com um grupo de distribuição alimentar nacional preocupado em aumentar o número de lojas pertencentes a uma cadeia de Supermercados de Proximidade.

Esta colaboração foi indispensável na recolha dos dados e na crítica de resultados. Na maioria das actividades realizadas, este grupo esteve profundamente envolvido, inclusivamente intervindo activamente em todas as fases do projecto e expondo os seus pontos de vista e opiniões, baseadas no extenso conhecimento do domínio que detêm. Esta constante interacção foi, na nossa opinião, o segredo do sucesso da implementação dos modelos desenvolvidos.

No entanto, no âmbito desta colaboração foram impostas restrições à revelação de algumas informações sobre os dados recolhidos. Nomeadamente, não é possível revelar nem a cadeia de lojas envolvidas no estudo, nem o grupo de distribuição com o qual se trabalhou. Igualmente não é possível mostrar mapas com a localização geográfica das lojas. Também não se revelam os valores de vendas por loja pelo que todos os valores relacionados, como desvios e parâmetros dos modelos, foram obtidos a partir de valores modificados. Também não se podem revelar nomes de lojas, mas a denominação apresentada é coerente em todo o texto da dissertação.

Assim, nesta dissertação, estas restrições são cuidadosamente seguidas a fim de não trair a confiança de quem tão amavelmente connosco colaborou. No entanto, as referidas restrições podem levantar problemas de reprodutibilidade dos resultados que se tenta minimizar ao apresentar dados agregados e/ou modificados. De qualquer modo, tem-se a preocupação de que tais restrições não afectem o rigor dos resultados apresentados.

Notação Matemática e Abreviaturas

Notação Matemática

α, β	parâmetros da expressão para o índice de precisão (IP_j);
a	índice identificativo da árvore de classificação;
A_j	atractividade gerada pelo ponto de venda j;
$aInfl_j$	área de influência definida por algoritmos de caminho mais curto para a loja j em hectares;
a_r	índice identificativo da regra proposicional (ou nó folha) r referente à árvore de classificação a;
$aVend_j$	área de vendas em metros quadrados para a loja \ localização potencial j;
B_{03j}	ordenada na origem da equação de previsão para as vendas da loja j no ano de 2003;
$dEdif_j$	densidade de edifícios construídos entre os anos de 1996 e 2001 em número de edifícios por 10 hectares para a área de influência definida por algoritmo de caminhos mais curtos a 2,5 minutos;
$d_{ij}=\|\|x_i-x_j\|\|$	distância, tempo ou custo de deslocação entre o polígono de procura i e o ponto de venda representando a oferta j;
d_{wj}	função de distância ponderada pelo peso w_j relativa ao ponto de venda j;
E_i	vendas potenciais provenientes da subzona i;
h	índice identificativo das n lojas em concorrência numa determinada região;
i	índice identificativo do polígono resultante da divisão da área de influência em subzonas homogéneas nos modelos gravitacionais;
IP_j	Índice de Precisão para o ponto de venda j;
j	índice identificativo do ponto de venda ou loja;
k	número de pontos de venda frequentados pelos clientes em simultâneo correspondendo igualmente à ordem dos diagramas de Voronoi;
$l={}^nC_k$	número de combinações de k pontos geradores no total de n pontos correspondente ao número de subconjuntos em P;
n	número finito de pontos no espaço associados a lojas, para gerar um diagrama de Voronoi é necessário um número mínimo de dois pontos;
$n_g^{a_r}$	número de observações no nó folha a_r pertencente ao grupo g;
$nAloj_j$	número de alojamentos com proprietário ocupante para a área de influência da loja j definida por diagramas de Voronoi de 1ª ordem;
$P=U_iP_i(k)$	conjunto de subconjuntos de k pontos geradores, para $k = 1$ reduz-se ao conjunto de pontos gerador dos diagramas de Voronoi simples;
$P_i(k)$	subconjunto i de k pontos geradores dum polígono de Voronoi de ordem k;

p_j	localização no espaço do ponto de venda j;
S_{ij}	fracção do potencial de vendas (ou quota de mercado) da zona i captada pelo ponto de venda j;
U_{ij}	função utilidade genérica entre a oferta do ponto de venda j e a procura proveniente do polígono i;
$V = \{V(p_1), V(p_2), \ldots, V(p_n)\}$	diagrama de Voronoi constituído pelo conjunto dos polígonos correspondentes a todos os pontos geradores de P;
$V(p_j)$	polígono de Voronoi gerado pelo ponto p_j;
$V(P_i(k))$	polígono de Voronoi multiplicativo de ordem k gerado pelo subconjunto i de k pontos geradores $P_i(k)$;
$\hat{W}_{03j}$	vendas anuais previstas para a loja j e para o ano de 2003;
w_j	peso superior a zero associando ao ponto de venda j;
x_j	coordenadas do ponto p_j;

Abreviaturas e Acrónimos

ADO	*ActiveX Data Objects*;
AHP	*Analytical Hierarchy Process* (processo hierárquico analítico);
AID	*Automatic Iteration Detector*;
ANOVA	*ANalysis Of VAriance* (análise de variância);
APAV	Análise e Previsão por Analogia de Vendas;
APED	Associação Portuguesa de Empresas de Distribuição;
APSI	Associação Portuguesa de Sistemas de Informação;
CART	*Classification And Regression Trees* (árvores de classificação e regressão);
CHAID	*Chi-square Automatic Interaction Detection*;
CMC	Algoritmo de Caminhos Mais Curtos sobre uma rede viária;
DDE	*Dynamic Data Exchange*;
DfBetas	Medida da variação dos coeficientes estimados por regressão atribuída a uma observação eliminada;
Eurostat	*STATistical office of the EUROpean communities* (agência de informação estatística da Comunidade Europeia);
ERP	*Enterprise Resource Planning*;
GIS	*Geographical Information System* (ver SIG);
GPS	*Global Positioning System* (sistema de posicionamento global);
HTML	*Hyper Text Markup Language;*
INE	Instituto Nacional de Estatística;
KBDSS	*Knowledge Based Decision Support Systems* (sistema de apoio à decisão baseado em conhecimento);

MCI	*Multiplicative Competitive Interactive model*;
MC-SDSS	*MultiCriteria Spatial Decision Support System* (sistema de apoio à decisão espacial multicritério);
MDS	*MultiDimensional Scaling*;
MNL	*MultiNomial Logit*;
MULTILOC	*MULTIple store LOCation model*;
MWVD	*Multiplicative Weighted Voronoi Diagrams* (diagramas de Voronoi multiplicativos ponderados);
OkMWVD	*Order* k MWVD (polígonos de Voronoi multiplicativos de ordem k);
OLE	*Object Linking and Embedding*;
OVD	*Ordinary Voronoi Diagram* (diagrama de Voronoi simples ou de primeira ordem);
PRESS	*PREdicted Sum of Squares*;
POS	*Point Of Sale* (ponto de venda);
QFD	*Quality Function Deployment;*
QUEST	*Quick Unbiased Efficient Statistical Tree*;
SAD	Sistema de Apoio à Decisão;
SDSS	*Spatial Decision Support Systems* (sistemas de apoio à decisão espacial ou geográfica);
SGBDOO	Sistemas Gestores de Bases de Dados Orientadas para Objectos;
SGBDR	Sistemas Gestores de Bases de Dados Relacionais;
SIG	Sistema de Informação Geográfica;
SLAM	*Store Location Assessment Model*;
UCDR	Unidades Comerciais de Dimensão Relevante;
VBA	*Visual Basic for Applications*;
WWW	*World Wide Web;*
XML	*eXtensible Markup Language*.

Formatações e Destaques

Itálico	destaca palavras ou expressões em língua estrangeira incluindo expressões em latim;
"Aspas"	destaca nomes de variáveis e expressões ou palavras que não devem ser confundidas com o texto;
Iniciais Maiúsculas	além da utilização habitual é também utilizado para realçar alguns nomes de grupos evitando o cansaço do excesso de aspas;
Carregado	destaca expressões e palavras que resumem o(s) parágrafo(s) ou termos definidos ou explicados nas linhas seguintes;
Times itálico	símbolos em notação matemática.

Capítulo I

INTRODUÇÃO

Este capítulo descreve, em traços largos, o contexto em que surge o problema e o ambiente vivido na distribuição em geral, sendo este tema mais extensivamente explorado no segundo capítulo. Descrevem-se ainda aspectos fundamentais para compreender esta dissertação como a motivação, o problema em estudo e os objectivos a atingir. Pretende-se demonstrar a necessidade de criação de modelos de apoio à decisão para localização de lojas de retalho alimentar de pequena a média dimensão por modelação de vendas em novas localizações. Faz-se igualmente uma descrição da estrutura da dissertação apresentada e discutem-se diferenças de nomenclatura entre as disciplinas de estatística, reconhecimento de padrões e análise de *marketing*.

> *«… new trends in retailing, commercial real estate development, and competitive forces require a new level of sophistication concerning where to best market a product or service»*
>
> Joseph R. Bagby
>
> (fundador da NACORE – *iNternational Association of COrporate Real Estate executives*, prefácio de Salvaneschi, 1996)

I.A. A Loja de Retalho e o Problema de Localização

O sector da distribuição tem vindo a ser dividido em dois subsectores de actividade muito interligados: o subsector retalhista e o grossista. Na verdade, esta divisão é artificial e resulta da cobertura de diferentes conjuntos de elos da cadeia logística. O grossista trataria dos primeiros elos da cadeia e o retalhista do contacto directo com o consumidor. As actividades e o tipo de negócio distinguem-se essencialmente por o subsector grossista ser do tipo *business to business* e o retalhista

do tipo *business to consumer*, estando na origem das respostas diferenciadas para as variáveis do *marketing mix* encontradas para cada subsector.

No entanto, as actividades básicas de transporte, gestão de inventários (*stocks*), divisão em quantidades apropriadas, transmissão de informação e serviços são muito semelhantes, pelo que a integração vertical da cadeia logística surgiu naturalmente tendo por consequência o desenvolvimento de grupos de distribuição com várias insígnias e cadeias de retalho. Uma **cadeia de retalho** pode ser definida como um conjunto de pontos de venda detidos pelo mesmo grupo de distribuição, com níveis de decisão comuns e uma logística integrada (Levy e Weitz, 2004).

A preocupação fundamental dos grupos de distribuição e, em geral, de todos os elos da cadeia de distribuição é a satisfação das necessidades do cliente, incluindo a criação de novas. Esta orientação para o cliente está no centro dos actuais conceitos de *marketing*[1] relacional (Gilbert, 2002), mas também da **logística empresarial** *(business logistics*). Por exemplo, uma das definições apresentadas por Ballou (2004) e atribuída ao *Council of Logistics Management*[2] coloca claramente toda a cadeia logística ao serviço do consumidor (pág. 4):

> *«Logistics Management is that part of Supply Chain Management that plans, implements, and controls the efficient, effective forward and reverse flow and storage of goods, services and related information between the point of origin and the point of consumption in order to meet customers' requirements».*

Assim, a loja de retalho adquiriu nos últimos anos uma relevância acrescida, podendo-se afirmar que quem controla o ponto de venda controla igualmente toda a cadeia logística já que os restantes elos da cadeia ficam dependentes do retalhista para chegarem ao consumidor (Levy e Weitz, 2004 e Rousseau, 1997). Apesar desta preponderância, os pontos de venda também estão sujeitos a fortes pressões. Pressões horizontais que provêm de outras cadeias semelhantes, num mercado que na maioria dos países é já muito saturado, e verticais provenientes de novas formatos de retalho como as vendas directas por catálogo ou o comércio electrónico.

Ao nível do retalho alimentar a variedade de formatos, de marcas e de insígnias demonstra bem a competitividade do sector. Na Figura 1 sugere-se uma tipologia de pontos de venda alimentares baseada em duas dimensões: preço \ nível de serviço e

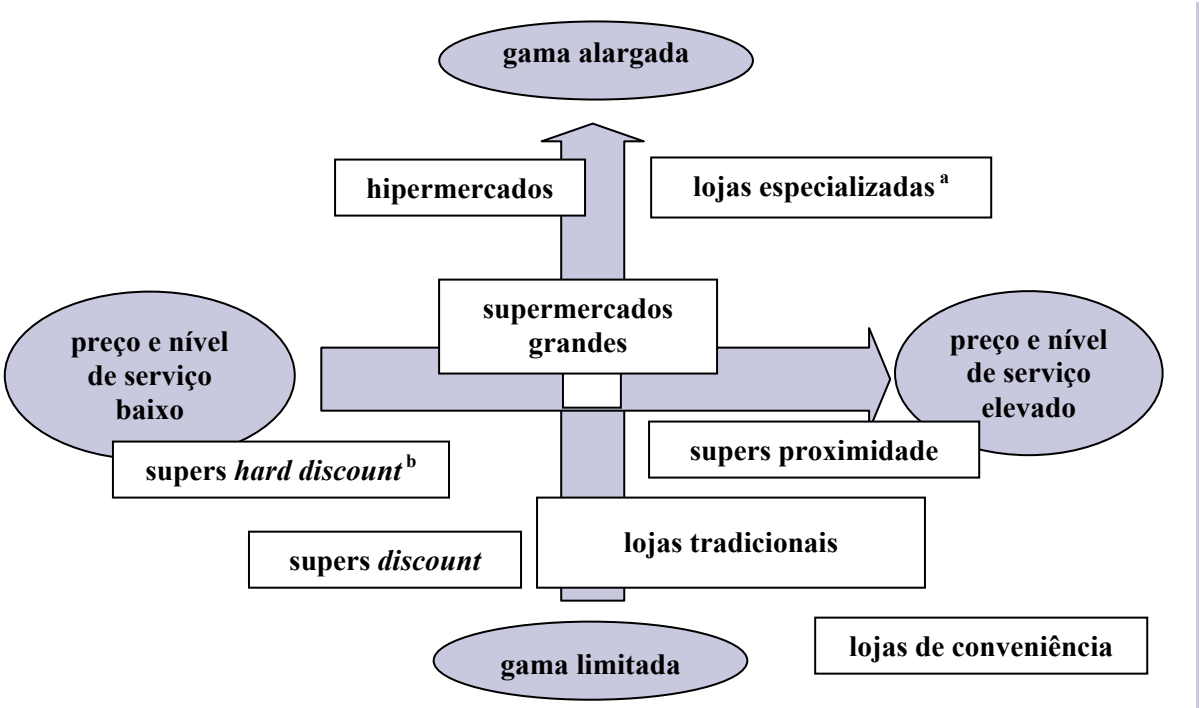
---

[1] Utiliza-se o termo *marketing* da língua inglesa ainda que a Diciopédia 2005 em DVD da Porto Editora (ISBN 972-0-65258-6) recomende o termo "mercadologia" que, no entanto, é pouco utilizado.

[2] Trata-se de uma associação profissional de gestores logísticos, educadores e profissionais com o objectivo de investigação, educação e promoção de troca de informações e conhecimento no domínio da logística, fundada em 1962. Outras informações podem ser consultadas no site clm1.org.

profundidade \ largura ou alcance da gama. A largura ou alcance da gama refere-se ao número de produtos disponíveis e a profundidade ao número de marcas de cada produto. Sublinhe-se no entanto que o posicionamento dos pontos de venda depende, em grande parte, da gestão local e do ambiente competitivo.

FIGURA 1 TIPOLOGIA DE PONTOS DE VENDA DE RETALHO ALIMENTAR SEGUNDO PREÇO E GAMA.
(Fonte: esquema reformulado a partir de uma ideia original de Rousseau, 1997)



[a] Note-se que a gama alargada das Lojas Especializadas se refere à profundidade da gama e não à sua largura. [b] Supers de *Hard Discount* apresentam uma gama de profundidade muito limitada ainda que a largura possa ser elevada.

Normalmente, o alcance da gama acompanha a profundidade da gama. Excepções são as **Lojas Especializadas** (as de maior dimensão também são chamadas de *category killers*) onde apenas se vende uma categoria de produtos normalmente com enorme profundidade de gama. No outro extremo temos os **Supermercados *Discount* e *Hard Discount*** caracterizados por profundidades de gama quase nulas, quase sempre só apresentando uma marca branca para cada tipo de produto, e níveis de serviço reduzidos ao mínimo. Exemplos de insígnias são para os supermercados *Discount* Dia \ Minipreço e para os *Hard Discount* Lidl e Plus.

Os **Hipermercados** são as maiores superfícies comerciais, correspondendo nos termos do decreto-lei nº 83/95 de 26 de Abril, aos estabelecimentos com área de exposição e vendas igual ou superior a 2.000 m^2 ou, no caso de estarem localizados em concelhos com menos de 30.000 habitantes, igual ou superior a 1.000 m^2. Estas lojas apresentam gamas tanto alargadas como profundas tanto em secções alimentares como

não alimentares, ainda que se verifique uma tendência recente de abertura de lojas especializadas que retiram do hipermercado parte da área não alimentar.

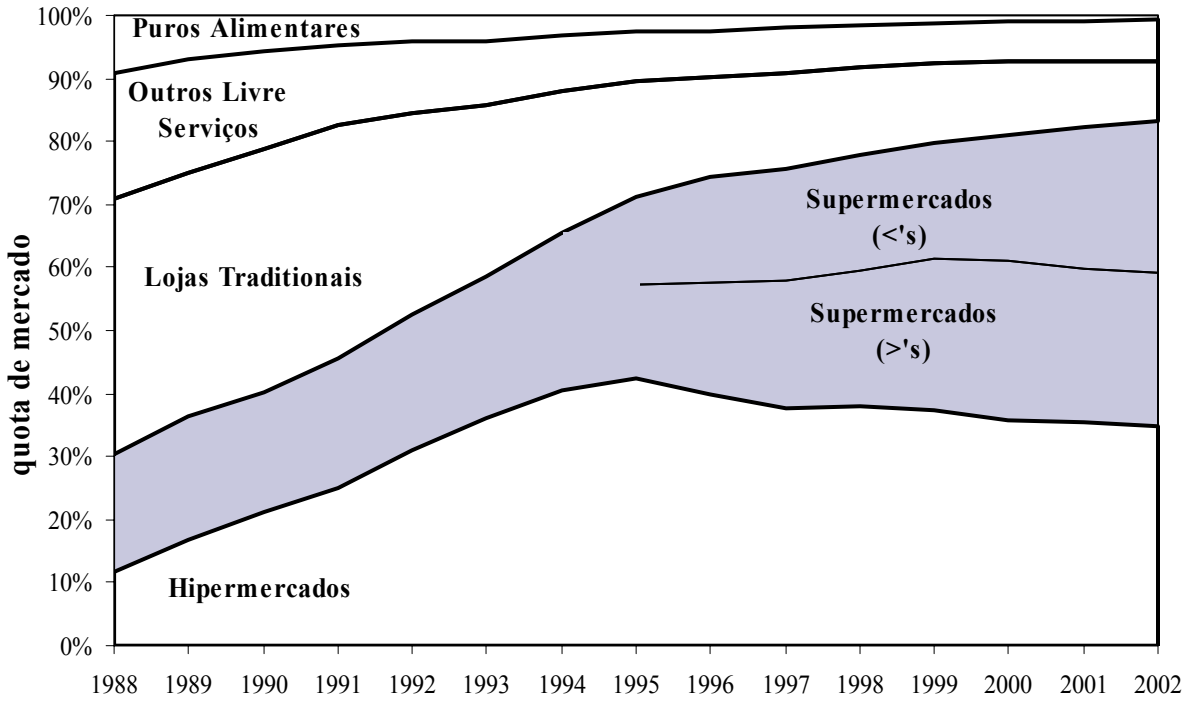
Na Figura 1 os **Supermercados Grandes** referem-se a lojas de dimensões intermédias (entre os hipermercados e os supermercados de proximidade), normalmente situadas fora dos centros das grandes cidades mas não fora da cidade. Como exemplo podem-se citar insígnias como Modelo e Intermarché. Pelo contrário, as **Lojas de Conveniência** situam-se quase exclusivamente em áreas de abastecimento de combustíveis ou dentro das grandes cidades, apresentam dimensões reduzidas mas um nível de serviço muito elevado, sendo caracterizadas por estarem abertas durante períodos muito alargados que podem chegar às 24 horas.

Por fim os **Supermercados de Proximidade** são a categoria mais difusa e com menos insígnias em Portugal mas que a Tesco Metro é um bom exemplo no Reino Unido e os Pingo Doce menores um exemplo nacional. Este tipo de lojas pretende oferecer uma alternativa de qualidade ao cliente evitando deslocações aos supermercados maiores com uma grama de produtos limitada mas com as marcas mais procuradas, com ênfase nos produtos frescos e elevados níveis de serviço. As **Lojas Tradicionais** são uma categoria mal definida de lojas de dimensões muito variáveis, se bem que em média sejam muito pequenas. A principal característica é a de não se integrarem em cadeias de retalho ainda que grande parte participe em algum tipo de associação de distribuição.

Note-se que esta tipologia não é consensual e, por exemplo, AC Nielson acrescenta a categoria de “puros alimentares” a qual é englobada na Figura 1 pelas Lojas Especializadas e divide as Lojas Tradicionais em “drogarias” e “mercearias”. Também a categoria de Lojas de Conveniência é englobada pela AC Nielson no grupo das “outros livre serviços”. Por outro lado, os “supermercados menores” incluem os Supermercados de Proximidade, os *Discount* e os *Hard Discount*.

No mercado Português, e desde que se dispõem de dados sobre a quota de mercado, os hipermercados e os supermercados têm crescido continuamente à custa dos restantes conceitos. De acordo com os dados da AC Nielsen Portugal, supermercados e hipermercados são hoje as estruturas comerciais mais importantes em Portugal Continental, se considerarmos o volume de vendas como indicador de referência. Recentemente os supermercados estão a superar os hipermercados em várias rubricas, tendo-se mesmo registado um crescimento acumulado nas suas vendas superior a 120%, entre 1990 e 1997. A partir desta data a quota de mercado dos supermercados superou a das grandes superfícies e tem crescido de forma sustentada, como se pode observar na Figura 2.

FIGURA 2 EVOLUÇÃO DA QUOTA DE MERCADO POR TIPO DE LOJA EM PORTUGAL.
(Fonte: AC Nielsen Portugal publicado na revista Distribuição Hoje, suplemento Atlas da Distribuição 2004)



Aliás, no ano de 1996, as pequenas e médias superfícies de retalho foram as únicas a registar um crescimento simultaneamente no número de lojas e no volume de vendas (aproximadamente mais 92 milhões de contos) e consequentemente a aumentar a sua quota de mercado de 28 para 34% no universo Nielsen. Em 1997 os Supermercados atingiram a liderança e em 1998 consolidaram a sua estratégia de expansão, em especial os supermercados de menores dimensões.

Como já foi notado os valores para os supermercados menores incluem vários formatos como as lojas *Discount* e *Hard Discount*, que têm ganho muita quota de mercado nos últimos anos. No entanto, também as lojas pequenas e de média dimensão dirigidas para classes mais elevadas, *i.e.* os Supermercados de Proximidade, têm tido importantes ganhos de quota como os lucros da cadeia Pingo Doce comprovam. Segundo o relatório anual da empresa, as vendas do Pingo Doce subiram 2% no ano transacto, apesar do enquadramento macroeconómico adverso e da crescente agressividade concorrencial. Este aumento é atribuído a uma generalizada redução de preços como reacção ao crescimento dos Supermercados *Hard Discount*[3]. Aliás, mais recentemente, a empresa decidiu concentrar o negócio em menos regiões geográficas e

[3] Informação retirada do sítio da empresa: www.jeronimomartins.pt em 18/11/2004.

em menos formatos de retalho, prevendo despender 140 a 150 milhões de euros ano, apostando na expansão e remodelação da rede de supermercados e no retalho especializado[4]. Assim, o futuro dos Supermercados de Proximidade parece promissor.

Ainda que o investimento inicial neste tipo de lojas de pequena a média dimensão seja reduzido, têm-se verificado cuidados especiais na localização destas lojas. Uma boa localização atrai mais consumidores, pelo que aumenta as vendas potenciais. No entanto, estes investimentos podem ser difíceis de rentabilizar, já que implicam retornos de investimento a mais longo prazo relativamente às lojas de maiores dimensões, devido ao fraco poder de atracção das lojas e principalmente, menores economias de escala com cadeias logísticas mais complexas e extensas (Birkin *et al.*, 2002 e Salvaneschi, 1996). Ver por exemplo o caso de um dos retalhistas mais inovadores no Reino Unido descrito em Smith (2004).

As pressões que as cadeias de lojas de distribuição alimentar enfrentam são tais que as decisões de localização não podem ser negligenciadas. As lojas representam locais onde volumes significativos de capital são investidos. Uma vez tomadas as decisões de localização são difíceis de alterar. Deste modo, as empresas não podem continuar a tomar decisões quanto ao quarto P (de *place*) do *marketing mix* de ânimo leve (Gilbert, 2002 e Salvaneschi, 1996). Trabalhos como os de Pioch e Byrom (2004) e Jones *et al.* (2003) confirmam a necessidade de uma boa localização, em especial em serviços mais padronizados e com atendimento menos personalizado, como é o caso das cadeias de supermercados. Neste contexto, o desenvolvimento de modelos e técnicas de apoio à decisão baseados em modelos quantitativos de previsão de vendas em novas localizações assume uma relevância acrescida.

I.B. Motivação, Definição do Problema, Objectivos e Estrutura

A motivação deste trabalho surgiu da necessidade, sentida pelo grupo de distribuição, de revitalizar uma pequena cadeia de lojas de retalho alimentar que se posicionara no mercado essencalmente como Supermercados de Proximidade, orientados para as classes de rendimentos médias a altas, ainda que originariamente tivessem tido uma orientação mais próxima das lojas *Discount* (ver Figura 1).

Ainda que actualmente o posicionamento estratégico seja claro, tanto para a cadeia existente como para as lojas a abrir futuramente, na verdade, alguma incerteza

[4] Revista Poupança Quinze, nº 233 de 27/7/2004, Lisboa: Edideco, pág. 7.

quanto a esse posicionamento no passado conduziu à abertura de lojas com características diferenciadas como é o caso de algumas lojas próximas dos Supermercados Grandes. Desta forma, foi sendo criada uma cadeia de lojas com dimensões e localizações heterogéneas cujo ponto comum é o facto de se localizarem todas nas áreas metropolitanas de Lisboa e Porto e quase todas dentro de cidades suburbanas. Esta cadeia de supermercados tem geralmente áreas alimentares e não alimentares, sendo a não alimentar responsável por uma pequena fracção das vendas da loja (entre 10 a 20%).

O problema essencial posto pelos especialistas do grupo de distribuição era a **comparação de localizações potenciais**. Após testes com modelos que eram usados para lojas de maiores dimensões localizadas mais longe do centro das cidades, chegaram rapidamente à conclusão que eram inadequados para este tipo de lojas. Na verdade as lojas de menores dimensões estão muito mais dependentes das vizinhanças próximas e tendem a apresentar valores de vendas mais difícieis de explicar uma vez que exigem uma análise mais fina.

Assim, o problema consiste em desenvolver modelos capazes de comparar localizações de pontos de venda de retalho alimentar correspondentes a lojas de pequena a média dimensão e com uma orientação típica de Supermercados de Proximidade. A este problema genérico foi acrescentada a restrição de que as localizações potenciais teriam de ser comparadas em termos de **vendas previstas**. É, aliás, esta última restrição imposta que justifica o título desta dissertação.

Ficou igualmente claro desde o início que, dada a reduzida dimensão da cadeia, com muito poucas lojas abertas, a **colaboração dos especialistas** seria ainda mais relevante do que se as circunstâncias fossem diferentes. Na verdade, a falta de dados quantitativos para validar os modelos teria de ser superada pelos conhecimentos profundos das lojas e da cadeia detidos por estes especialistas em localização. Os especialistas são, neste caso, analistas de *marketing* com formação em ciências sociais e gestão, responsáveis por todas as decisões de localização da cadeia em consideração e conhecedores de cada uma das lojas individualmente.

Foi ainda decidido que não se pretendia apoiar decisões estratégicas como a selecção de regiões do pais em que estes Supermercados de Proximidade deveriam ser instalados. Dado que a orientação estratégica da cadeia já estava definida, revelou-se concensual que estas lojas se deveriam localizar em zonas de grande expansão demográfica, ou zonas onde os consumidores apresentassem elevados rendimentos. No caso do continente português tal só se verifica nas zonas metropolitanas de Lisboa e do

Porto. As restantes zonas são cobertas por grandes lojas fora das cidades ou pequenas lojas em regime de *franchising*.

Assim, podem-se enumerar os seguintes objectivos para o trabalho que foi proposto e que é apresentado nesta dissertação.

(i) sistematizar, comparar, classificar e avaliar os **modelos descritos na literatura** sobre avaliação de desempenho de lojas de retalho e comparação de localizações potenciais;

(ii) definir uma **classificação das variáveis** a considerar nos problemas de previsão de vendas em novas localizações e recolher dados provenientes de várias origens que permitam cobrir todas as classes de variáveis identificadas;

(iii) utilizar e comparar diferentes modelos de **delimitação de áreas de influência** que permitam integrar variáveis demográficas em estudos de localização por análise espacial;

(iv) definir uma **tipologia de lojas** que permita compreender melhor o comportamento das diferentes lojas existentes e que possa ser utilizada nos modelos subsequentes;

(v) desenvolver modelos para apoiar decisões de **comparação de localizações potenciais** de novas lojas alimentares de pequena a média dimensão baseadas em previsão de vendas;

(vi) **integrar o conhecimento** da área detido pelos especialistas, tanto no desenvolvimento dos modelos e das metodologias, como na validação dos mesmos;

(vii) demonstrar que os **modelos adoptados** e a metodologia desenvolvida são, não apenas válidos, como os mais adequados e os que conduzem às melhores previsões, dadas as alternativas disponíveis e as limitações impostas;

(viii) por fim, o objectivo fundamental de todo o trabalho é a **geração de conhecimento** sobre este problema complexo que possa ser utilizado em momentos de decisão futuros.

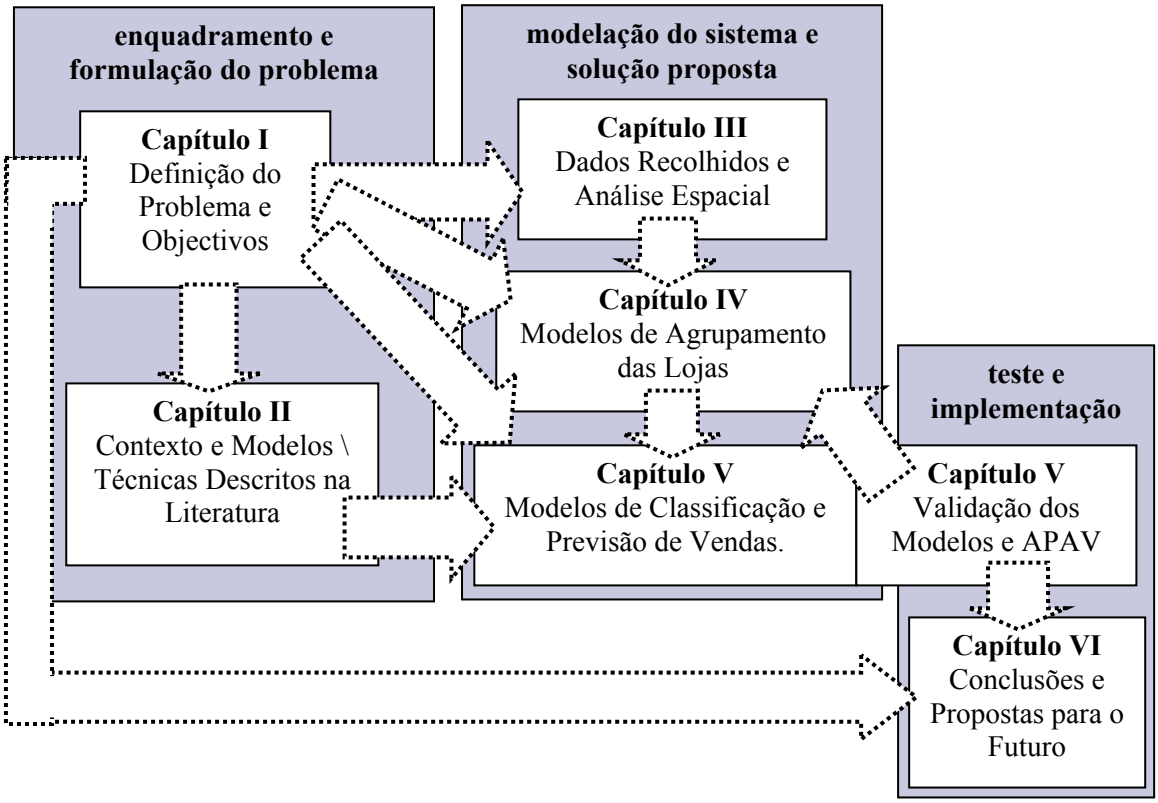
Assim, a **estrutura da dissertação** segue de perto a necessidade de preencher os objectivos identificados como se pode observar na Figura 3.

Neste capítulo apresentou-se uma descrição do problema que inclui já uma definição das fronteiras do sistema em estudo. No capítulo II faz-se uma abordagem mais completa ao contexto do problema e dos níveis de decisão envolvidos. Apresenta-se ainda uma sistematização dos modelos e técnicas da literatura, incluindo as metodologias baseadas em Sistemas de Informação Geográfica.

No capítulo III passa-se à fase de modelação do sistema e descrição da solução proposta. Neste capítulo, além de se sugerir uma classificação das variáveis usadas em problemas de avaliação de desempenho e localização de lojas de retalho, descrevem-se os dados recolhidos por diversos métodos, os testes de qualidade e consistência efectuados e os processos de integração utilizados. Descreve-se ainda o tratamento de

análise espacial efectuado com definição de áreas de influência por diferentes métodos que são comparados em termos de capacidade explicativa das variáveis obtidas.

FIGURA 3 ESTRUTURA DA DISSERTAÇÃO SEGUNDO TRÊS FASES.
(As setas referem-se às dependências mais relevantes entre os capítulos da dissertação)



No capítulo IV continua-se a modelação do problema, agora com definição de um modelo para compreender as diferenças entre grupos de lojas. Assim, define-se uma tipologia baseada na integração de conhecimento dos especialistas escolhida por comparação com outras tipologias desenvolvidas utilizando metodologias distintas. Este modelo de loja análoga é central neste trabalho e estruturante de todos os restantes modelos desenvolvidos.

O capítulo V tem uma dimensão superior aos restantes uma vez que inclui várias fases do processo. Assim, opta-se por modelos de análise de dados com fins descritivos e preditivos em detrimento de modelos mais normativos e desenvolvem-se modelos de classificação das lojas e de previsão baseados em regressão linear. Deste modo, completa-se a fase de modelação do sistema. No mesmo capítulo descreve-se um ambiente decisional baseado numa aplicação em folha de cálculo que permite obter previsões de vendas para localizações potenciais em comparação, e a que se chamou

APAV – Análise e Previsão por Analogia de Vendas. Assim, este capítulo inclui igualmente a fase de teste e validação dos modelos e da solução encontrada.

Por fim, apresentam-se as conclusões do trabalho efectuado e tenta-se provar que os objectivos propostos foram atingidos. Tenta-se igualmente, neste capítulo, destacar as contribuições que esta dissertação traz para o problema genérico de localização de lojas de retalho e apresentam-se vias a explorar no futuro. Esta última parte, centrada na preocupação com a constante melhoria das soluções encontradas, pode ser integrado no esquema da Figura 3 referente à implementação e actualização da solução proposta.

A dissertação termina com um conjunto de anexos onde se apresentam tabelas e gráficos que por serem demasiado extensos e por não serem essenciais para a compreensão do texto se remetem para esta área de consulta. Todos os capítulos incluem ainda um resumo inicial e secções introdutórias.

I.C. Algumas Considerações sobre a Nomenclatura

Nesta dissertação utilizam-se técnicas, métodos e algoritmos provenientes de dois domínios distintos mas que nos últimos anos têm verificado uma evolução convergente, nomeadamente a estatística multivariada e o reconhecimento de padrões (*pattern recognition*) ou aprendizagem automática (*machine learning*). Este facto complica a utilização de uma terminologia adequada, uma vez que cada um destes domínios usa as suas próprias designações. Mesmo quando os mesmos termos são utilizados por vezes têm significados distintos.

Na maioria dos textos de estatística multivariada o termo classificação engloba qualquer tipo de método usado para agrupar um conjunto de entidades em subgrupos. Assim, neste termo estariam englobados actividades complexas e multivariadas relacionadas com a análise de agrupamentos (*clusters*) ou actividades tão simples como agrupar segundo atributos conhecidos como o género ou classes de idades. Neste mesmo sentido, mais fundamentado no tempo, segue igualmente a definição apresentada na Diciopédia 2005 da Porto Editora[5]:

«acto, efeito ou processo de distribuir por classes»

ou a definição apresentada por Hartigan (1996) e atribuída a Webster:

[5] Diciopédia 2005 em DVD da Porto Editora, ISBN 972-0-65258-6.

> *«classification is (1) the act or process of classifying; (2) the systematic arrangement in groups or categories according to established criteria».*

Note-se, no entanto, que esta utilização da palavra classificação não é consensual entre todos os autores de estatística multivariada. Por exemplo, Everitt *et al.* (2001) utiliza a expressão "análise de *clusters*" como sinónimo de classificação efectuada por métodos numéricos, sugerindo mesmo que o primeiro pode ser mais abrangente do que o segundo (pág. 4):

> *«... nowadays cluster analysis is probability the preferred generic term for procedures which seek to uncover groups in data».*

Na mesma linha de pensamento surge a definição apresentada por Gordon (1999) onde classificação e análise de agrupamentos surgem mais uma vez como sinónimos, já que a palavra "*classification*" poderia sem perda de significado ser substituída pela expressão "*cluster analysis*":

> *«The subject of 'classification' is concerned with the investigation of sets of 'objects' in order to establish if they can validly be summarized in terms of a small number of classes of similar objects».*

Perante estas indefinições, o termo classificação tem-se tornado ao longo do tempo mal definido e confuso na literatura de estatística multivariada.

Também na literatura de análise de *marketing* o termo segmentação é utilizado como sinónimo de classificação ou mesmo, de forma ainda mais lata, incluindo neste conceito qualquer técnica que permita dividir entidades em grupos. Por exemplo, Wedel e Kamakura (2000) incluem nesta denominação técnicas como tabelas de contingência, tabelas cruzadas, regressão, análise discriminante, árvores de classificação ou modelos de mistura.

Pelo contrário, na bibliografia de reconhecimento de padrões o termo "classificação" é utilizado de forma muito mais restrita. Neste domínio do conhecimento, classificar corresponde a prever o valor de uma variável dependente ou *target*. Tal é fácil de entender, já que, se a variável for nominal, prever o valor para uma nova entidade corresponde a colocar um rótulo nessa entidade, e logo classifica-la no grupo de todas a que detêm esse rótulo. Nas palavras de Breiman *et al.* (1984) pág. 6:

> *«… the basic propose of a classification study can be either to produce an accurate classifier or to uncover the predictive structure of the problem».*

Este conceito vem na sequência de outros dois conceitos: aprendizagem supervisionada (*supervised learning*) ou não supervisionada (*non supervised learning*). Nesta terminologia, a calibração de um modelo de previsão ou a estimação de um

classificador é designado por treino ou aprendizagem (Marques, 1999). Assim, na aprendizagem supervisionada utiliza-se uma variável dependente com informação sobre as classes a que pertencem cada uma das entidades da amostra de treino. Neste conceito incluem-se técnicas da estatística multivariada como a regressão, análise discriminante e a regressão logística e técnicas novas da área de reconhecimento de padrões como as árvores de classificação e de regressão e as redes neuronais supervisionadas. Assim, o conceito de aprendizagem supervisionada conduz ao conceito de modelos de agrupamento baseados em relações de dependência, introduzido por Cardoso (2000), ou às técnicas preditivas de Wedel e Kamakura (2000).

Pelo contrário, na aprendizagem não supervisionada a divisão em classes baseia-se na procura de padrões ou de uma estrutura nos dados considerando em pé de igualdade todas as variáveis. Assim, enquadram-se neste conceito as técnicas de análise de *clusters*, os modelos de mistura e de segmentos latentes sem relações de dependência e as redes neuronais não supervisionadas. Cardoso (2000), denomina os modelos resultantes como modelos de agrupamento baseados em relações de interdependência e Wedel e Kamakura (2000) chama-lhes técnicas descritivas.

Tendo em conta que a definição apresentada na bibliografia de reconhecimento de padrões é mais precisa e clara, nesta dissertação adopta-se o termo “classificação” de forma restrita para técnicas como as árvores de classificação que utilizam aprendizagem supervisionada para prever um atributo nominal e construir modelos discriminantes lógicos (ver Figura 4).

No caso de se pretender prever uma variável contínua, utiliza-se a expressão “**árvores de regressão**” adoptada de Breiman *et al.* (1984). A expressão “análise de agrupamentos” é, assim, considerado independente de classificação. Aliás vários autores, na área da engenharia de sistemas e nomeadamente nos sistemas de apoio à decisão, utilizam nomenclaturas idênticas (ver por exemplo Sauter, 1997 e Turban *et al.,* 2005). Na Figura 4 utiliza-se ainda o termo segmentação no sentido lato descrito atrás.