

# As Normas de Informação Geográfica e o Raciocínio Espacial Qualitativo na Inferência de Informação Geográfica Qualitativa

Maribel Santos                      Luís Amaral  
Departamento de Sistemas de Informação  
Universidade do Minho  
Campus de Azurém  
4810 Guimarães  
Portugal  
Telf.: 253 510 259  
Fax: 253 510 250  
{maribel, amaral}@dsi.uminho.pt

**Resumo.** Este trabalho aborda a problemática da inferência automática de informação geográfica através do raciocínio espacial qualitativo. A implementação de um sistema de inferências é conseguida recorrendo a uma aplicação de *Data Mining*, o *Clementine*, na qual algoritmos inteligentes assimilam o conhecimento geográfico necessário à realização de tais inferências. O objectivo final é a criação de um ambiente de trabalho no qual seja possível integrar bases de dados geográficas e bases de dados não geográficas para a realização de *Data Mining* Espacial.

## 1 Introdução

As bases de dados geográficas existentes em inúmeros sistemas de informação geográfica armazenam grandes quantidades de informação espacial. Pesquisas nestes sistemas envolvem a derivação de relacionamentos espaciais que não estão explicitamente armazenados nas bases de dados.

O raciocínio espacial qualitativo tem sido proposto como um mecanismo complementar para a derivação automática de relacionamentos espaciais que não se encontram explícitos nas bases de dados geográficas [Abdelmoty e El-Geresy 1995]. O raciocínio qualitativo é baseado na manipulação de identificadores qualitativos como *Norte*, *Sul*, ... em detrimento de informação quantitativa como coordenadas ou distâncias entre pontos.

Se é uma realidade que a utilização de tais identificadores facilita o raciocínio, a existência de repositórios de informação geográfica onde tais indicadores surgem de uma forma normalizada, permite o desenvolvimento de mecanismos automáticos para a inferência de informação geográfica desconhecida ao utilizador.

Este trabalho aborda a problemática da **inferência automática de informação geográfica através do raciocínio espacial qualitativo**. A implementação de um sistema de inferências é conseguida recorrendo a uma aplicação de *Data Mining*, o *Clementine*, na qual algoritmos inteligentes assimilam o conhecimento geográfico necessário à realização de tais inferências. O objectivo final é a criação de um ambiente de trabalho no qual seja possível integrar bases de dados geográficas e bases de dados não geográficas para a realização de *Data Mining* Espacial.

O sistema utiliza bases de dados geográficas implementadas segundo os princípios estabelecidos pelo Comité Europeu de Normalização (CEN) nas normas **CEN TC 287 para Informação Geográfica**<sup>1</sup>, as quais desempenham um papel fundamental na definição e estruturação da informação utilizada neste estudo. A utilização de bases de dados geográficas cuja estrutura siga as recomendações do CEN permite-nos construir um sistema genérico, uma vez que o conjunto mínimo de informação que as mesmas devem disponibilizar é conhecido à partida.

Este artigo é organizado da seguinte forma. A secção 2 apresenta resumidamente as normas CEN TC 287 para informação geográfica. Na secção 3 são apresentados os princípios nos quais se rege o raciocínio espacial qualitativo, nomeadamente no que diz respeito aos relacionamentos do tipo direcção e distância. A secção 4 caracteriza a base de dados geográfica utilizada no processo de inferência descrito na secção 5. Na secção 6 são apresentadas algumas conclusões e propostas de trabalho futuro.

## 2 As normas CEN TC 287 para Informação Geográfica

As normas CEN TC 287 têm como principal objectivo permitir que a informação geográfica possa ser acedida por diferentes utilizadores, sistemas, aplicações e principalmente, de diferentes localizações. Requerem a implementação de oito esquemas, que são integrados posteriormente, dando origem ao esquema de aplicação [CEN/TC-287 1998a], ou seja, a especificação da informação para um domínio de aplicação. Os oito esquemas são: semântico, espacial, qualidade, metadados, posicionamento directo, identificadores geográficos, transferência e pesquisa e alteração. Mais detalhes sobre o objectivo de cada um destes esquemas podem ser encontrados em [Santos, et al. 1999].

Dos oito esquemas referidos anteriormente, destaca-se a importância do **esquema de identificadores geográficos** e do **esquema espacial** na inferência de relações espaciais existentes entre as entidades que constituem o espaço geográfico analisado. O esquema espacial [CEN/TC-287 1996] permite definir as primitivas geométricas e topológicas das entidades geográficas analisadas. O esquema de identificadores geográficos [CEN/TC-287 1998b] permite explicitar a hierarquia e relacionamentos existentes entre os identificadores utilizados. Tal é conseguido através da implementação de um sistema de identificadores geográficos.

### 2.1 O Esquema de Identificadores Geográficos

Num sistema de referência espacial utilizando identificadores geográficos, uma posição é indexada a uma localização recorrendo a um objecto real, sendo o seu identificador denominado de identificador geográfico. Neste contexto inserem-se nomes de ruas, de cidades, de monumentos, etc. Este sistema de referência é normalmente apelidado de *indirecto*, uma vez que não recorre à utilização de coordenadas. O documento prENV 12661 [CEN/TC-287 1998b] define como devem ser descritas referências espaciais que recorrem a utilização de identificadores geográficos e especifica, ainda, como deve ser construído um catálogo geográfico.

---

<sup>1</sup> Ressalva-se que neste momento a norma se encontra em fase de votação final pelo CEN, pelo que na realidade é apelidada de pré-norma apesar de ao longo do artigo ser sempre referida como norma.

O esquema de identificadores geográficos permite explicitar a hierarquia e os relacionamentos existentes entre os identificadores utilizados, através da implementação de um sistema de identificadores geográficos (Tabela 1).

O sistema de identificadores geográficos deve conter uma colecção estruturada de classes de localizações e os seus respectivos identificadores geográficos. As localizações deverão ser entidades identificáveis a partir do mundo real, podendo ser consideradas entidades com 0 ou mais dimensões. Deverão ainda ser independentes (existirem por si só) e contíguas (fornecendo uma subdivisão completa do domínio geográfico em causa). Os identificadores geográficos deverão identificar univocamente uma localização. O conjunto de instâncias de uma localização deve ser compilado num catálogo geográfico [CEN/TC-287 1998b].

Sistema de Identificadores Geográficos		
Classes de Localização	Distritos	Concelhos
Instâncias de Localização	Aveiro Braga Porto ...	Ílhavo Amares Guimarães ...
Identificadores Geográficos	AVR BRG PRT ...	110 301 308 ...

Tabela 1 – Sistema de Identificadores Geográficos

A descrição do sistema de identificadores inclui a descrição do sistema propriamente dito e a descrição de cada classe de localização. A implementação deste esquema permite a qualquer utilizador conhecer todos os detalhes associados ao sistema em causa. A interpretação dos diagramas EXPRESS [Schenck e Wilson 1994] especificados no documento prENV 12661 conduziu à elaboração do diagrama entidades-relacionamentos apresentado na Figura 1.

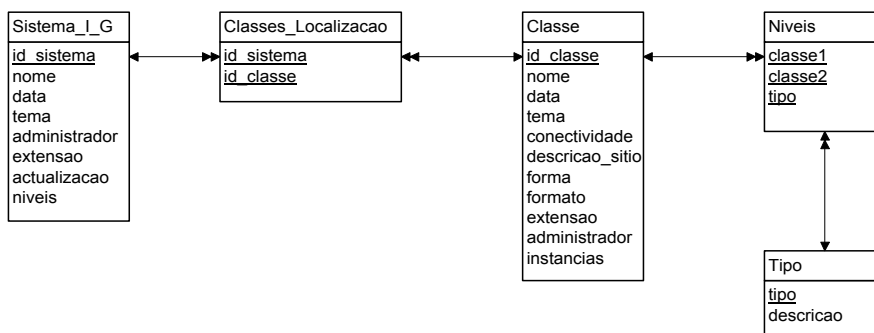


Figura 1 – Esquema de Identificadores Geográficos

Salienta-se que a referenciação geográfica da informação utilizada neste estudo é conseguida recorrendo a um **sistema de identificadores geográficos** que considera divisões administrativas de Portugal Continental, nomeadamente ao nível dos Concelhos e Distritos (ver secção 4). Sendo mandatária a construção do catálogo geográfico para todas as instâncias de localização utilizadas, os relacionamentos geográficos existentes entre as mesmas, ficam assim devidamente especificados. A interpretação do esquema proposto para o catálogo de localizações ([CEN/TC-287 1998b] Anexo C), conduziu à elaboração do diagrama entidades-relacionamentos apresentado na Figura 2.

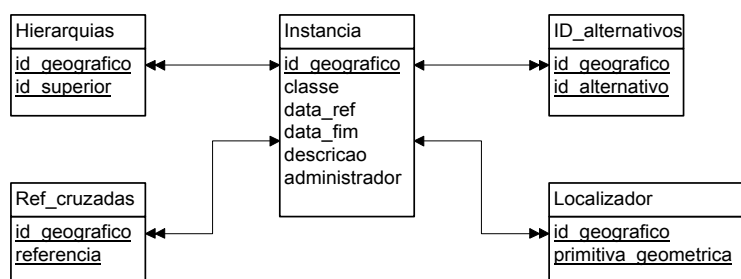


Figura 2 – Especificação do Catálogo Geográfico

## 2.2 O Esquema Espacial

Os aspectos espaciais da informação geográfica permitem que dados geográficos de diferentes origens possam ser integrados através da georreferenciação, ou ainda que os mesmos possam ser analisados utilizando operadores e funções espaciais. Estes aspectos requerem a utilização de estruturas especializadas para a geometria e a topologia.

No documento prENV 12160 [CEN/TC-287 1996] encontram-se definidos esquemas espaciais básicos (G1 a G8) derivados de um esquema espacial genérico denominado de G0. Sempre que nenhum dos esquemas G1 a G8 caracterizar correctamente a área geográfica em análise, o esquema G0 deverá ser utilizado como um *template* que dita as regras de construção do esquema espacial (é a partir deste esquema genérico que derivam os restantes esquemas G1 a G8).

Estes esquemas incluem aspectos geométricos e topológicos. A geometria permite descrever quantitativamente, através de coordenadas e funções matemáticas, os aspectos espaciais da informação geográfica. A topologia cobre a descrição qualitativa dos aspectos espaciais, permanecendo estes invariantes a qualquer deformação (por ex. aumento de escala) do espaço em análise. Estas últimas permitem a descrição dos aspectos topológicos dos objectos através da definição dos relacionamentos existentes entre os mesmos. A posição espacial de cada primitiva topológica pode ser descrita por uma ou mais primitivas geométricas.

A análise dos vários esquemas G1 a G8 permitiu constatar que o Mapa de topologia planar G1 caracteriza o espaço geográfico em análise neste trabalho. Este esquema deve ser utilizado sempre que se utilizem mapas planos com faces que

cubram completamente a superfície em estudo e em relação as quais não exista qualquer sobreposição. Neste esquema são permitidos nodos isolados ou terminais, mas não permitem a existência de nodos intermédios (as arestas que definem os limites administrativos não se podem sobrepor). As faces têm de ser definidas sem qualquer orifício.

O esquema espacial G1 possui duas ramificações: uma geométrica e uma topológica. Uma vez que neste caso de estudo a referência é obtida através da utilização de identificadores geográficos, só se utilizará a ramificação respeitante à topologia, apesar da Figura 3 apresentar as duas componentes. Pela análise da figura verifica-se que a vizinhança à esquerda e à direita das faces é conhecida através das arestas que as delimitam. Pretendendo-se neste estudo fazer uma análise mais detalhada dos relacionamentos espaciais existentes entre as entidades (nomeadamente entre as faces que cobrem a superfície em análise), o diagrama apresentando na Figura 3 foi alterado (na parte respeitante à primitivas topológicas) por forma a suportar mais duas entidades, **face** e **direcção**, as quais permitem explicitar os relacionamentos espaciais do tipo direcção e distância existentes entre as regiões consideradas (Figura 4).

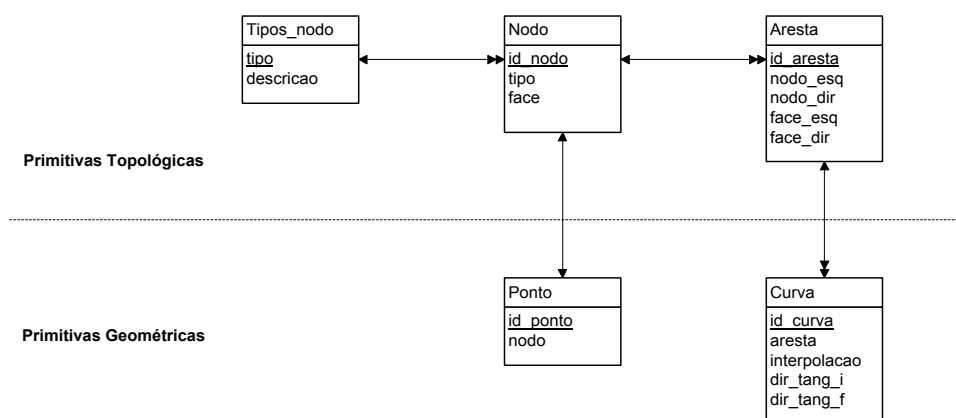


Figura 3 – Esquema espacial: primitivas geométricas e topológicas

Destaca-se que o novo diagrama apresentado na Figura 4 (evidenciando apenas a componente topológica) complementa as especificações proferidas pelas normas, acrescentando informação que é extremamente importante na percepção dos relacionamentos existentes entre regiões, neste caso divisões administrativas. Esta alteração tem como objectivo tornar o raciocínio cada vez mais exacto, uma vez que raciocinar qualitativamente em termos de distâncias requer a análise da direcção existente entre as regiões analisadas. Apesar disto, e sempre que tal se revelasse suficiente, a tabela *Aresta* permite inferir a localização existente entre as faces, verificando o posicionamento das mesmas em relação às arestas.

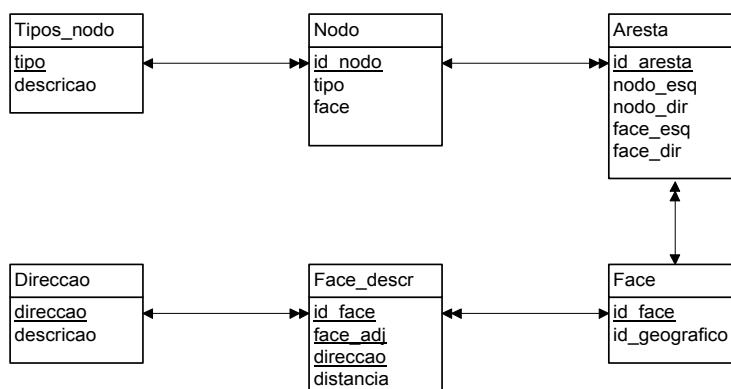


Figura 4 – Primitivas topológicas que integram o esquema espacial

### 3 O Raciocínio Espacial Qualitativo

No que diz respeito ao raciocínio espacial, as relações espaciais têm sido classificadas em vários tipos [Frank 1996; Papadias e Sellis 1994], incluindo **relações topológicas**, **relações de direcção** e **relações de distância**. Para cada uma destas relações existem tabelas de composição [Sharma 1996] que facilitam o raciocínio espacial permitindo a inferência de informação desconhecida. Sempre que as tabelas consideram um único tipo de relação representam **raciocínio espacial homogéneo**, já que apenas permitem a inferência de informação espacial do mesmo tipo (Ex: A Norte B, B Este C → A Nordeste C). A abordagem **heterogénea** permite partir de dois tipos de relações distintas e inferir o relacionamento desejado pelo utilizador (Ex: A contido em B, B Noroeste C → A Noroeste C).

Apesar de ambas abordagens serem úteis, apresentam limitações. Raciocinar qualitativamente em termos de distâncias requer um **raciocínio integrado** que conjugue a direcção e a distância entre as entidades analisadas. Por exemplo, os factos A muito distante B, B muito distante C não permite inferir o relacionamento que existe entre A e C. A pode estar próximo ou muito próximo de C, ou A pode estar distante ou muito distante de C, dependendo da orientação que existe entre A e B, e B e C.

A aproximação integrada requer a adopção de um conjunto de indicadores qualitativos que representem as direcções e distâncias em análise, assim como os respectivos intervalos de validade (para o caso das distâncias). Hong [Hong 1994] analisou os padrões que caracterizam a escolha do número de indicadores e a dimensão dos respectivos intervalos. Entre as várias combinações possíveis, destaca-se a definição de um **sistema de localização** baseado em oito símbolos para a direcção (Norte, Nordeste, Este, Sudeste, Sul, Sudoeste, Oeste, Noroeste) e quatro símbolos para a distância (muito próximo, próximo, distante, muito distante). Estes últimos são caracterizados por representar intervalos de dimensão regular nos quais existe um *ratio* ( $ratio = \text{comp}(dist_i) / \text{comp}(dist_{i-1})$ ) constante entre o comprimento de dois intervalos vizinhos (Tabela 2).

Os intervalos definidos permitem a definição de novos intervalos através da ampliação do intervalo original. Por exemplo, o conjunto de valores definido para o

*ratio* 4 pode ser multiplicado pelo factor 10, originando um novo conjunto de intervalos:  $dist_0$  (0, 10],  $dist_1$  (10, 50],  $dist_2$  (50, 210] e  $dist_3$  (210, 850]. Uma vez que o mesmo factor amplia todos os intervalos (relações quantitativas), as inferências qualitativas permanecem inalteradas seja qual for o factor utilizado.

<i>Ratio</i>	$dist_0$	$dist_1$	$dist_2$	$dist_3$
1	(0, 1]	(1, 2]	(2, 3]	(3, 4]
2	(0, 1]	(1, 3]	(3, 7]	(7, 15]
3	(0, 1]	(1, 4]	(4, 13]	(13, 40]
4	(0, 1]	(1, 5]	(5, 21]	(21, 85]
5	(0, 1]	(1, 6]	(6, 31]	(31, 156]
6	(0, 1]	(1, 7]	(7, 43]	(43, 259]
7	(0, 1]	(1, 8]	(8, 57]	(57, 400]
8	(0, 1]	(1, 9]	(9, 73]	(73, 585]
9	(0, 1]	(1, 10]	(10, 91]	(91, 820]
10	(0, 1]	(1, 11]	(11, 111]	(111, 1111]
20	(0, 1]	(1, 21]	(21, 421]	(421, 8421]
50	(0, 1]	(1, 51]	(51, 2551]	(2551, 127551]
100	(0, 1]	(1, 101]	(101, 10101]	(10101, 1010101]

**Tabela 2 – Limites para os intervalos de referência qualitativa** (Adaptado de Hong, J.-H., M. J. Egenhofer, e A. U. Frank, *On the Robustness of Qualitative and Direction Reasoning*, Proceedings of Auto-Carto 12, Charlotte, North California, 1995).

Após a caracterização dos intervalos, Hong [Hong 1994] definiu as tabelas de composição para um sistema de localização com estas características, verificando ainda a robustez das inferências obtidas. Tais tabelas permitem a sua utilização na inferência de novas localizações desde que as propriedades dos intervalos sejam respeitadas.

A notação simbólica utilizada por Hong é apresentada na Figura 5. Existindo várias tabelas de composição (que dependem do *ratio* entre intervalos utilizado), a Tabela 3 (onde **O** denota identidade – mesma posição) apresenta o conjunto de inferências possíveis para um sistema de localização, com *ratio* 4, baseado em quatro símbolos para a distância e oito para a direcção. Pela análise da tabela verifica-se que os casos apresentados  $\Delta_{dir0}$ ,  $\Delta_{dir1}$ ,  $\Delta_{dir2}$ ,  $\Delta_{dir3}$  e  $\Delta_{dir4}$  representam localizações com a mesma direcção até localizações com direcções opostas respectivamente. Os restantes casos devem ser gerados partindo do pressuposto que  $\Delta_{dir1} = \Delta_{dir7}$ ,  $\Delta_{dir2} = \Delta_{dir6}$  e  $\Delta_{dir3} = \Delta_{dir5}$ .

Analisando as composições especificadas na Tabela 3, e seguindo os critérios definidos por Hong, o conjunto de todas as inferências possíveis foi gerado. Como resultado obteve-se uma matriz com 32x32 células que será utilizada em conjunto com os relacionamentos geográficos explícitos existentes entre regiões na inferência de novos relacionamentos espaciais. A próxima secção descreve a base de dados geográfica utiliza assim como os relacionamentos que se encontram explícitos na mesma.

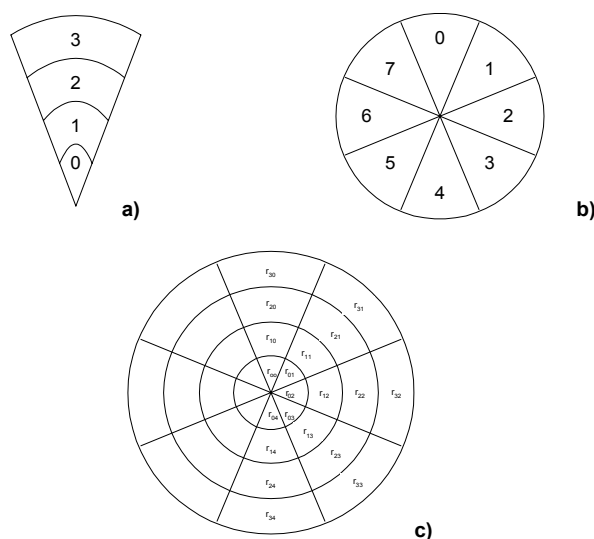


Figura 5 – a) Representação qualitativa das distâncias; b) Representação qualitativa das direcções; c) Representação simbólica de um sistema de localização baseado em 4x8 símbolos (Adaptado de Hong, J.-H., *Qualitative Distance and Direction Reasoning in Geographic Space*, PhD thesis, University of Maine, 1994, p.76).

$\Delta dir0$		$\Delta dir1$		$\Delta dir2$		$\Delta dir3$		$\Delta dir4$						
$r_{00};r_{00}$	$\rightarrow$	$r_{00}$	$r_{00};r_{01}$	$\rightarrow$	$r_{00}$	$r_{00};r_{02}$	$\rightarrow$	$r_{01}$	$r_{00};r_{03}$	$\rightarrow$	$r_{02}$	$r_{00};r_{04}$	$\rightarrow$	$r_{03}$
$r_{10};r_{00}$	$\rightarrow$	$r_{10}$	$r_{10};r_{01}$	$\rightarrow$	$r_{10}$	$r_{10};r_{02}$	$\rightarrow$	$r_{10}$	$r_{10};r_{03}$	$\rightarrow$	$r_{10}$	$r_{10};r_{04}$	$\rightarrow$	$r_{10}$
$r_{20};r_{00}$	$\rightarrow$	$r_{20}$	$r_{20};r_{01}$	$\rightarrow$	$r_{20}$	$r_{20};r_{02}$	$\rightarrow$	$r_{20}$	$r_{20};r_{03}$	$\rightarrow$	$r_{20}$	$r_{20};r_{04}$	$\rightarrow$	$r_{20}$
$r_{30};r_{00}$	$\rightarrow$	$r_{30}$	$r_{30};r_{01}$	$\rightarrow$	$r_{30}$	$r_{30};r_{02}$	$\rightarrow$	$r_{30}$	$r_{30};r_{03}$	$\rightarrow$	$r_{30}$	$r_{30};r_{04}$	$\rightarrow$	$r_{30}$
$r_{10};r_{10}$	$\rightarrow$	$r_{20}$	$r_{00};r_{11}$	$\rightarrow$	$r_{11}$	$r_{00};r_{12}$	$\rightarrow$	$r_{12}$	$r_{00};r_{13}$	$\rightarrow$	$r_{13}$	$r_{00};r_{14}$	$\rightarrow$	$r_{14}$
$r_{20};r_{10}$	$\rightarrow$	$r_{20}$	$r_{10};r_{11}$	$\rightarrow$	$r_{21}$	$r_{10};r_{12}$	$\rightarrow$	$r_{11}$	$r_{10};r_{13}$	$\rightarrow$	$r_{12}$	$r_{10};r_{14}$	$\rightarrow$	$r_{14}$
$r_{30};r_{10}$	$\rightarrow$	$r_{30}$	$r_{20};r_{11}$	$\rightarrow$	$r_{20}$	$r_{20};r_{12}$	$\rightarrow$	$r_{20}$	$r_{20};r_{13}$	$\rightarrow$	$r_{20}$	$r_{20};r_{14}$	$\rightarrow$	$r_{20}$
$r_{20};r_{20}$	$\rightarrow$	$r_{30}$	$r_{30};r_{11}$	$\rightarrow$	$r_{30}$	$r_{30};r_{12}$	$\rightarrow$	$r_{30}$	$r_{30};r_{13}$	$\rightarrow$	$r_{30}$	$r_{30};r_{14}$	$\rightarrow$	$r_{30}$
$r_{30};r_{20}$	$\rightarrow$	$r_{30}$	$r_{00};r_{21}$	$\rightarrow$	$r_{21}$	$r_{00};r_{22}$	$\rightarrow$	$r_{22}$	$r_{00};r_{23}$	$\rightarrow$	$r_{23}$	$r_{00};r_{24}$	$\rightarrow$	$r_{24}$
$r_{30};r_{30}$	$\rightarrow$	$r_{30}$	$r_{10};r_{21}$	$\rightarrow$	$r_{21}$	$r_{10};r_{22}$	$\rightarrow$	$r_{22}$	$r_{10};r_{23}$	$\rightarrow$	$r_{23}$	$r_{10};r_{24}$	$\rightarrow$	$r_{24}$
			$r_{20};r_{21}$	$\rightarrow$	$r_{31}$	$r_{20};r_{22}$	$\rightarrow$	$r_{21}$	$r_{20};r_{23}$	$\rightarrow$	$r_{22}$	$r_{20};r_{24}$	$\rightarrow$	$r_{24}$
			$r_{30};r_{21}$	$\rightarrow$	$r_{30}$	$r_{30};r_{22}$	$\rightarrow$	$r_{30}$	$r_{30};r_{23}$	$\rightarrow$	$r_{30}$	$r_{30};r_{24}$	$\rightarrow$	$r_{30}$
			$r_{00};r_{31}$	$\rightarrow$	$r_{31}$	$r_{00};r_{32}$	$\rightarrow$	$r_{32}$	$r_{00};r_{33}$	$\rightarrow$	$r_{33}$	$r_{00};r_{34}$	$\rightarrow$	$r_{34}$
			$r_{10};r_{31}$	$\rightarrow$	$r_{31}$	$r_{10};r_{32}$	$\rightarrow$	$r_{32}$	$r_{10};r_{33}$	$\rightarrow$	$r_{33}$	$r_{10};r_{34}$	$\rightarrow$	$r_{34}$
			$r_{20};r_{31}$	$\rightarrow$	$r_{31}$	$r_{20};r_{32}$	$\rightarrow$	$r_{32}$	$r_{20};r_{33}$	$\rightarrow$	$r_{33}$	$r_{20};r_{34}$	$\rightarrow$	$r_{34}$
			$r_{30};r_{31}$	$\rightarrow$	$r_{31}$	$r_{30};r_{32}$	$\rightarrow$	$r_{31}$	$r_{30};r_{33}$	$\rightarrow$	$r_{32}$	$r_{30};r_{34}$	$\rightarrow$	$r_{34}$

Tabela 3 – Tabela de Composição para o *ratio* 4 (Adaptado de Hong, J.-H., *Qualitative Distance and Direction Reasoning in Geographic Space*, PhD thesis, University of Maine, 1994, pp.139-144).

## 4 A Base de Dados Geográfica

O caso de estudo analisado é constituído por uma base de dados geográfica com os limites administrativos de Portugal Continental ao nível dos Concelhos e Distritos. Esta base de dados armazena as entidades e atributos provenientes da integração do esquema espacial com o esquema de identificadores geográficos. Entre



os atributos constantes destaca-se a direcção e distância (ou seja, a localização) existente entre Concelhos (Figura 6).

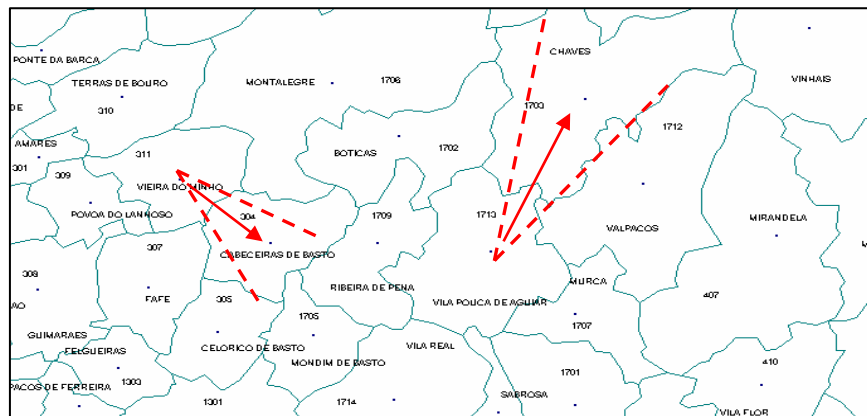


Figura 6 – Direcção e distância entre Concelhos (Mapa: “Atlas do Ambiente – DGA”)

Analisando a distância máxima que pode existir entre dois Concelhos, verificamos que a utilização do *ratio* 4 e a sua subsequente ampliação por um factor de 15, permite abranger todo o espaço analisado (Figura 7). Assim, os identificadores qualitativos utilizados e os respectivos intervalos de validade são: mp (0, 15], p (15, 75], d (75, 315], md (315, 1275].

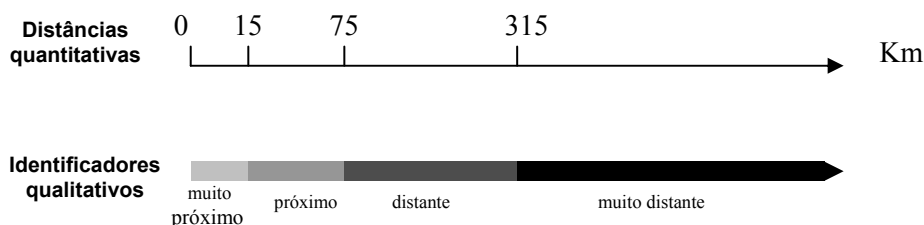


Figura 7 – Correspondência entre valores quantitativos e qualitativos

Os trabalhos desenvolvidos até ao momento permitiram construir uma base de dados geográfica, especificada segundo as recomendações do CEN, que armazena, entre outros, os **limites administrativos ao nível dos Concelhos e Distritos** (caracterizados recorrendo a um sistema de identificadores geográficos) e a **localização** (direcção e distância) existente entre alguns dos Concelhos (informação proveniente do esquema espacial, que apenas permite definir a localização entre Concelhos adjacentes). Na secção seguinte descreve-se o processo automático de inferência, implementado recorrendo a uma aplicação de *Data Mining*.

## 5 A Inferência de Informação Geográfica Qualitativa

O *Clementine* é uma aplicação de *Data Mining* baseada em programação visual, que inclui técnicas de aprendizagem automática como indução de regras ou redes neuronais [ISL 1998a; ISL 1998b]. O processo de descoberta de conhecimento é definido no *Clementine* através da construção de uma *stream*, na qual cada operação sobre os dados é representada por um nodo.

Uma vez que o objectivo deste trabalho é a incorporação do raciocínio espacial qualitativo no processo de descoberta de conhecimento, nomeadamente na derivação de relações espaciais que não estão explicitamente armazenadas na base de dados, duas etapas foram já executadas. Em primeiro lugar, a tabela de composição para a inferência de novas localizações (matriz 32x32) foi utilizada como fonte de informação para um nodo C5.0<sup>2</sup>. Através da *stream* construída (à esquerda na Figura 8), o processo de inferência foi assimilado e gerado um conjunto de regras de decisão (à direita na Figura 8) que retratam o conhecimento explícito na referida matriz. Estas regras podem agora ser utilizadas automaticamente na inferência de novos relacionamentos espaciais entre regiões.

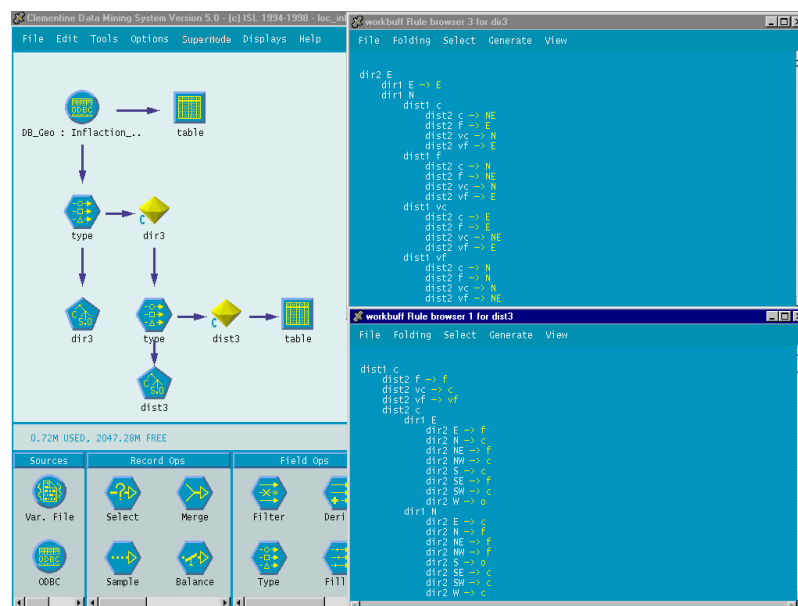


Figura 8 – Árvore de decisão gerada para o sistema de localizações

Posteriormente, e aqui podem existir duas abordagens, é possível inferir todos os relacionamentos espaciais implícitos existentes entre Concelhos ou, seleccionar o conjunto de Concelhos relevantes para a análise e só para estes gerar os relacionamentos desconhecidos.

Para o segundo caso referido, uma amostra de relacionamentos espaciais explícitos foi seleccionada (pertencentes aos Concelhos que integram o Distrito de

<sup>2</sup> Este nodo implementa o algoritmo C5.0 que permite a construção de árvores de decisão que permitem prever o valor de um dado atributo (*output*) baseado em diversos atributos de entrada (*input*).

Braga). Destaca-se que para 45 relacionamentos explícitos foram gerados 113 novos relacionamentos, através de um processo que ciclicamente (Figura 9) infere novas localizações entre Concelhos. A Figura 10 evidencia um extracto da tabela de resultados gerada no *Clementine* (com um total de 158 relacionamentos) e ainda um mapa da região que permite a verificação dos resultados.

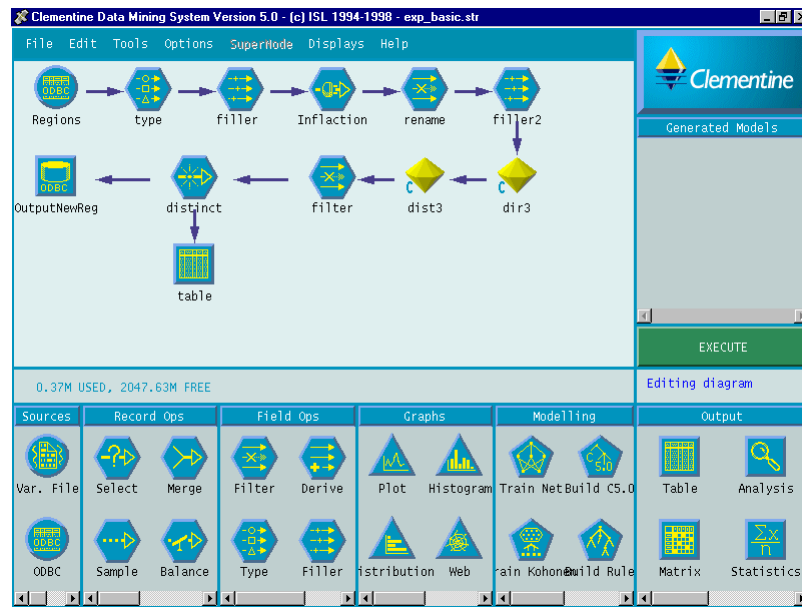


Figura 9 – Stream que ciclicamente infere novas relações

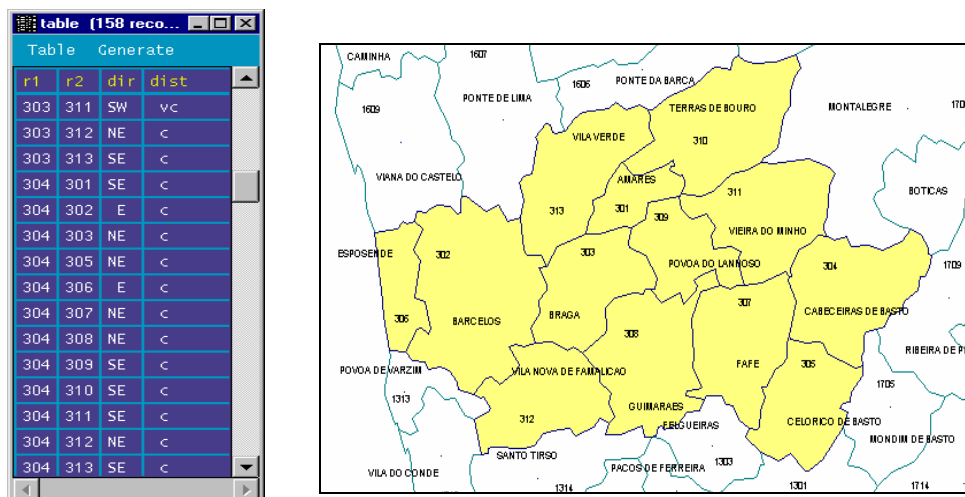


Figura 10 – Inferência de novas localizações entre Concelhos (Mapa: “Atlas do Ambiente DGA”)

O sistema permite ainda inferir a localização existente entre Distritos partindo do nível hierárquico inferior, o dos Concelhos. Salienta-se que o sistema

parte do conjunto de relacionamentos explícitos que existem entre Concelhos, executa ascensão hierárquica (através do conhecimento existente na hierarquia de conceitos) e aplica novamente as técnicas de indução de regras para descobrir a direcção e distância existente entre os Distritos analisados (aqui sem recorrer a qualquer conhecimento geográfico. O algoritmo apenas considera, na tomada de decisão, os casos conhecidos para o nível dos Concelhos).

Apesar do sistema partir de um conjunto restrito de localizações explícitas, a ascensão hierárquica e a análise dos dados resultantes pelo algoritmo C5.0 permite induzir correctamente a localização existente entre Distritos [Santos, et al. 1999]. A partir desta informação, e seguindo o procedimento descrito anteriormente para o caso dos Concelhos, é possível aplicar os mecanismos de inferência qualitativos já assimilados (árvore de decisão apresentada na Figura 8), no sentido de conhecer os restantes relacionamentos existentes entre Distritos.

Este processo de inferência espacial  $\rightarrow$  ascensão hierárquica  $\rightarrow$  inferência espacial pode ser aplicado repetidamente para todos os níveis especificados no sistema de identificadores geográficos. A grande vantagem é apenas ser necessário especificar alguns dos relacionamentos espaciais existentes entre as entidades pertencentes a classe que se encontra no nível hierárquico inferior.

## 6 Conclusões

Neste artigo abordou-se a problemática da inferência de informação geográfica qualitativa através da utilização de um sistema de localização baseado em 8 símbolos para a direcção e 4 símbolos para a distância. A base de dados utilizada foi construída seguindo os princípios definidos na norma CEN TC 287 para Informação Geográfica, nomeadamente através da implementação do esquema espacial e do esquema de identificadores geográficos.

A base de dados geográfica obtida foi explorada no *Clementine*, o que permitiu assimilar, através de algoritmos de indução de regras, o conjunto de princípios nos quais se baseia o raciocínio espacial qualitativo. Estas regras foram posteriormente utilizadas na inferência automática de relações espaciais do tipo direcção e distância entre entidades (através de um processo que ciclicamente infere novas localizações a partir do conhecimento disponível).

Em termos de trabalho futuro destaca-se que o sistema evoluirá por forma a integrar mais um tipo de relação espacial (a topologia) no raciocínio. Está equacionada ainda a integração da base de dados geográfica daqui resultante com uma base de dados demográfica. Pretende-se a aplicação dos algoritmos de *Data Mining* disponíveis no *Clementine* às duas bases de dados, com o objectivo de descobrir os padrões que caracterizam os dados demográficos analisados, assim como identificar os relacionamentos espaciais implícitos existentes entre os dados geográficos e os dados não geográficos analisados (*Data Mining* Espacial).

## 7 Referências

Abdelmoty, A. I., e B. A. El-Geresy, *A General Method for Spatial Reasoning in Spatial Databases*, Proceedings of the fourth International Conference on Information and Knowledge Management, Baltimore, ACM, 1995.

- CEN/TC-287, *Geographic Information: Data Description, Spatial Schema*, Comité Europeu de Normalização, prENV 12160, 1996.
- CEN/TC-287, *Geographic Information: Data Description, Rules for application schemas*, Comité Europeu de Normalização, WI 006, 1998a.
- CEN/TC-287, *Geographic Information: Referencing, Geographic Identifiers*, Comité Europeu de Normalização, prENV 12661, 1998b.
- Frank, A. U., "Qualitative Spatial Reasoning: cardinal directions as an example", *International Journal of Geographical Information Systems*, 10, 3 (1996), 269-290.
- Hong, J.-H., *Qualitative Distance and Direction Reasoning in Geographic Space*, PhD thesis, University of Maine, 1994.
- ISL, *Clementine, Reference Manual, Version 5.0*, 1998a.
- ISL, *Clementine, User Guide, Version 5.0*, Integral Solutions Limited, 1998b.
- Papadias, D., e T. Sellis, "On the Qualitative Representation of Spatial Knowledge in 2D Space", *Very Large Databases Journal, Special Issue on Spatial Databases*, 3, 4 (1994), 479-516.
- Santos, M., L. Amaral, e P. Pimenta, *A Descoberta de Conhecimento em Bases de Dados Geográficas através da Explicação Semântica*, GISBrasil'99 - V Congress and Exhibition of Latin America Geo-processing Users, Salvador, Brasil, 19-23 July, 1999.
- Schenck, D., e P. Wilson, *Information Modeling: The EXPRESS Way*, Oxford University Press, Oxford, 1994.
- Sharma, J., *Integrated Spatial Reasoning in Geographic Information Systems: Combining Topology and Direction*, Tese de Doutorado, University of Maine, 1996.