
Enhancing Spatial Association Rule Mining in Geographic Databases

Vania Bogorny

Instituto de Informática - UFRGS

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Roteiro

- Introdução e Motivação
- Objetivos
- Proposta e Principais Contribuições
- Considerações Finais

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Bancos de Dados Geográficos

Street

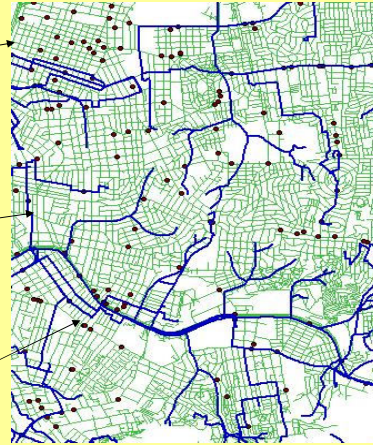
Gid	Name	Shape
1	Ijuí	Multiline [(x1,y1),(x2,y2),...]
2	Lavras	Multiline [(x1,y1),(x2,y2),...]

WaterResource

Gid	Name	Shape
1	Jacuí	Multiline [(x1,y1),(x2,y2),...]
2	Guaíba	Multiline [(x1,y1),(x2,y2),...]
3	Uruguai	Multiline [(x1,y1),(x2,y2),...]

GasStation

Gid	Name	VolDiesel	VolGas	Shape
1	BR	20000	85000	Point[(x1,y1)]
2	IPF	30000	95000	Point[(x1,y1)]
3	Esso	25000	120000	Point[(x1,y1)]



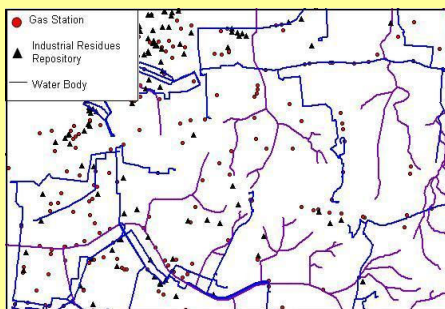
Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

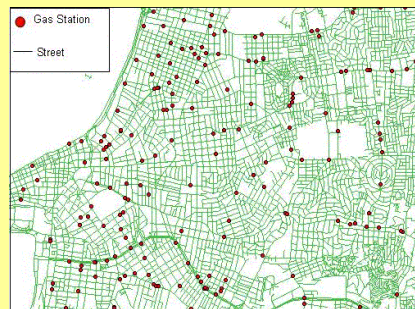
Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Dependências Geográficas

Relacionamentos espaciais não óbvios



Dependências bem conhecidas



contains(GasStation) → contains(Street) (100%)
contains(BusStop) → intersects(Street) (100%)

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Regras de Associação (Apriori)

$X \rightarrow Y$

$Suporte = \#(X \cup Y) / \# total$

$Confiança = s(X \cup Y) / s(X)$

Dataset

Tid	Itemset
1	A, C, D, T, W
2	C, D, W
3	A, D, T, W
4	A, C, D, W
5	A, C, D, T, W
6	C, D, T

1) Conjuntos Frequentes

Set k	Alguns conjuntos frequentes (suporte 50%)
k=1	{A}, {C}, {D}, {T}, {W}
k=2	{A,C}, {A,D}, {A,T}, {A,W}, {C,D},
k=3	{A,C,D}, {A,C,W}, {A,D,T}, {A,D,W},
k=4	{A,C,D,W}, {A,D,T,W}

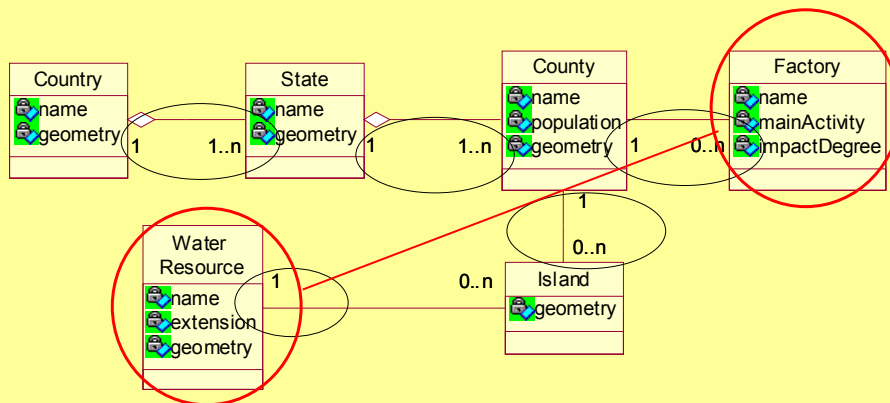
2) Regras de Associação

$A \rightarrow C$ (s= 3/6) (c=3/4)
 $AD \rightarrow W$ (s=4/6) (c=4/4)

Objetivos

- Utilizar conhecimento do domínio para reduzir o número de regras sem conhecimento novo
- Tornar mais eficiente o processo de mineração de RA

Dependências Geográficas em Esquemas de BD

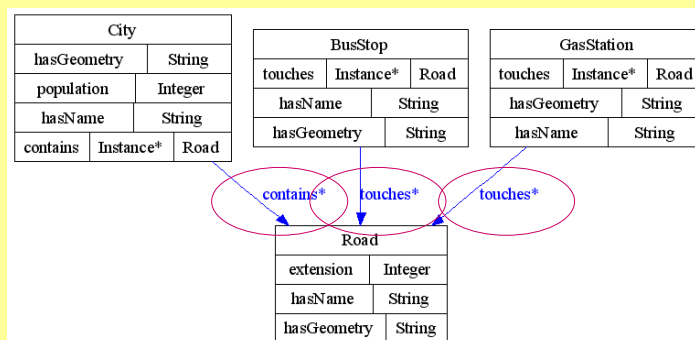


Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Dependências Geográficas em Ontologias



Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Conjunto das Contribuições

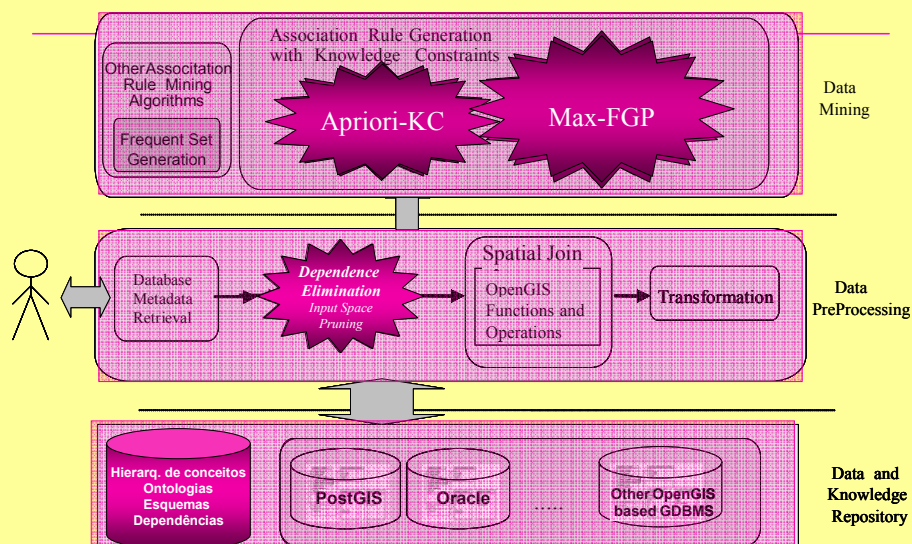
1. Dependências entre o objeto de referência e os objetos relevantes
2. Dependências entre objetos relevantes
3. Regras redundantes
4. Dependências hierárquicas, mineração em diferentes níveis de granularidade
5. Pares de elementos com o mesmo tipo de objeto geográfico
6. Otimização do pré-processamento utilizando ontologias
7. Redução de dependências utilizando ontologias
8. Automatização do pré-processamento

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Visão Geral das Contribuições



Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Problema 1 - Dependências Geográficas entre o Objeto de Referência e os Objetos Relevantes

Dependência {City, TreatedWaterNet}

Tuple (city)	Spatial Predicates
1	contains(Port), contains(Hospital), contains(TreatedWaterNet), contains(Factory), crosses(WaterBody)
2	contains(Hospital), contains(TreatedWaterNet), crosses(WaterBody)
3	contains(Port), contains(TreatedWaterNet), contains(Factory), crosses(WaterBody)
4	contains(Port), contains(Hospital), contains(TreatedWaterNet), crosses(WaterBody)
5	contains(Port), contains(Hospital), contains(TreatedWaterNet), contains(Factory), crosses(WaterBody)
6	contains(Hospital), contains(TreatedWaterNet), contains(Factory)

Minconf=70%

Suporte 100%

contains(Hospital) → contains(TreatedWaterNet)

Min Sup %	Total FrequentSets/ Rules	Rules with Dependence / Rules without Dependence	FrequentSets with dependence / FrequentSets without dependence
20	31 / 180	130 / 50	16 / 15
50	25 / 96	72 / 24	13 / 12

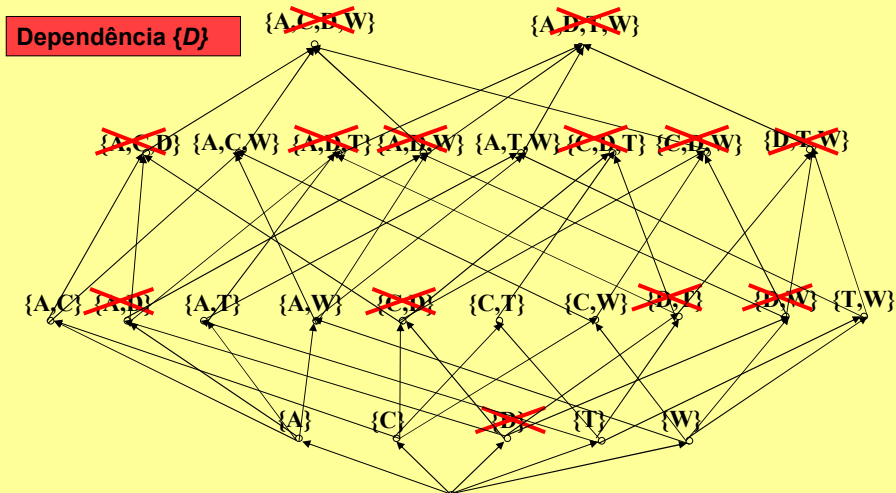
Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Contribuição 1 – Pré-processamento (método)

SEKE, 2006

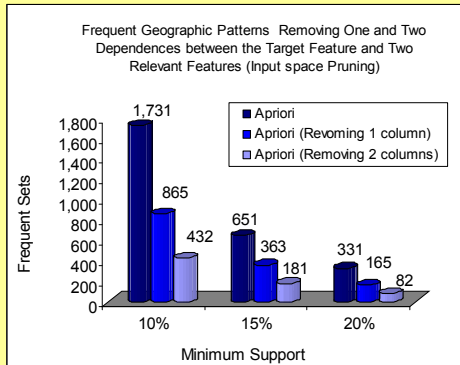


Congresso da SBC
Julho 2007

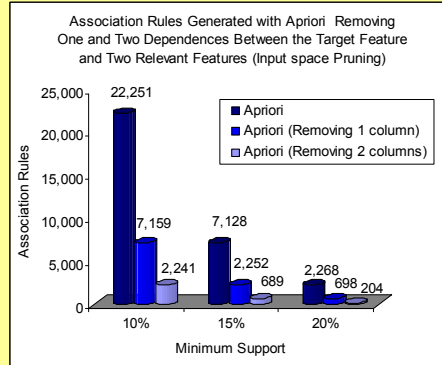
Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Contribuição 1 - Pré-Processamento (experimento)



1 dependência reduziu 50%
2 dependências reduziu 70%



1 dependência reduziu 70%
2 dependências reduziu 90%

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Problema 2 - Dependências Geográficas entre os Objetos Relevantes

Dependência {Port, WaterBody}

Tuple (city)	Spatial Predicates
1	contains(Port), contains(Hospital), contains(TreatedWaterNet), contains(Factory), crosses(WaterBody)
2	contains(Hospital), contains(TreatedWaterNet), crosses(WaterBody)
3	contains(Port), contains(TreatedWaterNet), contains(Factory), crosses(WaterBody)
4	contains(Port), contains(Hospital), contains(TreatedWaterNet), crosses(WaterBody)
5	contains(Port), contains(Hospital), contains(TreatedWaterNet), contains(Factory), crosses(WaterBody)
6	contains(Hospital), contains(TreatedWaterNet), contains(Factory)

Minsup=50%
25 Conjuntos Frequentes (6 contêm a dependência)

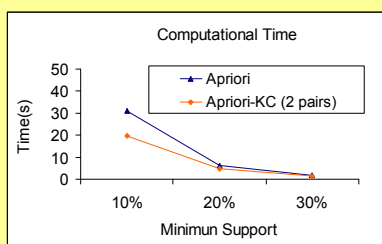
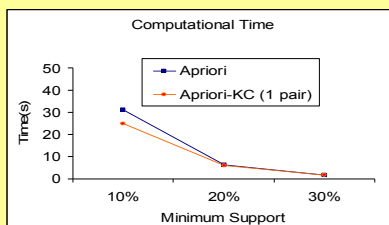
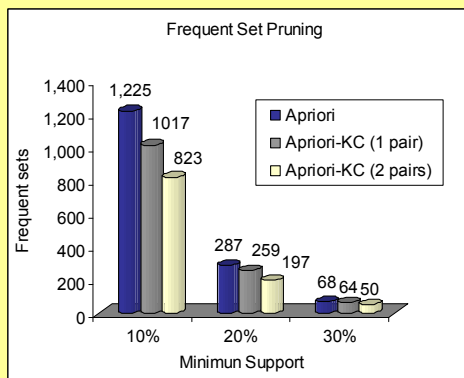
Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Contribuição 2 – Experimento

(IEEE-IS, 2006) (Springer, 2007) (IEEE-STD, 2007)



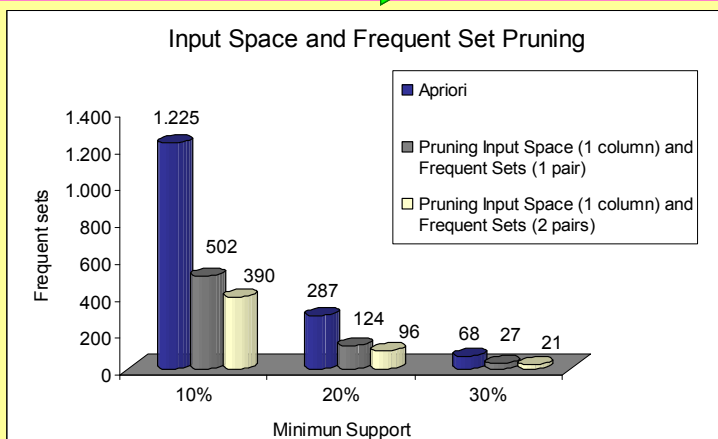
Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Eliminação Total das Dependências

(ACMGIS, 2006) (IJGIS, 2007)



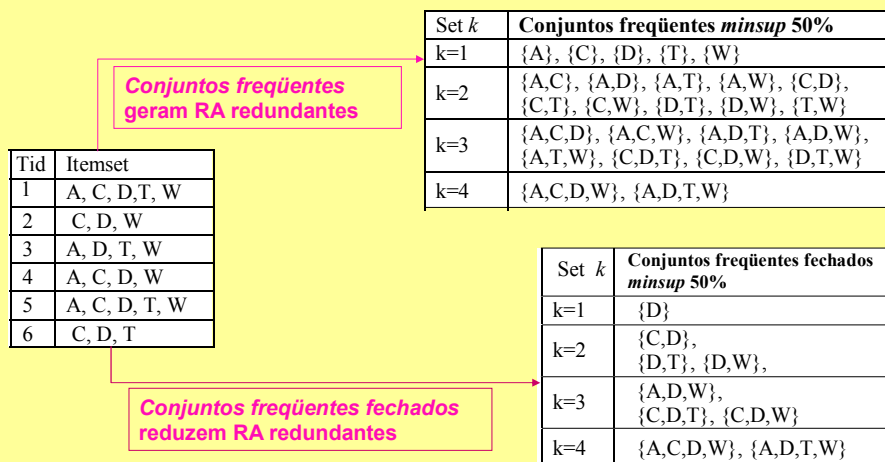
Redução de 60% e 70%

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Problema 3 – Regras Redundantes



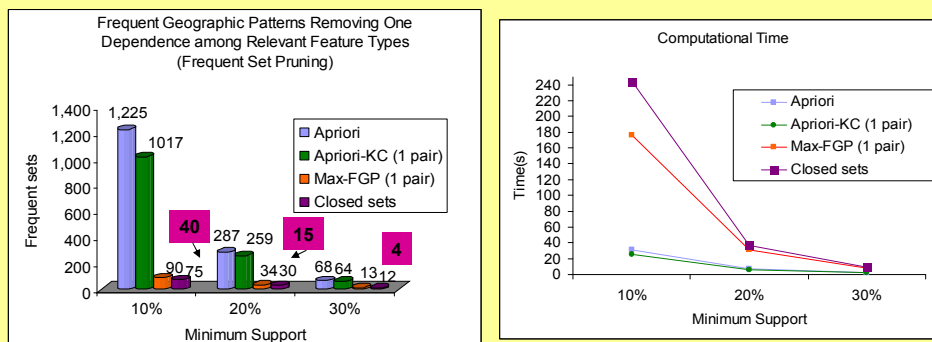
Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Contribuição 3 – Eliminação de Conjuntos Redundantes

(IEEE-ICDM, 2006)



- Suporte mínimo 10% e 20%, metade dos conjuntos fechados contém dependências
- Método proposto reduziu em mais de 80% o total de conjuntos frequentes

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Contribuições Adicionais

- O uso de ontologias como base de conhecimento reduz:
 - tempo de pré-processamento (junção espacial) [\(KDO, 2005\)](#)
 - número de padrões bem conhecidos [\(IDEA GROUP, 2007\)](#)
- Automatização do pré-processamento (Weka-GDPM) [\(WAAMD, 2006\)](#)
 - geração automática de dados em 2 níveis de granularidade sem necessidade de hierarquias de conceito [\(SEKE, 2005\)](#)
 - interface para definição e eliminação automática de dependências entre o objeto de referência e objetos relevantes [\(WSL, 2007\)](#)

Trabalhos Relacionados

	Pré-processamento	Dependências Geográficas	Regras com o mesmo tipo de feição geográfica	Conjuntos redundantes
Koperski, 1995 Clementini, 2000 Mennis, 2005	calcula relacionamentos mais genéricos e depois mais específicos	-	-	-
Appice e Malerba 2003, 2005	calcula todos os relacionamentos	elimina a posteriori	-	-
Huang 2004, Yoo, 2005, 2006	distância	-	-	-
Abordagens Conjuntos Fechados	-	-	-	elimina redundâncias mas não dependências
Nossa Proposta	ontologia- calcula relacionamentos semanticamente consistentes dependências – elimina dependências	elimina a priori	elimina a priori	gera conjuntos máximos sem dependências e sem redundâncias

Considerações Finais

- ♦ Foram identificados vários aspectos que geram regras de associação sem conhecimento novo
- ♦ Apresentamos diferentes abordagens em que o conhecimento do domínio é utilizado para reduzir o número de regras com padrões óbvios bem como o número final de regras
- ♦ A maior vantagem da solução proposta é a sua simplicidade:
 - ♦ dependências e combinações que geram padrões não interessantes são eliminados de forma eficiente e eficaz

Publicações (Nacionais)

1. Bogorny, V.; Palma, A.; Kuijpers, B.; Alvares, L.O.: *Spatial Data Mining: From Theory to Practice with Free Software*. In: International Workshop on Free Software **WSL**. Porto Alegre, 2007.
2. Bogorny, V.; Palma, A.; Engel, P.; Alvares, L.O. Weka-GDPM: Integrating Classical Data Mining Toolkit to Geographic Information Systems. In: SBBB Workshop on Data Mining Algorithms and Applications, **WAAMD**, Florianópolis, 2006.
3. Bogorny, V.; Engel, P.; Alvares, L. O.. *Spatial Data Preparation for Knowledge Discovery*. In: **IFIP Academy On The State Of Software Theory And Practice - Phd Colloquium**. Porto Alegre, Brazil 2005.
4. Bigolin, N. M.; Bogorny, V.; Alvares, L. O. A Query Language for Data Mining in Object Oriented Geographic Databases. In: **Clei XXIV** - Conferência Latino-Americana de Informática. La Paz: Universidad Mayor de San Andrés, 2003.
5. Bogorny, V.; Nievinski, F.; Bigolin, N. Um modelo de dados geográficos para integração de dados de saúde pública. In: **WSL**. Porto Alegre: 2003. p.119 - 122.

Publicações (Workshops Internacionais)

1. Bogorny, V.; Engel, P. M.; Alvares, L.O. Towards the Reduction of Spatial Join for Knowledge Discovery in Geographic Databases using Geo-Ontologies and Spatial Integrity Constraints. In: WORKSHOP ON KNOWLEDGE DISCOVERY AND ONTOLOGIES OF THE **ECML/PKDD**, **KDD**, 2., Porto, 2005.
2. Bogorny, V.; Moelans, B.; Alvares, L.O.: *Filtering frequent geographic patterns with qualitative spatial reasoning*. In: WORKSHOP ON SPATIO-TEMPORAL DATA MINING OF THE **IEEE-ICDE** **STDM**. Istanbul, 2007.

Publicações (Conferências Internacionais Qualis B)

1. Bogorny, V.; Engel, P. M.; Alvares, L.O. GeoARM: an interoperable framework to improve geographic data preprocessing and spatial association rule mining. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, **SEKE**, 18., San Francisco, 2006.

Publicações (Conferências Internacionais Qualis A)

1. Bogorny, V.; Valiati, J.; Camargo, S.; Engel, P.; Alvares, L. O.: Mining Maximal Generalized Frequent Geographic Patterns with Knowledge Constraints. In: IEEE INTERNATIONAL CONFERENCE ON DATA MINING, **IEEE-ICDM**, 6, Hong-Kong, 2006. (tx. Act. < SIGKDD, 15%)
2. Bogorny, V.; Camargo, S.; Engel, P. M.; Alvares, L.O. Towards elimination of well known geographic domain patterns in spatial association rule mining. In: IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS, **IEEE-IS**, 3, London. IEEE Computer Society, 2006.
- ★ 3. Bogorny, V.; Camargo, S.; Engel, P. M.; Alvares, L.O. Towards elimination of redundant and well known geographic domain patterns in spatial association rule mining. In: Selected for the special edition to be published by **SPRINGER-VERLAG** in 2007.
4. Bogorny, V.; Camargo, S.; Engel, P.; Alvares, L. O.: Mining Frequent Geographic Patterns with Knowledge Constraints. In: ACM INTERNATIONAL SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, **ACM-GIS**, 14., Arlington, 2006.
5. Bogorny, V.; Engel, P. M.; Alvares, L.O.: A Reuse-Based Spatial Data Preparation Framework for Data Mining. In INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, **SEKE**, 17., Taipei, Taiwan, 2005.

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Publicações (Revista e Capítulos de Livro)

1. Bogorny, V.; Kuijpers, B.; Alvares, L.O.: *Reducing Non-Interesting Spatial Association Rules in Geographic Databases using Background Knowledge: a Summary of Results*. In: **IJGIS** INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 2007. Taylor and Francis (Fator de impacto 1.36).
1. Bogorny, V.; Engel, P. M.; Alvares, L.O. Enhancing the Process of Knowledge Discovery in Geographic Databases using Geo-Ontologies. In: NIGRO, H. O.; CISARO, S.G.; XODO, D. (Ed.). Data Mining with Ontologies: Implementations, Findings, and Frameworks. **Idea Group**, 2007 (in press).
2. Rinzivillo, S., Turini, F., Bogorny, V., Koerner, C., Kuijpers, B. and May, M. Knowledge Discovery from Geographical Data, In Mobility, Privacy, Geography: a knowledge discovery perspective, 2007, **Springer**. (in Press).

Congresso da SBC
Julho 2007

Concurso de Teses e Dissertações

Apoio: CAPES, CNPq, UFRGS,
GeoPKDD e SBC

Trabalhos Derivados

Total = 16 publicações com comitê de programa

Graduação:

- Leandro Gonçalves (UFRGS)

Mestrado:

- Anke Jager (Universidade de Hasselt, Bélgica)

- Andrey Palma (UFRGS)

Pós-doutorado:

Convite para Pós-doutorado de 2 anos (Universidade de Hasselt) no contexto do projeto europeu GeoPKDD (*Geographic Privacy-Aware Knowledge Discovery and Delivery*)

Resultados Adicionais

1. Membro de Comitê de programa:
IEEE-ICDE (STDM 2007)
IEEE-ICDM (SSTDM 2007)
2. Membro de Comitê de programa Nacionais
GeoInfo 2007
WTIC 2007
3. Revisor de artigos da revista IJGIS