

DENDROIDH: AGRUPANDO CIDADES POR SEMELHANÇA DE INDICADORES

DENDROHDI: CLUSTERING CITIES BY SIMILARITY INDICATORS

DENDROIDH: INTERCALACIÓN DE CIUDADES POR INDICADORES DE SIMILITUD

Ricardo Alexandre Afonso¹, Raquel da Silva Cabral¹,
Vinicius Cardoso Garcia² e Alexandre Álvaro³

Resumo: Objetivo: as cidades inteligentes estão surgindo mediante a necessidade de otimização de recursos e ampliação do bem-estar dos seus habitantes. Atualmente não existem dados claros sobre como comparar cidades inteligentes com base em indicadores que utilizem dados públicos, principalmente na área de Saúde. **Método:** este trabalho propõe a utilização de indicadores de IDH para comparar estatisticamente e agrupar cidades com semelhança de indicadores, e assim, oferecer aos seus gestores, a possibilidade de adotar estratégias de gestão baseadas em visualização de dados dispostos em dendrogramas. **Resultado:** foram realizados cálculos com a utilização de uma ferramenta estatística em bases de dados públicas para obter dendrogramas de dados. **Conclusão:** o agrupamento de cidades por semelhança de indicadores se mostrou promissor para comparar e medir cidades com semelhantes características.

Palavras-chave: Mineração de Dados; Cidades Saudáveis; Base de Dados.

Abstract: Objective: Smart cities are emerging by the need to optimize resources and expansion of the welfare of its inhabitants. Currently there are no clear data on how to compare smart cities based on indicators using public data, especially in the health area. **Method:** this paper proposes the use of HDI indicators to compare statistically and group towns with similar indicators, and thus, offer its managers, the possibility of adopting management strategies based on data visualization arranged in dendrograms. **Result:** Calculations were performed using a statistical tool in public databases for dendrograms data. **Conclusion:** the grouping of cities by similarity indicators showed promise to compare and measure cities with similar characteristics.

Keywords: Data Mining; Healthy Cities; Data base.

Introdução

Segundo Chourabi⁽¹⁾ as Cidades Inteligentes estão surgindo na medida em que as soluções tecnológicas permitem ampliar o poder de otimizar recursos e melhorar os processos gerenciais. Entretanto, essas soluções necessitam de bases de dados e de tratamento para que esses dados se tornem semanticamente passíveis de entendimento, e recuperação de informações, seja isso automatizado ou não. E a visualização destes dados pelos seus gestores precisa ser simples e intuitiva.

1 Professor Adjunto do NCEX (Núcleo de Ciências Exatas), UFAL, Arapiraca, AL, Brasil.

2 Professor Adjunto do CIn (Centro de Informática), UFPE, Recife, PE, Brasil.

3 Professor Adjunto do DComp (Departamento de Computação), UFSCar, Araraquara, SP, Brasil.

É possível afirmar que ao comparar cidades com características socioeconômicas e políticas semelhantes, a utilização dos meios de gestão democráticos depende diretamente da característica de cada região, de modo que indicadores de saúde pública revelem nuances do desenvolvimento dessas cidades ⁽²⁾.

A análise desenvolvida neste trabalho representa a busca pela compreensão sobre a relação entre Território, População e Desenvolvimento, a partir de indicadores analíticos que articulam estas três dimensões visando encontrar um padrão de semelhança entre os municípios avaliados. Esta análise baseia-se em evidências empíricas provenientes de dados coletados em um conjunto de dados envolvendo aglomerações de cidades brasileiras por semelhança de variáveis encontradas em bases de dados de saúde pública ⁽⁷⁾.

Este trabalho encontra-se organizado da seguinte forma: a próxima seção apresenta o Método conceitual que serve de base para o estudo sobre a relação entre as dimensões estabelecidas para comparação de cidades. Na Seção seguinte (Resultados e Discussão) apresenta-se o comportamento da amostra de dados, com aplicação ao conjunto de cidades do estado de Alagoas que serviu como exemplo de aplicação deste método. E na Seção final (Conclusão) são apresentadas as considerações finais.

Métodos

Essa Seção apresenta os indicadores utilizados para comparar cidades inteligentes brasileiras, e enfoca sobretudo os indicadores de IDH para pontuar o desenvolvimento das cidades. Para permitir uma real comparação entre as cidades e suas características individuais, foram propostas três dimensões (Território, População e Desenvolvimento), com suas respectivas variáveis (Tabela 1):

(A) Dimensão Território: esta dimensão utiliza as variáveis Área Territorial e Densidade Demográfica. Esta dimensão tem como objetivo comparar o território de cada município, uma vez que a área total pode ter influência direta na forma como serão implementadas as estratégias de infraestrutura e serviços públicos em detrimento da extensão deste município;

(B) Dimensão População: nesta dimensão são apresentados os dados referentes às variáveis de População Urbana e Rural. Com a comparação destas variáveis, é possível compreender aspectos referentes a urbanização de cidades e os fatores que possam impedir um comparativo com cidades de características mais agrárias;

(C) Dimensão Desenvolvimento: nesta dimensão as variáveis de desenvolvimento humano municipal são detalhadas em Renda, Longevidade e Educação. Esta dimensão representa a comparação mais complexa entre os municípios, pois engloba variáveis que embora sejam distintos compõem uma fórmula (IDH) responsável por indicar o desenvolvimento do município.

Tabela 1: Características dos domínios e variáveis de comparação

Dimensão	Variável	Dado	Fonte
Território	Área Territorial	km ²	IBGE ⁽⁹⁾
	Densidade demográfica	hab./km ²	
População	População Urbana População Rural	Habitantes	IBGE ⁽⁵⁾
Desenvolvimento	Renda	Idh-r	PNUD ⁽¹⁰⁾
	Longevidade	Idh-l	
	Educação	Idh-e	

Dimensão: (A) Território

Segundo estatísticas do Instituto Brasileiro de Geografia e Estatística ⁽⁹⁾, o Brasil possui uma população de 202.768.562 habitantes, distribuídos em uma área de 8.515.767,049 km², que resulta em uma densidade demográfica de 23,8 habitantes por quilômetro quadrado. Em virtude da história econômica do Brasil, as maiores taxas de densidade demográfica do Brasil encontram-se na região Sudeste, seguida pelo Sul e pelo Nordeste (Figura 4.1), sendo respectivamente (em hab/km²): Sudeste (67,77), Sul (38,38), Nordeste (27,33), Centro-Oeste (5,86) e Norte (2,66).

Dimensão: (B) População

Nesta dimensão são exploradas as variáveis referentes a quantidade de habitantes nas zonas urbanas e rurais das cidades avaliadas. Atualmente, é uma característica da maior parte dos municípios brasileiros, que a população que vive em áreas rurais seja bem menor do que a que vive em áreas urbanas, embora até a década de 1950, a população rural tenha sido consideravelmente maior do que a urbana ⁽⁵⁾.

Dimensão: (C) Desenvolvimento

O Índice de Desenvolvimento Humano (IDH) é uma medida resumida do progresso a longo prazo em três dimensões básicas do desenvolvimento humano: renda, educação e saúde. O IDH pretende ser uma medida geral e sintética que, apesar de ampliar a perspectiva sobre o desenvolvimento humano, não abrange nem esgota todos os aspectos de desenvolvimento ⁽⁶⁾.

Desde 2010, quando o Relatório de Desenvolvimento Humano completou 20 anos, novas metodologias foram incorporadas para o cálculo do IDH. Atualmente, os três pilares que constituem o IDH (saúde, educação e renda) são mensurados da seguinte forma: **a)** Uma vida longa e saudável (saúde) é medida pela expectativa de vida; **b)** O acesso ao conhecimento (educação) é medido por: i) média de anos de educação de adultos, que é o número médio de anos de educação recebidos durante a vida por pessoas a partir de 25 anos; e ii) a expectativa de anos de escolaridade para crianças na idade de iniciar a vida escolar; **c)** O padrão de vida (renda) é medido pela Renda Nacional Bruta (RNB) per capita expressa em poder de paridade de compra (PPP) constante, em dólar, tendo 2005 como ano de referência.

A próxima Seção apresenta uma aplicação deste catálogo de métricas que utiliza o IDH para comparar cidades, e de que forma a utilização da técnica de análise multivariada pode auxiliar no agrupamento de cidades por semelhanças.

Resultados e Discussão

O objetivo desta seção é apresentar a aplicação de catálogo métricas usando Análise Multivariada e Análise de Cluster para uma comparação das cidades com base em áreas previamente estabelecidas. Esta prática permite selecionar cidades com características semelhantes, como se pretende mostrar através da comparação e agrupamento do IDH.

Análise Multivariada

A análise multivariada de uma forma bem geral refere-se a todos os métodos estatísticos que analisam simultaneamente múltiplas medidas em cada indivíduo ou objeto sob investigação ⁽³⁾. O resultado destas análises pode ser o agrupamento de indivíduos por semelhança. A análise de clusters normalmente envolve dois passos. O primeiro é a medida de alguma forma de semelhança ou associação entre as entidades para determinar quantos grupos existem, realmente, na amostra e esse será chamado nessa análise de Passo 1.

Passo 2: (Analisando Clusters)

Uma forma de detalhar melhor e realizar uma análise mais aprofundada dos agrupamentos de cidades por semelhança de indicadores é separá-los em agrupamentos por amostragem. A Figura 2 apresenta o agrupamento das 50 cidades alagoanas de acordo com a dimensão Desenvolvimento (Educação, Longevidade e Emprego).

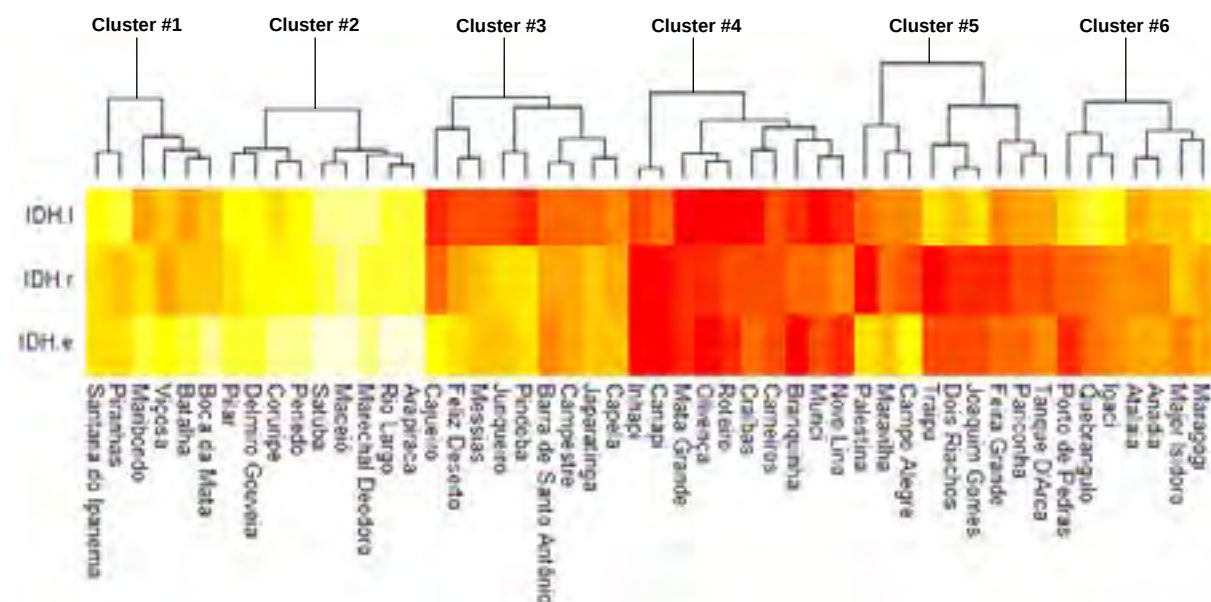


Figura 2: Detalhamento do *Cluster* de cidades por semelhança de IDH

Nesse agrupamento a granularidade de variáveis é maior, e com isso, a semelhança entre municípios se torna mais evidente na medida em que a variação entre dados é menor. Observa-se através da Tabela 2 que o *Cluster* #5 se destaca com a maior média, e nesse caso, representa os municípios com melhores variáveis que indicam áreas de melhor desenvolvimento no estado.

Tabela 2 – Média e variação do agrupamento de Desenvolvimento.

Cluster #1 – Média 2,78 ± 0,081
Maragogi, Major Isidoro, Anadia, Atalaia, Igaci, Quebrangulo, Porto de Pedras
Cluster #2 – Média 2,58 ± 0,132
Tanque D'Arca, Pariconha, Feira Grande, Dois Riachos, Traipu, Campo Alegre, Maravilha, Palestina
Cluster #3 – Média 2,09 ± 0,075
Novo Lino, Murici, Branquinha, Carneiros, Craibas, Roteiro, Olivença, Mata Grande, Canapi, Inhapi
Cluster #4 – Média 2,67 ± 0,102
Capela, Japaratinga, Barra de Santo Antônio, Pindoba, Junqueiro, Messias, Feliz Deserto, Cajueiro
Cluster #5 – Média 3,62 ± 0,210
Arapiraca, Rio Largo, Marechal Deodoro, Maceió, Satuba, Penedo, Coruripe, Delmiro Gouveia, Pilar
Cluster #6 – Média 3,06 ± 0,063
Boca da Mata, Batalha, Viçosa, Maribondo, Piranhas, Santana do Ipanema

Conclusão

As Cidades Inteligentes estão emergindo, e junto com elas surge a necessidade de obter e transformar os dados gerados por elas para melhorar a vida dos seus habitantes. O objetivo deste trabalho foi apresentar um catálogo de métricas que comparou cidades brasileiras, e em especial as cidades nordestinas, considerando para isso três dimensões (Território, População e Desenvolvimento). Os dados dessas cidades foram agrupados de acordo com a semelhança de variáveis de cada domínio e os dados resultantes foram apresentados em forma de dendrogramas, que foram o objeto de estudo deste trabalho, focando os aspectos de Desenvolvimento das cidades baseadas nos indicadores de desenvolvimento humano (IDH).

Pretende-se com esse trabalho apresentar uma possibilidade de medição e comparação de cidades inteligentes, utilizando para isso esses domínios pré-estabelecidos (Território, População e Desenvolvimento). Dessa forma, a comparação de cidades vai atender a um padrão de comparação mais igualitária considerando assim, as diferenças regionais, de perfil populacional e desenvolvimento da região na qual o município está inserido.

Como trabalho futuro, estão sendo desenvolvidas rotinas de conversão como um serviço na web, no qual o usuário indicará as fontes de dados e a rotinas produzirão os gráficos com os resultados da comparação dos dados.

Agradecimentos

Este trabalho foi parcialmente financiado pela FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco), concessão IBPG-0499-1.03/11.

Referências

- [1] Chourabi, H.; et al. "Understanding Smart Cities: An Integrative Framework," System Science (HICSS), 2012 45th Hawaii International Conference on , vol., no., pp.2289-2297, 4-7 Jan. 2012. doi: 10.1109/HICSS.2012.615.
- [2] Santos, M. A Natureza do Espaço: Técnica e Tempo, Razão e Emoção (4 ed. Vol. 2. reimpr). 2006. São Paulo: Editora da Universidade de São Paulo.
- [3] Field, A. Descobrendo estatística utilizando o SPSS. Tradução Lorí Viali. 2. ed. Porto Alegre: Artmed, 2009. 688p.
- [4] Britto, J., Stallivieri, F., Campos, R. R., & Vargas, M. Padrões de aprendizagem, cooperação e inovação em aglomerações produtivas no Brasil: uma análise multivariada exploratória. 2007. ENCONTRO NACIONAL DE ECONOMIA-ANPEC, Recife.
- [5] IBGE. Censo Populacional 2010 Censo Populacional 2010 Instituto Brasileiro de Geografia e Estatística (IBGE).
- [6] Frei, Fernando. Índice de Desenvolvimento Municipal-IDM: uma alternativa para mensuração do desenvolvimento humano nos municípios do Estado de São Paulo. Diss. Universidade de São Paulo. Faculdade de Saúde Pública. Departamento de Epidemiologia, 2002.
- [7] Afonso, R. A., Silva, W. D., Tomas, G. H. R. P., Gama, K., Oliveira, A., Álvaro, A., & Garcia, V. C. Br-SCMM: Modelo Brasileiro de Maturidade para Cidades Inteligentes. 2013. Simpósio Brasileiro De Sistemas De Informação - SBSI.
- [8] Racine, Jeffrey S. RStudio: A Platform-Independent IDE for R and Sweave. Journal of Applied Econometrics, v. 27, n. 1, p. 167-172, 2012
- [9] IBGE. IBGE apresenta nova área territorial brasileira [Internet]. [Atualizado 2016 Junho 26; Citado 2016 Set 06] Disponível em: <http://goo.gl/lgPqWI>

[10] PNUD, IPEA. FJP. Atlas do Desenvolvimento Humano no Brasil [Internet]. [Atualizado 2016 Set 06; Citado 2016 Set 06] Disponível em: <http://goo.gl/cdSocd>

Contato

Ricardo Alexandre Afonso
Rua Amélia Nunes Correia, 200
Arapiraca – Alagoas
(82) 9184-8184
ricardo.afonso@arapiraca.ufal.br

