

Nonlinear dimensionality reduction through optimal transport between incomparable spaces

Por Rafael Pereira Eufrazio, Eduardo Fernandes Montesuma e Charles C. Cavalcante

Fonte: OpenAlex · IFCE · [Publicação original](#)

INOVAÇÃO	APLICABILIDADE	POTENCIAL ECONÔMICO	MATURIDADE
8/10	7/10	7/10	Média

Proposta do método GW-MDS (Gromov-Wasserstein Multidimensional Scaling) para redução dimensional não linear em espaços incomparáveis. A técnica utiliza transporte ótimo para preservar a geometria relacional dos dados, superando limitações de métodos tradicionais como MDS e Isomap, especialmente em regimes de poucas amostras.

NÚCLEO DA ANÁLISE

Ideia de startup ou produto

Startup de 'Data Fusion' focada em uma API que utiliza GW-MDS para permitir que empresas de HealthTech e Farmacêutica cruzem dados clínicos e moleculares de diferentes plataformas para descoberta de padrões em doenças raras.

Aplicações práticas

Alinhamento de dados em bioinformática (perfis genômicos distintos), integração de sensores em IoT industrial com diferentes métricas, e sistemas de recomendação que necessitam comparar itens com atributos não sobrepostos.

Potencial de mercado

Alto. A capacidade de processar e alinhar dados heterogêneos é crítica para avanços em IA Multimodal e Indústria 4.0, onde dados de diferentes fontes precisam ser comparados sem transformações forçadas.

FUNDAMENTO CIENTÍFICO

Problema abordado

Dificuldade de alinhar e visualizar dados provenientes de espaços heterogêneos ou incomparáveis (onde não há um espaço de características comum), além da vulnerabilidade de métodos clássicos à 'maldição da dimensionalidade' e baixa robustez em datasets pequenos.

Metodologia

Desenvolvimento de uma formulação probabilística que utiliza a distância Gromov-Wasserstein para criar correspondências suaves (soft correspondences) entre espaços. Implementação de uma variante semi-relaxada (srGW-MDS) que fixa apenas a marginal da distribuição original, integrando distâncias geodésicas baseadas em grafos e Euclidianas via um esquema de otimização alternada.

Principais descobertas

O método srGW-MDS demonstra alta robustez em cenários de baixa amostragem (low-sample regimes) e capacidade superior de preservar estruturas geodésicas complexas em comparação com abordagens rígidas de mapeamento um-para-um.

LEITURA PARA GESTÃO PÚBLICA

Criação de sistemas integrados de inteligência urbana que permitem cruzar dados de mobilidade, saúde e economia (que possuem estruturas distintas) para otimizar políticas públicas setoriais sem a necessidade de padronização prévia complexa.

PARCERIA UNIVERSIDADE × EMPRESA

Colaboração entre o IFCE e empresas de tecnologia de grande porte (corporates) ou hospitais de pesquisa para validação do algoritmo srGW-MDS em bases de dados reais e escassas, acelerando a transferência de tecnologia.

OPORTUNIDADES DERIVADAS

3 direções estratégicas identificadas

STARTUP · IMPACTO ALTO · CIÊNCIA DE DADOS

HeteroData Align

Spin-off comercializando bibliotecas de software especializadas em redução dimensional para datasets incomparáveis, focada inicialmente no setor de P&D farmacêutico.

PARCERIA · IMPACTO MÉDIO · INDÚSTRIA 4.0

Parceria IFCE e Indústria 4.0

Projeto colaborativo para aplicar a técnica de transporte ótimo na calibração e fusão de sensores em linhas de produção inteligentes onde os espaços de medição variam ao longo do tempo.

POLÍTICA PÚBLICA · IMPACTO MÉDIO · GOVTECH

Integração de Dados Governamentais

Implementação de piloto para unificar bases de dados de secretarias estaduais distintas, permitindo visualização relacional de problemas sociais sem a necessidade de unificação forçada de schemas.

ABSTRACT ORIGINAL

Abstract We study a dimensionality reduction method grounded in optimal transport theory, leveraging the Gromov–Wasserstein (GW) distance to align the relational geometry of data across incomparable spaces. Building on recent advances that connect GW and semi-relaxed GW formulations with multidimensional scaling, we adopt a probabilistic perspective in which MDS and Isomap can be viewed through soft correspondences defined by optimal transport plans rather than rigid one-to-one mappings. Our approach, termed Gromov–Wasserstein MDS (GW-MDS), seeks a low-dimensional embedding whose pairwise distances match the structural relationships of the original space under a GW objective. We also investigate a semi-relaxed variant (srGW-MDS) that fixes only the marginal over the original data distribution, allowing greater flexibility in the embedding space. We further analyze the integration of Euclidean and graph-based geodesic distance within this framework and provide a convergence analysis for the proposed alternating optimization scheme. Experiments on synthetic manifolds and real datasets demonstrate the impact of geodesic structure preservation and highlight the robustness of the semi-relaxed formulation in low-sample regimes.

MATÉRIA PARA LEIGOS

Para leigos: Simplificando dados complexos preservando suas conexões estruturais

O cenário atual

No mundo atual, lidamos com uma enorme quantidade de informações que muitas vezes possuem diversas características simultâneas, um problema conhecido como "maldição da dimensionalidade". Para visualizar e entender esses dados, é necessário reduzir sua complexidade, transformando-os em versões mais simples (com menos dimensões) sem perder suas características essenciais. Métodos tradicionais já existem, mas podem encontrar dificuldades ao tentar comparar espaços de dados que não têm uma correspondência direta óbvia entre si.

O que os pesquisadores fizeram

Pesquisadores do IFCE desenvolveram um novo método de redução de dimensionalidade não linear baseado na teoria de transporte ótimo. O trabalho se utiliza da distância de Gromov–Wasserstein (GW) para alinhar a geometria relacional dos dados, mesmo quando esses dados estão em espaços que não são diretamente comparáveis. Eles propuseram o método Gromov–Wasserstein MDS (GW-MDS) e uma variante chamada GW-MDS semi-relaxado (srGW-MDS), conectando essas formulações com técnicas clássicas como o escalonamento multidimensional (MDS) e o Isomap.

Como funciona na prática

A abordagem inova ao adotar uma perspectiva probabilística. Em vez de forçar mapeamentos rígidos de "um para um" entre pontos de dados, o método utiliza "correspondências suaves" definidas por planos de transporte ótimo. O GW-MDS busca criar uma representação simplificada (embedding) onde as distâncias entre os pontos mantenham as mesmas relações estruturais do espaço original, sob um objetivo de transporte ótimo.

Já a variante semi-relaxada (srGW-MDS) fixa apenas a distribuição dos dados originais, permitindo maior flexibilidade no espaço de destino. O estudo também integra o uso de distâncias euclidianas e distâncias geodésicas (que calculam o caminho mais curto em uma superfície curva, como a terra) dentro desse framework.

Resultados e evidência

Os pesquisadores conduziram experimentos utilizando variedades sintéticas (dados simulados) e conjuntos de dados reais. Os testes demonstraram o impacto positivo da preservação da estrutura geodésica. Um resultado

importante observado foi a robustez da formulação semi-relaxada (srGW-MDS) em regimes de baixa quantidade de amostras, ou seja, o método se manteve eficaz mesmo quando havia poucos dados disponíveis para análise. O estudo também forneceu uma análise matemática de convergência para o esquema de otimização proposto.

Implicações práticas

Esta técnica oferece uma ferramenta poderosa para simplificar dados complexos mantendo a integridade de suas conexões internas. A capacidade de trabalhar com espaços "incomparáveis" permite analisar estruturas de dados que não compartilham as mesmas medidas ou escalas diretas. A robustez do método com poucos dados é particularmente valiosa para aplicações práticas onde a coleta de grandes volumes de informação é difícil ou cara.

Limitações e próximos passos

O paper não detalha limitações específicas do método além da necessidade de realizar uma análise de convergência para o esquema de otimização alternado proposto. Da mesma forma, o material fornecido não explicita quais seriam os próximos passos para o desenvolvimento futuro dessa pesquisa.

QUEM SÃO OS PESQUISADORES

Quem são os pesquisadores

O estudo foi desenvolvido por Rafael Pereira Eufrazio, Eduardo Fernandes Montesuma e Charles C. Cavalcante, todos vinculados ao IFCE (Instituto Federal do Ceará). O material fornecido não detalha a formação acadêmica específica, títulos ou trajetória profissional prévia dos autores. Limita-se a informar que eles propuseram os métodos GW-MDS e srGW-MDS, realizaram a análise de integração de distâncias e conduziram os experimentos descritos no abstract.

PALAVRAS-CHAVE

Geodesic · Embedding · Isomap · Dimensionality reduction · Nonlinear dimensionality reduction · Pairwise comparison · Multidimensional scaling · Curse of dimensionality · Robustness (evolution) · Euclidean space