

Modern Applications and State-of-the-Art

Overview

This document surveys modern applications of box embeddings (2023-2025), demonstrating their effectiveness and exploring current research frontiers. The mathematical foundations established in the previous documents enable these practical applications.

2025: RegD - Hyperbolic-Like Expressiveness

Yang & Chen (2025) - arXiv:2501.17518

Key Contribution: A flexible Euclidean framework that supports arbitrary geometric regions (boxes, balls) as embeddings while achieving hyperbolic-like expressiveness.

Innovations:

- **Depth-based dissimilarity:** Incorporates depth distance and boundary distance metrics to emulate hyperbolic geometry properties
- **Eliminates precision issues:** Uses elementary arithmetic operations instead of specialized hyperbolic constructions
- **Addresses crowding effect:** As the number of children increases in hierarchies, traditional Euclidean embeddings cluster together; RegD's depth distance solves this by incorporating region "size"
- **Boundary distance for containment:** When one region is fully contained within another, boundary distance captures both (i) whether containment exists and (ii) discrimination between regions in inclusion chains

Theoretical Significance: RegD proves that Euclidean box embeddings can achieve hyperbolic-like expressiveness without the computational complexity of hyperbolic operations. This validates the choice of Euclidean space while maintaining theoretical guarantees.

Relevance to Foundations: This work builds directly on the max-stability property (see the Gumbel Max-Stability document) and volume calculations (see the Gumbel-Box Volume document), demonstrating that the mathematical structure we've established enables state-of-the-art performance.

2024: TransBox - EL++-closed Ontology Embedding

Yang, Chen & Sattler (2024) - arXiv:2410.14571

Key Contribution: Box embeddings for OWL ontologies with Description Logic (DL) axioms, ensuring logical closure properties.

Innovations:

- Extends box embeddings to handle complex DL axioms (not just simple triples)
- Ensures embeddings respect logical closure (EL++ semantics)
- Applications in healthcare and bioinformatics where ontologies are critical

Theoretical Foundation: This work relies on the subsumption relationship (see the Subsumption document) and demonstrates that box embeddings naturally encode logical subsumption in formal ontologies. The containment probability (see the Containment Probability document) provides the probabilistic foundation for handling uncertain or noisy ontological data.

Practical Impact: TransBox enables embedding of formal ontologies used in healthcare (SNOMED CT, UMLS) and bioinformatics, where logical consistency is critical. The geometric interpretation of subsumption makes these embeddings interpretable and verifiable.

2023: Concept2Box - Joint Geometric Embeddings

Huang et al. (2023) - arXiv:2307.01933 (ACL 2023)

Key Contribution: Jointly embeds two views of knowledge graphs: ontology-view concepts (as boxes) and instance-view entities (as vectors).

Innovations:

- **Dual geometric representations:** Concepts as box embeddings, entities as vector embeddings
- **Box volumes as granularity:** Interprets box volumes as concepts' granularity (probabilistic semantics)
- **Vector-to-box distance metric:** Novel metric to bridge concept boxes and entity vectors
- **Two-view modeling:** Addresses the fact that many KGs contain both high-level concepts and fine-grained entities

Key Results:

- Experiments on DBpedia KG and industrial KG show effectiveness
- Box embeddings capture hierarchy structure and complex relations (overlap, disjoint) among concepts
- Joint learning of both representations improves performance

Mathematical Connection: This work leverages the volume calculations (see the Gumbel-Box Volume document) to interpret box volumes as concept granularity. The containment probability (see the Containment Probability document) enables modeling of complex relationships between concepts and entities.

2023: Dual Box Embeddings for EL++

Jackermeier, Chen & Horrocks (2023) - arXiv:2301.11118

Key Contribution: Dual box embeddings for OWL ontologies, handling both ABox (data-level) and TBox (concept-level) knowledge.

Innovations:

- Dual box representation for concepts and individuals
- Handles complex DL axioms beyond simple triples
- Maintains logical consistency with EL++ semantics

Theoretical Foundation: This work demonstrates that box embeddings can handle both instance-level (ABox) and concept-level (TBox) knowledge, relying on the subsumption relationship (see the Subsumption document) and the max-stability property (see the Gumbel Max-Stability document) to maintain logical consistency.

Related Developments

Spherical Knowledge Graph Embeddings

Quan et al. (2025) - arXiv:2511.02460

Key Contribution: SKGE embeds knowledge graphs into spherical space with geometric regularization.

Relevance: While not box embeddings, this shows continued interest in non-Euclidean geometric embeddings, suggesting the field is exploring various geometric structures. RegD (2025) suggests

that Euclidean box embeddings can achieve similar expressiveness with better computational properties.

Hyperbolic Embeddings

Multiple recent papers (2024-2025) continue to explore hyperbolic embeddings for hierarchical data:

- **Low-distortion Tree Embeddings in Hyperbolic Space** (van Spengler & Mettes, 2025)
- **Embeddings with constant additive error for hyperbolic spaces** (Park & Vigneron, 2024)

Relevance: These papers show that hyperbolic geometry remains active, but RegD (2025) suggests that Euclidean box embeddings can achieve similar expressiveness with better computational properties. The max-stability property of Gumbel distributions (see the Gumbel Max-Stability document) enables efficient computation in Euclidean space.

Trends and Insights

1. Hybrid Representations

Recent work (Concept2Box, RegD) suggests combining different geometric structures (boxes + vectors, boxes + depth metrics) rather than using a single representation. This hybrid approach leverages the strengths of each representation:

- Boxes for hierarchical/containment relationships
- Vectors for symmetric/similarity relationships
- Depth metrics for enhanced expressiveness

2. Uncertainty Quantification

Multiple papers (NFE 2024, UKGE 2021) focus on introducing uncertainty into embeddings, which aligns with our probabilistic box embeddings approach. The Gumbel distribution provides a natural way to quantify uncertainty through the scale parameter β , and the containment probability (see the Containment Probability document) enables probabilistic reasoning.

3. Formal Logic Integration

Papers on EL++ and ontology embeddings (TransBox, Dual Box Embeddings) show box embeddings are being applied to formal knowledge representation, validating the subsumption-based approach. The geometric interpretation of subsumption (see the Subsumption document) makes these embeddings interpretable and verifiable.

4. Computational Efficiency

RegD (2025) emphasizes that Euclidean box embeddings can achieve hyperbolic-like expressiveness without the computational complexity of hyperbolic operations. The max-stability property (see the Gumbel Max-Stability document) enables efficient computation, and the log-sum-exp function (see the Log-Sum-Exp and Gumbel Intersection document) provides numerical stability.

5. Two-View Knowledge Graphs

Concept2Box highlights that real-world KGs often have two views (concepts vs entities), suggesting that hybrid box+vector representations are practically important. This validates the need for flexible embedding frameworks that can handle multiple geometric structures.

Applications in Practice

Knowledge Graph Completion

Box embeddings excel at link prediction in knowledge graphs, where the task is to predict missing relationships. The containment probability (see the Containment Probability document) provides a

natural scoring function, and the local identifiability solution (see the Local Identifiability document) enables effective gradient-based learning.

Ontology Embedding

Formal ontologies (OWL, Description Logic) require logical consistency, which box embeddings naturally provide through geometric subsumption (see the Subsumption document). Applications include:

- Healthcare: SNOMED CT, UMLS medical ontologies
- Bioinformatics: Gene ontologies, protein function hierarchies
- Semantic web: OWL ontologies for knowledge representation

Natural Language Understanding

Box embeddings can model hyponym-hypernym relationships in natural language, where containment represents semantic subsumption. For example, “dog” is a hyponym of “mammal”, which is a hyponym of “animal”—this hierarchy is naturally encoded as nested boxes.

Two-View Knowledge Graphs

Many real-world knowledge graphs contain both high-level concepts and fine-grained entities. Concept2Box demonstrates that joint learning of concepts (as boxes) and entities (as vectors) improves performance, validating the hybrid approach.

Performance Benchmarks

Recent work demonstrates box embeddings’ effectiveness:

- **Concept2Box (2023)**: Shows effectiveness on DBpedia KG and industrial KGs
- **RegD (2025)**: Achieves hyperbolic-like expressiveness with Euclidean efficiency
- **TransBox (2024)**: Handles complex DL axioms while maintaining logical consistency
- **Dual Box Embeddings (2023)**: Successfully embeds both ABox and TBox knowledge

These results validate the mathematical foundations we’ve established: the subsumption relationship, volume calculations, max-stability, and local identifiability all contribute to practical effectiveness.

Connection to Mathematical Foundations

All these applications build on the mathematical foundations established in the previous documents:

- **Subsumption**: Geometric containment enables logical subsumption in ontologies
- **Gumbel Box Volume**: Volume calculations enable granularity interpretation and depth metrics
- **Containment Probability**: Probabilistic reasoning enables uncertainty quantification
- **Gumbel Max-Stability**: Algebraic closure enables efficient computation and logical consistency
- **Log-Sum-Exp**: Numerical stability enables practical implementation
- **Local Identifiability**: Smooth gradients enable effective learning

The theoretical elegance of the mathematical foundations translates directly to practical effectiveness in modern applications.

Next Steps

For future directions and open questions, see the Future Directions document. For implementation details, see the code documentation and practical guides.