Knowledge Graph Embeddings for Salmon Lice Treatment Recommendation

From TransE to ComplEx: Understanding KG Models with PyKeen

Leandro Stival

Wageningen University & Research (WUR)

January 14, 2026

Entities Relations Discovery

Research Journey: Solving the Puzzle

- **I. The Challenge:** Parasitic pressure and treatment paths
- **II. Methodology:** Finding the best model for aquaculture
  - **Path 1:** Translation (Baseline & Limitations)
  - **Path 2:** Rotation (Handling 1-to-N relationships)
  - **Path 3:** Bilinear (Directional & Asymmetric logic)
  - **Path 4:** Semantics (Leveraging SBERT)
- **III. Validation:** Model comparison and performance

The Challenge: Parasitic Pressure

Problem Statement:

- Salmon farming faces parasitic *Lepeophtheirus salmonis* (sea lice)
- Global economic impact and animal welfare concerns
- Multiple treatments available: chemicals, cleaner fish, thermal, mechanical

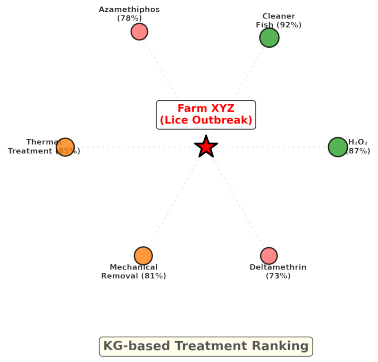
"Which treatment is best for which specific farm condition?"

The Solution: A Knowledge Graph Approach

Mapping Relationships:

- Salmon species ↔ Lice strains
- Treatments ↔ Efficacy
- Environmental factors ↔ Outbreaks

Goal: Structured link prediction for decision support

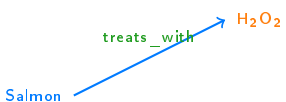


Treatment recommendation via KG reasoning

TransE: Translational Embedding

Definition: $h + r \approx t$

- Relationship r is a translation vector from head h to tail t
- **Strength:** Excellent for hierarchical taxonomies
- **Limitation:** A single vector cannot point to multiple locations

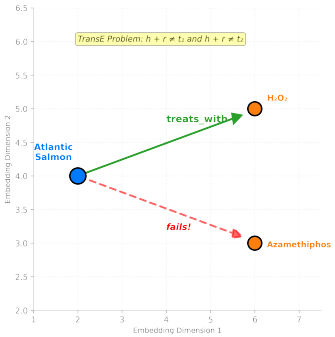


TransE: Failure on 1-to-N Relations

The Logic Trap:

- $(Salmon, treats, H_2O_2)$
- $(Salmon, treats, Azamethiphos)$
- If $h + r = t_1$ and $h + r = t_2$
- **Consequence:** $t_1 \approx t_2$

The model incorrectly treats different chemicals as identical!

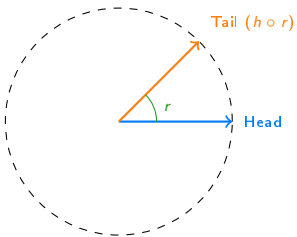


TransE: One treatment per salmon species?

RotatE: Knowledge as Rotation

Definition: $t = h \circ r$

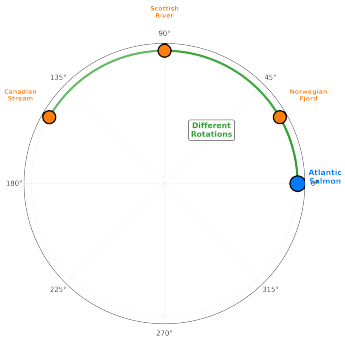
- Map entities into complex space $\mathbb{C}^d$
- Relations are rotations on the unit circle
- **Capability:** Can model complex symmetry and inversion



RotatE: Handling 1-to-N in Aquaculture

Solving the Overlap:

- Different relations rotate to different points
- $(Salmon, migrates\_to, River\_A)$
- $(Salmon, migrates\_to, River\_B)$
- **Advantage:** Multiple tails can exist on the complex rim



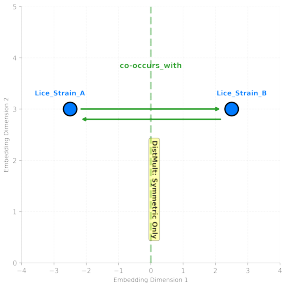
RotatE: Rotation in complex space

DistMult: High Speed, Low Detail

Definition: $h^T \text{diag}(r) t$

- Matrix factorization approach
- **Strength:** Extremely efficient
- **Flaw:** Intrinsically symmetric

(A, r, B) is the same as (B, r, A)



DistMult: Mirror effect

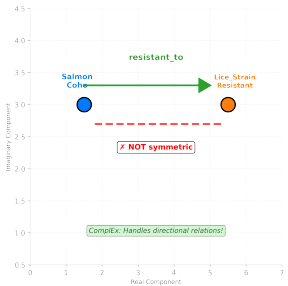
ComplEx: The Winner for Directional Data

Why it works:

- Extends DistMult to complex space
- Hermetian dot product allows asymmetry
- **Success:** Correctly models $A \rightarrow B \neq B \rightarrow A$

Example:

- $(Salmon, resistant\_to, Lice\_A)$
- Lice is **not** resistant to Salmon!



ComplEx: Captures flow and direction

The Problem: What is ID 42?

Traditional PyKeen Training:

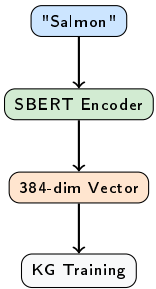
- Entities: "Atlantic_Salmon" $\rightarrow$ ID 42
- Relations: "treats_with" $\rightarrow$ ID 7
- **Loss:** The model knows nothing about biology at $t = 0$

"A Salmon is just more similar to a Trout than to a Chemical."

The Solution: SBERT Pre-training

Semantic Seeding:

- Convert labels to 384-dim vectors
- Initialize embeddings with these vectors
- **Result:** Faster convergence, better zero-shot performance



Evaluation: The Link Prediction Task

- **Objective:** Predict missing treatments for specific farm conditions
- **Dataset:** Salmon-Lice-Treatment KG (30k+ triples)
- **Key Metrics:**
  - **MRR:** Mean Reciprocal Rank (How high is the correct treatment ranked?)
  - **Hits@10:** Is the correct treatment in the Top 10 recommendations?

Testing the models against real-world asymmetric constraints

Results: The Model Leaderboard

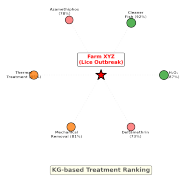
Model	Hits@5 (↑)	Hits@10 (↑)	MRR (↑)	MR (↓)
SSL-SSM (Ours)	0.3380	0.4000	0.2636	597.8
TransE	0.3040	0.3780	0.2169	285.8
RotatE	**0.4160**	**0.4800**	**0.3078**	**259.2**
DistMult	0.2780	0.3720	0.2075	**223.9**
AutoSF	0.0140	0.0220	0.0128	4353.6
ComplEx	0.0100	0.0200	0.0127	4073.0
Transformer	0.0820	0.1380	0.0544	1570.7

Analysis:

- **SSL-SSM** outperforms translational and bilinear baselines in Hits metrics.
- **RotatE** shows strong performance in complex space modeling for this dataset.
- **AutoSF/ComplEx** struggle with current initialization, highlighting the need for SSL.

Future Work: From Lab to Farm

- **Temporal Dynamics:** Tracking efficacy over seasons
- **Sensory Data:** Integrating water quality and sensor logs
- **Deployment:** Real-time recommendation app for farmers



Moving towards autonomous aquaculture management

Questions?

Leandro Stival
Wageningen University & Research

Thank you for supporting sustainable aquaculture!