**PHẦN LÝ THUYẾT CHI TIẾT CÁC THUẬT TOÁN TRONG BÀI TOÁN TÌM KIẾM ẢNH DỰA TRÊN ĐỒ THỊ TRI THỨC**

**1. Thuật toán TransE (Translating Embeddings for Modeling Multi-relational Data)**

**1.1. Khái niệm**

TransE là một phương pháp embedding đơn giản nhưng hiệu quả, được đề xuất bởi Bordes et al. (2013) nhằm ánh xạ các thực thể (entities) và quan hệ (relations) trong đồ thị tri thức vào cùng một không gian vector Euclid. Ý tưởng chính của TransE là mô hình hóa mối quan hệ như một phép tịnh tiến trong không gian vector: với mỗi bộ ba (head, relation, tail), nếu quan hệ giữa head và tail là hợp lệ thì:

h+r≈t\mathbf{h} + \mathbf{r} \approx \mathbf{t}

Trong đó:

* h,t∈Rk\mathbf{h}, \mathbf{t} \in \mathbb{R}^k: vector embedding của thực thể head và tail.
* r∈Rk\mathbf{r} \in \mathbb{R}^k: vector embedding của quan hệ.

**1.2. Hàm mất mát**

Hàm mất mát chuẩn của TransE dựa trên Margin Ranking Loss, với mục tiêu tối thiểu hóa khoảng cách giữa h+r\mathbf{h} + \mathbf{r} và t\mathbf{t} trong khi tối đa hóa khoảng cách với các bộ ba sai lệch (negative samples):

L=∑(h,r,t)∈T∑(h′,r,t′)∈T′[γ+d(h+r,t)−d(h′+r,t′)]+\mathcal{L} = \sum\_{(h, r, t) \in \mathcal{T}} \sum\_{(h', r, t') \in \mathcal{T}'} \left[ \gamma + d(h + r, t) - d(h' + r, t') \right]\_+

Trong đó:

* γ\gamma: margin (thường là hằng số như 1.0)
* d(⋅)d(\cdot): khoảng cách, thường dùng L1 hoặc L2-norm.
* [z]+=max⁡(0,z)[z]\_+ = \max(0, z)

**1.3. Ưu điểm và hạn chế**

* Ưu điểm: đơn giản, tốc độ huấn luyện nhanh.
* Hạn chế: không biểu diễn tốt các quan hệ 1-N, N-1, hoặc N-N do sử dụng cùng một không gian embedding cho cả thực thể và quan hệ.

**2. Thuật toán TransR (Learning Entity and Relation Embeddings for Knowledge Graph Completion)**

**2.1. Khái niệm**

TransR (Lin et al., 2015) mở rộng TransE bằng cách ánh xạ thực thể và quan hệ vào hai không gian vector riêng biệt. Cụ thể, mỗi quan hệ có một ma trận chiếu riêng Mr∈Rdr×deM\_r \in \mathbb{R}^{d\_r \times d\_e} để chiếu các thực thể sang không gian của quan hệ.

hr=Mr⋅h,tr=Mr⋅t\mathbf{h}\_r = M\_r \cdot \mathbf{h}, \quad \mathbf{t}\_r = M\_r \cdot \mathbf{t}

Sau đó thực hiện phép tịnh tiến như TransE trong không gian quan hệ:

hr+r≈tr\mathbf{h}\_r + \mathbf{r} \approx \mathbf{t}\_r

**2.2. Hàm mất mát**

L=∑[γ+∥Mr⋅h+r−Mr⋅t∥−∥Mr⋅h′+r−Mr⋅t′∥]+\mathcal{L} = \sum \left[ \gamma + \| M\_r \cdot h + r - M\_r \cdot t \| - \| M\_r \cdot h' + r - M\_r \cdot t' \| \right]\_+

**2.3. Ưu điểm và hạn chế**

* Ưu điểm: xử lý tốt hơn các quan hệ phức tạp so với TransE.
* Hạn chế: tốn nhiều bộ nhớ và thời gian tính toán hơn do sử dụng ma trận chiếu riêng cho từng quan hệ.

**3. R-GCN (Relational Graph Convolutional Network)**

**3.1. Khái niệm**

R-GCN là một biến thể của GCN được thiết kế đặc biệt cho đồ thị tri thức có nhiều loại quan hệ. Trong R-GCN, mỗi loại quan hệ có một ma trận chuyển riêng, và thông tin được lan truyền giữa các node thông qua các quan hệ này.

**3.2. Công thức lan truyền**

hi(l+1)=σ(∑r∈R∑j∈Nir1ci,rWr(l)hj(l)+W0(l)hi(l))\mathbf{h}\_i^{(l+1)} = \sigma \left( \sum\_{r \in \mathcal{R}} \sum\_{j \in \mathcal{N}\_i^r} \frac{1}{c\_{i,r}} \mathbf{W}\_r^{(l)} \mathbf{h}\_j^{(l)} + \mathbf{W}\_0^{(l)} \mathbf{h}\_i^{(l)} \right)

Trong đó:

* Nir\mathcal{N}\_i^r: tập các nút láng giềng của nút i với quan hệ r.
* Wr(l)\mathbf{W}\_r^{(l)}: ma trận chuyển riêng cho quan hệ r ở lớp l.
* W0(l)\mathbf{W}\_0^{(l)}: ma trận tự chuyển (self-loop).
* σ\sigma: hàm kích hoạt (thường là ReLU).

**3.3. Ưu điểm và hạn chế**

* Ưu điểm: khai thác tốt cấu trúc liên kết và quan hệ trong đồ thị.
* Hạn chế: số lượng ma trận chuyển tăng tuyến tính theo số quan hệ → dễ quá tải với KG lớn.

**4. GCN with BoVW + SIFT**

**4.1. Khái niệm**

Mô hình này áp dụng GCN vào đồ thị similarity giữa các ảnh, trong đó:

* Mỗi node là một ảnh
* Đặc trưng ban đầu là vector histogram từ BoVW (Bag of Visual Words), được trích xuất qua SIFT + KMeans
* Các node được nối nếu ảnh tương đồng cao (cosine similarity)

**4.2. Phép tính BoVW**

* SIFT trích xuất các keypoints di∈R128\mathbf{d}\_i \in \mathbb{R}^{128} từ mỗi ảnh
* Dùng KMeans để gộp các keypoints vào K cluster → "visual words"
* Mỗi ảnh được biểu diễn bởi histogram vector xi∈RK\mathbf{x}\_i \in \mathbb{R}^K

**4.3. Phép tính lan truyền GCN**

H(l+1)=σ(D^−1/2A^D^−1/2H(l)W(l))\mathbf{H}^{(l+1)} = \sigma \left( \hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)} \right)

Trong đó:

* A^=A+I\hat{A} = A + I: ma trận kề cộng thêm self-loop
* D^\hat{D}: ma trận bậc của A^\hat{A}
* H(0)=X\mathbf{H}^{(0)} = X: BoVW input feature

**5. Faster-RCNN (tiền xử lý tạo triplet từ ảnh)**

**5.1. Khái niệm**

Faster-RCNN là mô hình two-stage object detection mạnh mẽ:

* Giai đoạn 1: RPN sinh vùng nghi vấn (proposal)
* Giai đoạn 2: ROI pooling và classification từng proposal

**5.2. Các phép tính**

* Tính toán đặc trưng ảnh bằng backbone CNN → feature map F∈RC×H×WF \in \mathbb{R}^{C \times H \times W}
* ROI pooling trên FF với bounding box (x,y,w,h)(x, y, w, h)
* Fully connected → phân loại đối tượng và tọa độ

**5.3. Ứng dụng vào KG**

Kết quả phát hiện các đối tượng có thể sinh ra triplet kiểu:

* (person, on, horse)
* (dog, next to, chair)

Dữ liệu đầu vào cho KG embedding hoặc reasoning.

Tài liệu này cung cấp nền tảng lý thuyết chi tiết và chuẩn học thuật để triển khai các thuật toán embedding và reasoning trong bài toán tìm kiếm ảnh dựa trên đồ thị tri thức. Có thể sử dụng cho luận văn đại học, seminar chuyên ngành hoặc mở rộng nghiên cứu sau đại học.