**Phương pháp đánh giá một mô hình word embedding**

Chia làm hai loại:

* **Đánh giá nội tại (Intrinsic Evaluation):** Đánh giá chất lượng biểu diễn vector dựa trên các nhiệm vụ không phụ thuộc vào tác vụ NLP cụ thể (như phân tích độ tương đồng từ).
* **Đánh giá ngoại tại (Extrinsic Evaluation):** Sử dụng nhúng từ làm đầu vào cho các tác vụ NLP (như phân tích cú pháp, nhận diện thực thể).

**1. Đánh giá nội tại (Intrinsic Evaluation)**

**Đặc điểm:**

* Đánh giá độc lập với các tác vụ NLP cụ thể.
* Tập trung vào việc đo lường các đặc tính bên trong của biểu diễn vector như độ tương đồng ngữ nghĩa hoặc cú pháp.
* Nhanh chóng và ít tốn tài nguyên tính toán hơn so với đánh giá ngoại tại.
* Chủ yếu được thực hiện trên các tập dữ liệu chuẩn, nơi các cặp từ hoặc nhóm từ đã được gắn nhãn thủ công.

**Tiêu chí phổ biến:**

1. **Độ tương đồng từ (Word Similarity):**
   * Đo lường mức độ tương đồng giữa các vector của cặp từ, thường sử dụng **cosine similarity**.
   * Ví dụ: "car" và "automobile" có độ tương đồng cao hơn so với "car" và "banana".
   * Datasets: WS-353, SimLex-999, Rare-Word (RW).
2. **Tương tự từ (Word Analogy):**
   * Kiểm tra xem mô hình có thể tìm ra quan hệ giữa các cặp từ không.
   * Ví dụ: *king - man + woman = queen*.
   * Datasets: Google Analogy, MSR Analogy.
3. **Phân loại khái niệm (Concept Categorization):**
   * Yêu cầu mô hình nhóm từ vào các danh mục khái niệm.
   * Ví dụ: phân nhóm "dog, cat" vào động vật và "banana, apple" vào thực phẩm.
4. **Phát hiện từ lạc lõng (Outlier Detection):**
   * Tìm từ không thuộc nhóm từ đã cho.
   * Ví dụ: Trong nhóm "dog, cat, banana", "banana" là từ lạc lõng.
5. **QVEC:**
   * Đánh giá độ tương quan giữa các vector nhúng từ và các vector ngôn ngữ học do con người tạo ra.

**Ưu điểm:**

* Dễ dàng triển khai và nhanh chóng.
* Có thể sử dụng để so sánh nhiều mô hình khác nhau một cách độc lập.

**Nhược điểm:**

* Không phải lúc nào cũng phản ánh đúng hiệu suất của mô hình trong các tác vụ thực tế.
* Phụ thuộc nhiều vào tập dữ liệu đánh giá (có thể thiên lệch).

**2. Đánh giá ngoại tại (Extrinsic Evaluation)**

**Đặc điểm:**

* Đo lường hiệu suất của mô hình nhúng từ khi được sử dụng trong các tác vụ NLP thực tế.
* Mô hình word embedding được dùng làm đầu vào cho các hệ thống xử lý như phân tích cú pháp, nhận diện thực thể, hoặc dịch máy.
* Yêu cầu tài nguyên tính toán lớn hơn và phức tạp hơn.

**Tiêu chí phổ biến:**

1. **Gán nhãn từ loại (Part-of-Speech Tagging - POS):**
   * Xác định chức năng ngữ pháp của từ (danh từ, động từ, tính từ,...).
   * Ví dụ: "The cat is sleeping" → ["The/DT", "cat/NN", "is/VB", "sleeping/VBG"].
2. **Chunking:**
   * Gán nhãn các cụm từ ngữ pháp, ví dụ như cụm danh từ, cụm động từ.
   * Ví dụ: "The cat is sleeping" → ["The cat/NP", "is sleeping/VP"].
3. **Nhận diện thực thể (Named-Entity Recognition - NER):**
   * Xác định các thực thể (người, địa điểm, tổ chức) trong văn bản.
   * Ví dụ: "Barack Obama was born in Hawaii" → ["Barack Obama/PER", "Hawaii/LOC"].
4. **Phân tích cảm xúc (Sentiment Analysis):**
   * Phân loại cảm xúc của văn bản (tích cực, tiêu cực, hoặc trung tính).
   * Ví dụ: Đánh giá bình luận phim trên IMDb.
5. **Dịch máy (Neural Machine Translation - NMT):**
   * Sử dụng nhúng từ làm đầu vào để dịch văn bản từ ngôn ngữ nguồn sang ngôn ngữ đích.
   * Ví dụ: "Hello, how are you?" → "Bonjour, comment ça va ?".

**Ưu điểm:**

* Cung cấp kết quả thực tế và đáng tin cậy hơn, vì nó phản ánh hiệu suất của mô hình trong các ứng dụng thực tế.
* Dễ so sánh các mô hình dựa trên hiệu suất cuối cùng.

**Nhược điểm:**

* Tốn nhiều thời gian và tài nguyên để thực hiện.
* Hiệu suất phụ thuộc vào chất lượng của các thành phần khác trong hệ thống (như mô hình dự đoán, dữ liệu huấn luyện).

**\*Nếu kết hợp một mô hình xử lý ngôn ngữ thuần, không chuyên dùng cho cả hình ảnh. Thì quy trình xử lý phức tạp hơn:**

* **Kết hợp Word2Vec/BERT với mô hình hình ảnh:**
  + Sử dụng Word2Vec hoặc BERT để tạo embedding mô tả, sau đó huấn luyện thêm mô hình ánh xạ giữa các không gian vector văn bản và hình ảnh (VD: dùng mạng nơ-ron fully connected).
* **Fine-tuning CLIP:** Với tập dữ liệu mô tả-hình ảnh đặc thù, bạn có thể fine-tune CLIP để đạt hiệu suất cao hơn.