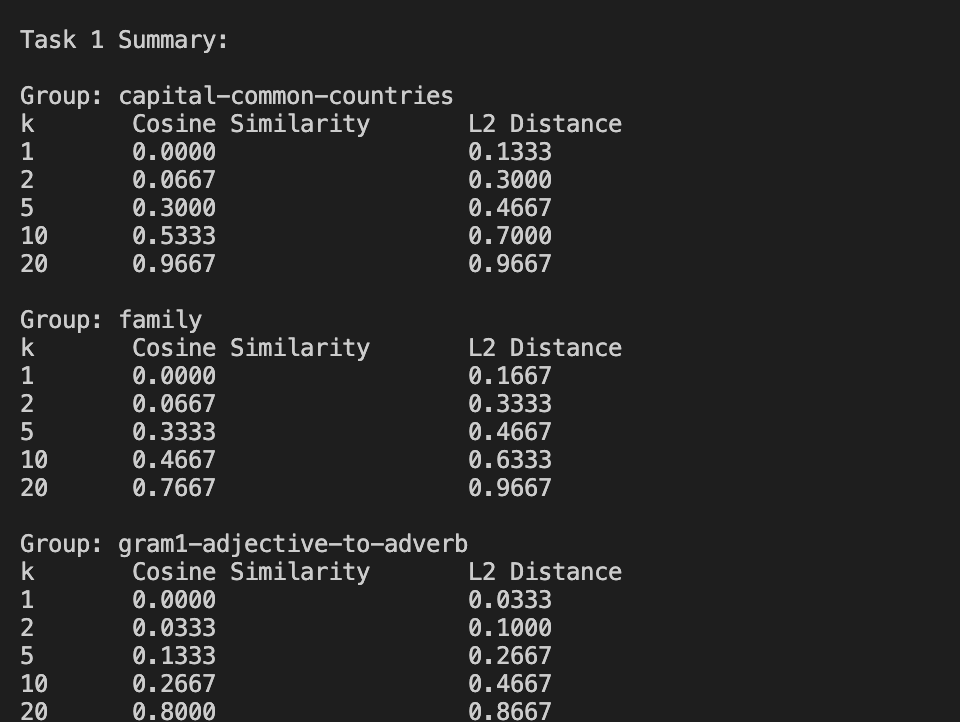
Contextual Embeddings for Word Analogies: Analysis Report

**Pre-Experiment Hypotheses**

We anticipated that BERT's contextual embeddings would underperform traditional static embeddings on pure word analogy tasks, given that models like Word2Vec were specifically designed for vector arithmetic. We hypothesized that the choice between cosine similarity and L2 distance might reveal fundamental differences in how BERT encodes relational information. Additionally, we expected varying performance across semantic relationships (capitals:countries), familial relationships, and morphological relationships (adjective:adverb).

**Key Findings**

****

Our experiment revealed that L2 distance consistently outperformed cosine similarity across all categories, with Top-1 accuracy of 3.3-16.7% for L2 versus 0% for cosine, Top-5 accuracy of 26.7-46.7% for L2 versus 13.3-33.3% for cosine, and Top-10 accuracy of 46.7-70.0% for L2 versus 26.7-53.3% for cosine. This suggests that in BERT's embedding space, both direction and magnitude of relationship vectors are important for analogical reasoning. When examining performance across categories at Top-5 accuracy, capital-country relationships showed the highest performance (L2: 46.7%, Cosine: 30.0%), family relations performed similarly (L2: 46.7%, Cosine: 33.3%), while adjective-to-adverb transformations demonstrated the lowest performance (L2: 26.7%, Cosine: 13.3%). This indicates that BERT better captures semantic relationships than morphological ones, reflecting its training objective of predicting contextual words rather than modeling explicit linguistic structure. Our k-value analysis revealed poor Top-1 accuracy (0-16.7%) but good Top-20 accuracy (87-97%), with the largest accuracy gains occurring between k=5 and k=10, demonstrating a non-linear improvement pattern.

**Theoretical Implications**

The word analogy vector arithmetic approach (a-b ≈ c-d) applies to BERT's embeddings but requires preserving magnitude information through L2 distance rather than normalizing with cosine similarity. BERT encodes analogical relationships implicitly rather than explicitly, explaining why correct answers typically appear in the broader neighborhood (k>5) rather than as the closest match. The model's contextual training produces stronger representations for semantic relationships than for syntactic or morphological ones, which aligns with its pre-training objective. These findings highlight the fundamental differences between how contextual and static embedding models organize relational information, offering insights into both the limitations and unique properties of language models for analogical reasoning tasks. The superior performance of L2 distance across all categories reinforces the importance of considering both directional and magnitude information when working with contextual embeddings for relational tasks.