

確率除去を組み込んだ検索拡張された言語モデルによる帰納的關係予測

兼岩研究室 岩本有生

知識グラフの補完

知識グラフ：知識の体系的なつながりをグラフとして表現

トリプル：各データは（主語，述語，目的語）で表現

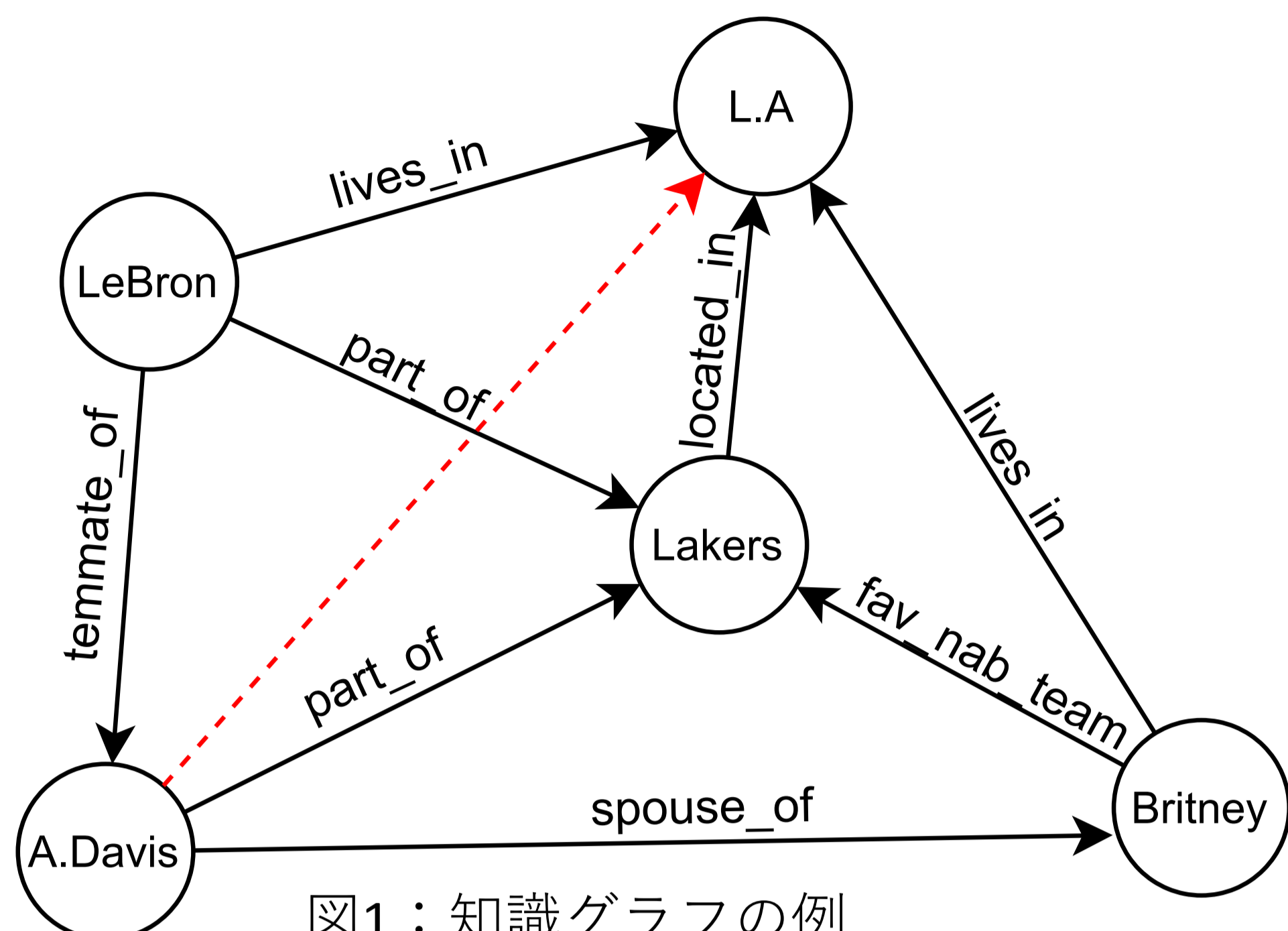


図1：知識グラフの例

知識グラフ補完：知識グラフの不足した情報（赤矢印）を予測し補完

先行研究

GRAILやBERTRL

あるトリプルの予測を行う際に、入力近傍の部分グラフを入力とする（ロジカルルールを学習）

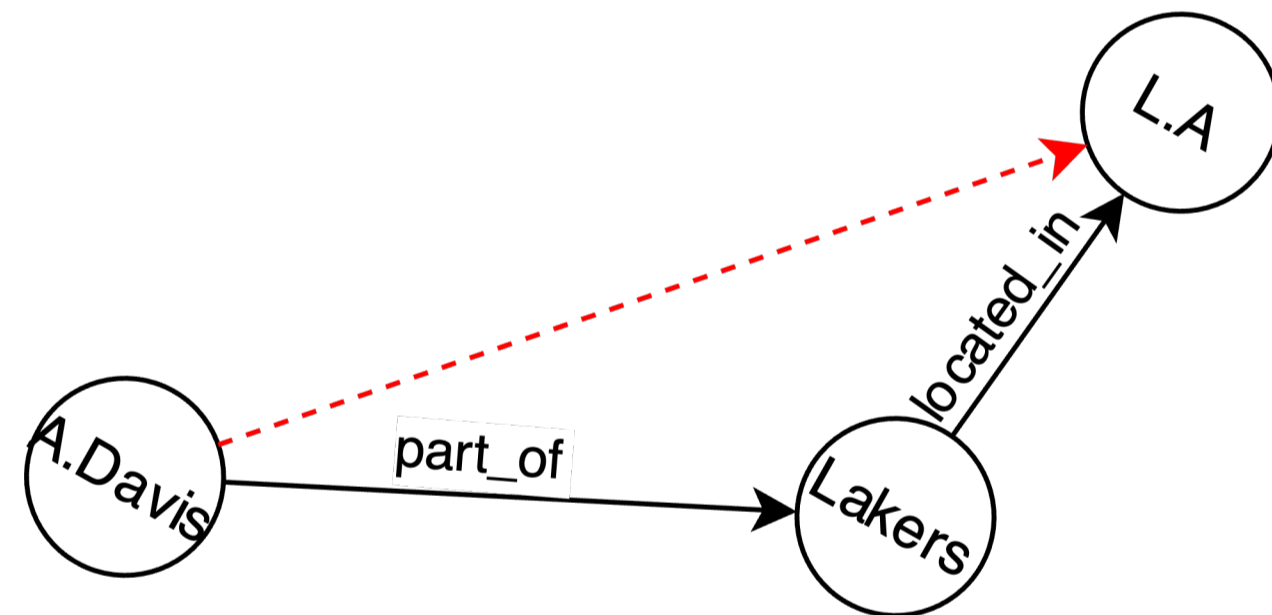
ロジカルルール

トリプルの予測に有効な、論理的なルール
未知のエンティティ（ノード）に対して一般化可能

$$(x, \text{part_of}, y) \wedge (y, \text{located_in}, z) \rightarrow (x, \text{lives_in}, z)$$

欠点

- 知識グラフのサイズに依存した計算コスト
- 大規模な知識グラフに適用できない



提案手法

問題点

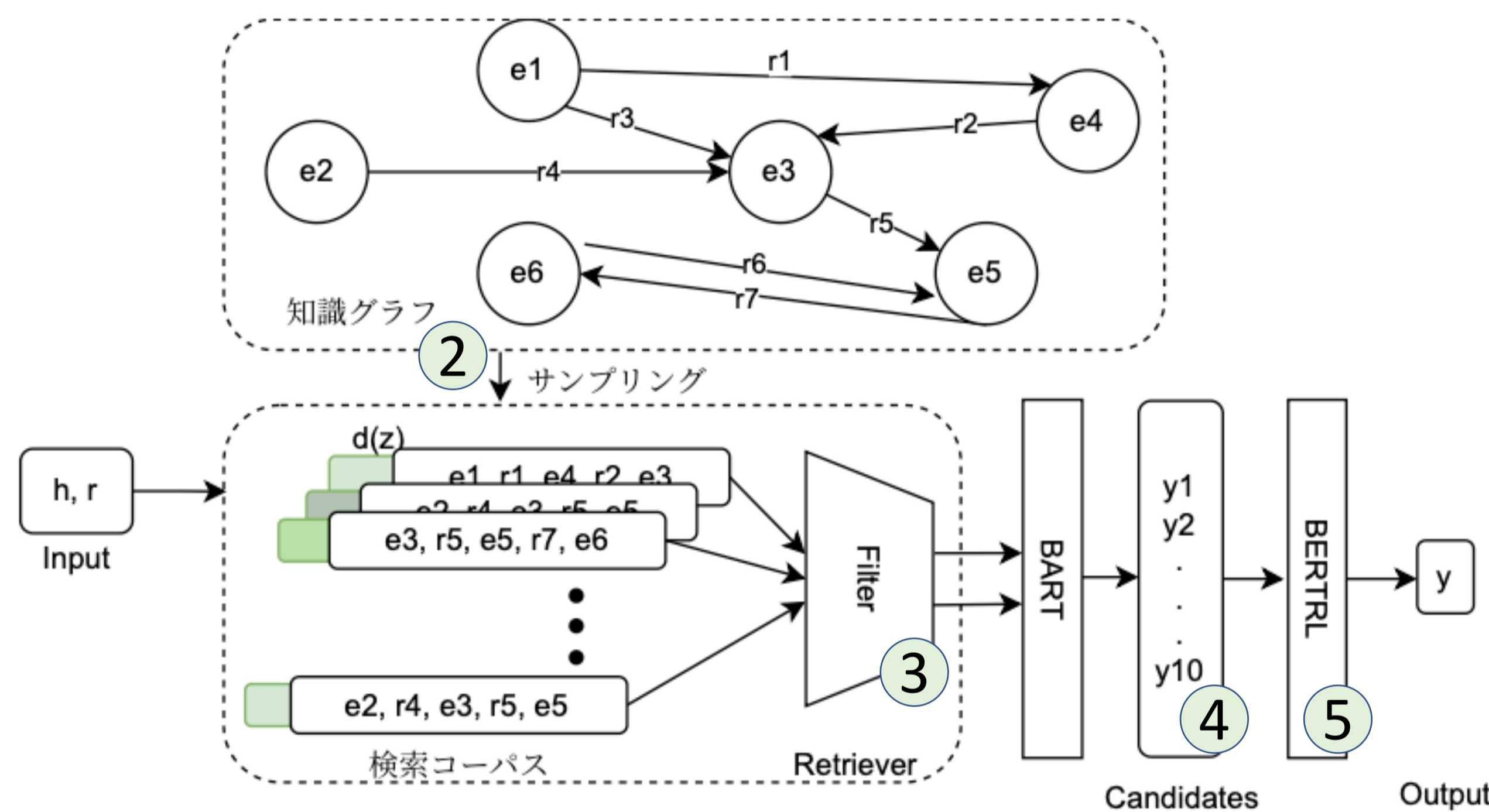
- 部分グラフを考慮すると計算量が知識グラフのサイズに比例して増加
- 各トリプルのスコアを計算すると計算量が知識グラフのエンティティ数に比例して増加

目的

- 部分グラフを考慮しつつ、知識グラフのサイズに依存しない高速な手法の提案
- 精度の向上

提案手法の概要

- 検索拡張された言語モデル（RAG）を用いる
- 検索コーパスを知識グラフより構築
- 検索結果の確率的除去，RAGの精度向上
- RAGにより解候補を生成
- 再ランキングによって，さらなる精度向上（ある条件下でランキングするモデルよりも精度が向上する）

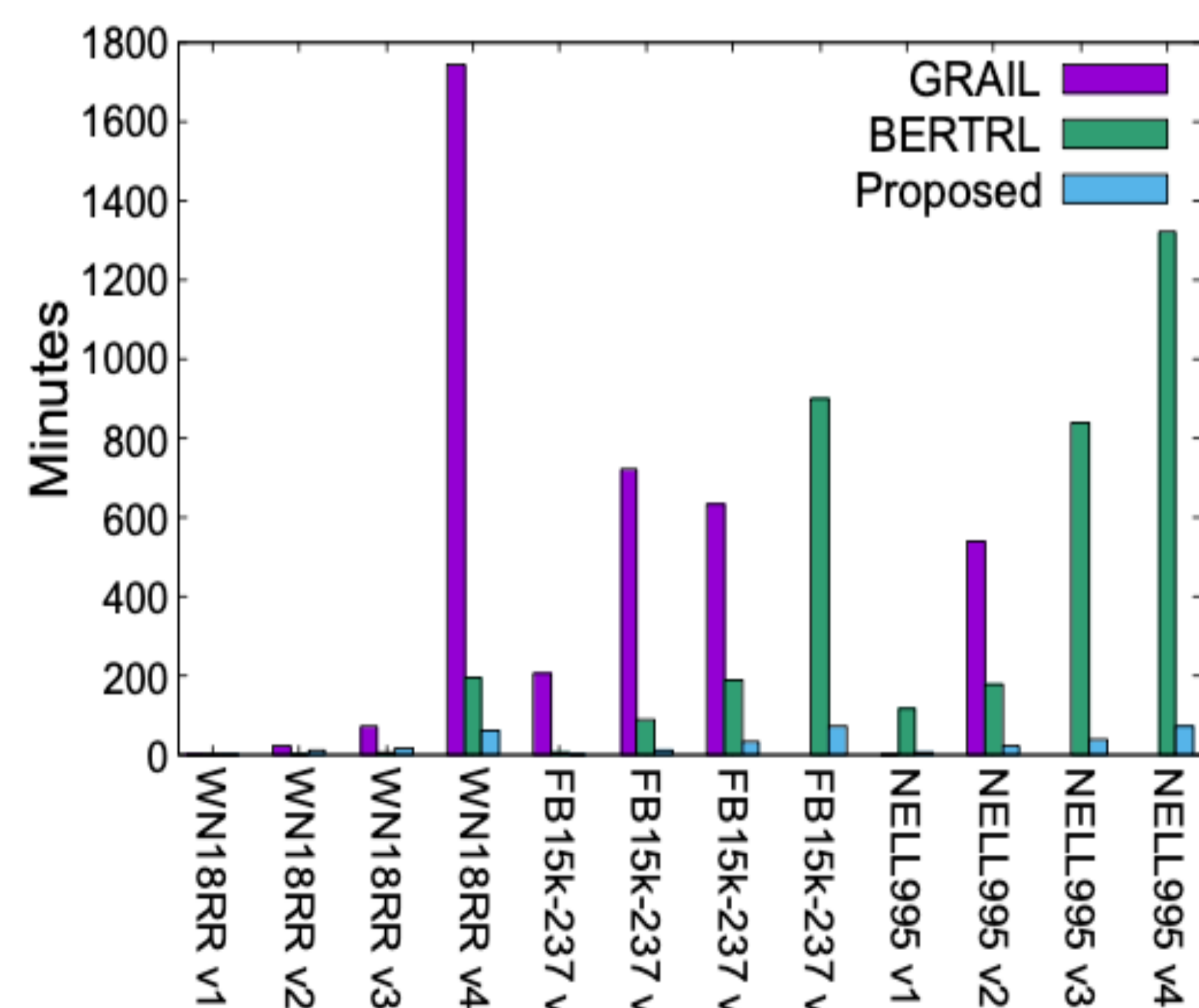


図：提案手法の概要図

実験

表：Hits@1の数値（Hits@1：正解を1位で予測できた割合）

Method	WN18RR		FB15k-237		NELL-995	
	v3	v4	v3	v4	v3	v4
GRAIL	0.23	0.479	0.151	-	0.097	-
BERTRL	0.349	0.484	0.26	0.202	0.269	0.109
CoMPLE	0.255	0.480	-	-	0.210	-
MorsE	0.201	0.252	0.134	0.114	0.038	0.014
KRST	0.508	-	-	-	-	-
GenKGC	0.086	0.048	0.108	0.126	0.151	0.063
(Proposed 1hop)	0.289	0.374	0.305	0.265	0.272	0.156
(Proposed 3hop)	0.149	0.125	0.184	0.178	0.220	0.073
(Proposed tails)	0.264	0.265	0.197	0.195	0.231	0.068
(Proposed relations)	0.053	0.052	0.044	0.044	0.053	0.0082



図：テストデータ評価時の実行時間

まとめと今後の課題

- 精度を保ちつつ高速な手法の提案
 - WN18RR以外で精度を更新
 - 推論速度を改善
- 検索精度の向上による，関係予測の精度向上
- 生成モデルと再ランキング用モデルの連携または統合（知識蒸溜）