人工知能学会第二種研究会資料
汎用人工知能研究会
SIG-AGI-031-4

競合文脈を用いたプロトタイプ拡張による少数ショットテキスト含意

Prototype Augmented Few-Shot Textual Entailment with Competing Contexts

田梓岐 * 兼岩憲
Zuchi Den Ken Kaneiwa

電気通信大学 情報理工学域 コンピュータサイエンスプログラム
Computer Science Program, School of Informatics and Engineering, The University of Electro-Communications

Abstract: Prototype networks are a standard few-shot learning paradigm that construct class representatives from support examples and classify queries by similarity to those prototypes. In the field of Textual Entailment (TE), many few-shot methods have combined Large Language Models (LLMs) with prototypical ideas to adapt quickly with limited labeled data. However, existing few-shot TE approaches often fail to exploit relative information among texts. In realistic multiple-choice QA settings a single premise is paired with several competing hypotheses which are also important for making the correct judgment. In this paper we propose UFO-CC that explicitly incorporates competitive context and robust prototype weighting to mitigate these problems. UFO-CC improves stability and discriminative power through three stages: (i) append competing contexts to each TE pair so the encoder learns relative differences via cross-attention; (ii) reweight supports by similarity and a Context Gates-derived consistency score to build stable prototypes; (iii) linearly fuse prototypes and average multiple head logits to reduce noise and stabilize predictions. In the experiments, we show that our UFO-CC outperforms the UFO-ENTAIL models in average accuracy on benchmark QA datasets.

1 はじめに

テキスト含意（Textual Entailment, TE）[1] は自然言語文の意味関係を判定するタスクである．このタスクは問われる前提参照文と仮説の組に対して後者が前者から導かれるか否かを識別する．そのようなテキスト含意は質問応答や情報抽出など自然言語理解の多くの下位課題と密接に関連しており，特に選択式質問応答（Question Answering, QA）タスク においては，多数の候補文を前提と照合して正答を選ぶプロセスがテキスト含意的判定の繰り返しと見なせる．

一方で，伝統的な QA システムは大規模なラベル付けデータが不可欠であるが，特にドメインが変化する場面や新たなタスクに対しては大量のラベルを付与することが難しい．そのため，テキスト含意のタスクに適応可能な汎化表現が得られる少数ショット学習（few-shot learning）の正確性を向上させる研究は，ラベリングコストの制約がある実用システムにとって非常に重要である．

テキスト含意に対して今まで少数ショットについて数多くの研究がなされてきた．プロトタイプ [2] に基づく手法は，少数ショット学習における標準的な枠組みの一つである．Prototypical Networks はサポート例の平均でクラスごとの代表ベクトルを構築する有効な枠組みを提供する一方で，固定的なプロトタイプはクエリ依存性や候補間の競合情報を扱いにくいという限界がある．Attentive Prototype[3] は，サポート例に対して注意力や重み付けを導入することで，より重要なサポート例を優先することで性能を改善してきた．

さらに、近年自然言語処理の分野では，RoBERTa[4] などの事前学習済み大規模言語モデル（Large Language Models，LLMs）[5] を利用した少数ショットテキスト含意の研究が注目を集めている．UFO-ENTAIL[6] は，テキスト含意を少数ショット学習の統一枠組みとして位置づけ，LLMs とプロトタイプを組み合わせて，少数

*連絡先： 電気通信大学 情報理工学域 コンピュータサイエンスプログラム
〒 182-8585 東京都調布市調布ヶ丘 1-5-1
E-mail: den@sw.cei.uec.ac.jp

ラベルで下流タスクに適応する実践的な枠組みを提案している．Context-TE[7] は，他候補を評価対象の文脈として付与して判定の情報量を増やす手法であり，特に類似性の高い誤答が混入する場合に有効である．ただし，単純な候補の連結や均等な平均集約はノイズ混入を招きやすく，競合文脈を如何に選別するかが課題である．

これらの少数ショット手法は，様々なテキスト含意タスクに有効であるが，実用的な QA 設定では単一の前提に対して複数の候補仮説が存在し，これらは互いに競合的でかつ類似性の高い情報を共有する．このとき二つの問題が生じる．第一に，平均化に基づくプロトタイプは競合候補由来のノイズで代表表現が希釈されやすく，特に少数ショットでは有益な信号が埋もれてしまう．第二に，複数候補が語彙や表層表現を共有する場合，単純な類似度に基づく評価は語彙的に近い誤答を高く評価してしまう．

たとえば，同一の前提文に対して「人物 A が行動を起こした理由」を問う正答と，「人物 A が行動を起こした結果」や「他者の意図」を述べる誤答が，表層的な語彙は類似していても，意味的な方向性が異なるため，単純平均では誤った中心に引き寄せられる．また，「because」「so that」といった因果接続詞を含む複数の選択肢は，いずれも「理由説明文」として類似度が高く計算される一方，実際に含意関係を持つのはごく一部である．これらの要因が重なると，モデルの精度低下に加えて，実際の QA 応用での信頼性が低下する．

本研究では，UFO-ENTAIL 型の少数ショット分類器に基づき，候補文脈間の相対関係をプロトタイプベースの少数ショット処理に取り込む統合的枠組み UFO-CC(UFO-Competitive Context) を提案する．まず，ベースとなる UFO-ENTAIL は大規模な自然言語推論データセットで RoBERTa を事前学習する．さらに，推論の安定性を向上させるために文脈重み付きプロトタイプが文脈をより適切に扱い，モデル内部の複数ヘッドがそれぞれ独自の判断を行う．最後に複数アテンションヘッドを統合することで，既存手法より高い学習精度を実現する．提案手法は図 1 に示す三つのステージから構成される．

1. 同じ前提文に対する他の選択肢を競合文脈として仮説に付け加えることで，エンコーダーにクエリを相対的に理解させる．これにより，Transformer[8] の相互注意機構を介して文脈間の相対差分を直接学習でき，類似選択肢間の微妙な差を捉えやすくなる．

2. クエリと各サポート例との意味的類似度および文脈的整合度をもとに，サポートごとの重要度を算出し，その重みに基づいてプロトタイプを構築する．

3. 二つのプロトタイプを線形融合し，複数のアテンションヘッドは各々が独立に判断を行い，それらの出力の平均により統合する．

本研究では，提案手法を長文読解 QA のデータセット MCTest500[9] で評価した．既存手法の UFO-ENTAIL と比較して，ベースラインより一貫して優れた性能を示し，詳細なアブレーションにより各構成要素の寄与を確認した．

2 関連研究

2.1 競合文脈拡張（Competitive Context）

テキスト含意タスクは，前提 P と仮説 H の関係（含意・矛盾・中立）を判定するタスクであり，文脈化ペア (P, H)[10] を入力としてエンコーダにより表現 $\mathbf{v}(P, H)$ を得る．しかし，実際の多選択肢 QA タスクでは，同一の前提に対して複数の候補仮説 $H_1, H_2, \ldots, H_n$ が存在し，これらの間には意味的な競合関係が生じる．Context-TE は特定の仮説 H_1 を評価する際に，従来の文脈化ペア (P, H) を拡張し，他の k 個の競合文脈 $c_1, \ldots, c_k$ を同一入力列としてモデルに付加することで，文脈の相対的な情報を同時に考慮する．このとき，従来の文脈化ペアを拡張して

$$(P,\ H,\ c_1,\ c_2,\ \ldots,\ c_k)$$

のような拡張文脈化ペアを形成する．これにより，モデルは Transformer の相互注意機構を介して文脈間の相対差分を直接学習でき，より判別的に特徴を抽出する．特に少数ショット設定下では，限られたデータから一般化性能を高める有効な文脈拡張手法と言える．

2.2 ヘッド投票（Head Voting）

ヘッド投票とは，Transformer モデル内部の複数アテンションヘッドがそれぞれ異なる言語的特徴を捉えて，独立したヘッド単位出力を投票する手法である．

NoVO[11] はヘッド投票の考え方を Transformer 構造に定式化した手法である．各アテンションヘッドを部分的に独立した識別器として扱い，それぞれが局所的な意味表現に基づいて独自の予測を出力し，最終的なクラスを投票によって決定する．特定ヘッドの局所的偏りの影響を平均化により軽減することを目的とする．

2.3 コンテキストゲート（Context Gates）

複数情報源の信頼度を統合する際，入力依存で重みを調整する注意制御機構であるコンテキストゲートが用

いられている．代表的には複数の画像情報を，ゲート機構によって動的に統合する Gated Multimodal Unit[12] や，入力や状況に応じて各要素の重みを動的に調整する Adaptive Weighting[13] がある．これらはいずれもシグモイド関数による平滑化を介して，情報の通過量を制御する．一般形として，2 つの入力 x_1, x_2 に対し，

$$g = \sigma(w^\top [x_1, x_2] + b)$$

で得られるゲート値 $g \in [0, 1]$ により，統合表現を

$$z = g \cdot x_1 + (1 - g) \cdot x_2$$

と定義する．この構造は，情報源ごとに最適な重みを決定できる点で有効である．

2.4 UFO-ENTAIL

UFO-ENTAIL は，Prototypical Networks に基づいて設計されている，ソース領域の大規模データ S とターゲット領域 T の少数ショット例を同時に利用する少数ショットテキスト含意学習の枠組みである．プロトタイプとはあるクラスの典型的な特徴を凝縮したベクトル表現を指し，クラス j に対して，サポート例集合 s_m^i を RoBERTa で符号化した埋め込みの平均によりプロトタイプ p^j を構築する．

$$p_t^j = \frac{1}{n} \sum_{i=1}^{n} \mathrm{RoBERTa}(s_m^i), \qquad t \in \{S, T\}$$

ここで n は各クラス j に対する少数ショットの個数を表す．特に，n 個のサポート例を用いてタスクに適応する少数ショット学習を n-shot と呼ぶ．クエリ q は推論対象となる文脈化ペア y_m をエンコーダに入力し，その出力表現 $q = \mathrm{RoBERTa}(y_m)$ を用いる．クエリ q と各クラスのプロトタイプとの類似度を計算し確率的なクラス分類を行う．入力特徴 I はクエリとプロトタイプ p を結合して構築する．

$$I = [p,\ q,\ p \circ q,\ p - q]$$

ここで $p \circ q$ は両者の局所的な一致情報を，$p - q$ は差分情報を表す．結合特徴 I を PrototypeNet に入力し，ソース領域 S の大規模汎用的な情報とターゲット領域 T のタスク固有の情報それぞれに対して，$\{\mathrm{Entailment}, \mathrm{Neutral}, \mathrm{Contradiction}\}$ の 3 クラスに対応するスコア $z = \mathrm{PrototypeNet}(I) \in \mathbb{R}^3$ を出力する．学習は S と T の損失を合算して最適化する．

UFO-ENTAIL は，エンコーディングを基盤とする標準的かつ拡張性の高い少数ショットテキスト含意モデルであるため，本研究ではベースラインとして採用する．ソース領域での処理手法は保持し，ターゲット領域における処理のみを拡張の対象とする．

3 UFO-CC

本章では，UFO-ENTAIL を基盤に競合文脈拡張プロトタイプと文脈重み付きプロトタイプを統合し，注意ヘッド統合で最終判定を行う少数ショットテキスト含意手法を提案する．

図 1 に示すように，まず前提と仮説から構成される文脈化ペアに，前提と複数候補から競合文脈を抽出し，候補間の相互作用を取り込んだ拡張文脈化ペアを追加する．次に，クエリ依存の類似度と信頼度ゲートでサポート例を再重み付けした両者の文脈重み付きプロトタイプを構築する．これら二種類のプロトタイプを線形に融合し，最後に複数の注意ヘッドからのスコアを平均して最終解を決定する．

3.1 競合文脈による拡張文脈化ペア

従来のテキスト含意タスクで用いられる前提文と仮説からなる文脈化ペア $s_i = (P, H)$ のサポート例を競合文脈で拡張し，サポート間の相対的情報を補う競合文脈拡張による文脈化ペアを導入する．

各 P に対して複数の候補仮説が存在する場合，自身の文脈集合から優先して合計 k 個の競合文脈を $c_1, \ldots, c_k$ 選ぶ．ここで，もし要求する競合文脈数 k が文脈集合の要素数 $|P|$ を超える場合（$k > |P|$），不足分は少数ショットで用いられていない前提の候補群からランダムに抽出して補填する．最後に，抽出した k 個の競合文脈を文脈化ペアに [SEP] 区切り連結して拡張文脈化ペア $s_i^{\mathrm{aug}} = (P, H, c_1,\ c_2,\ \ldots,\ c_k)$ を構築する．

3.2 文脈重み付きプロトタイプ

少数ショット設定では，各クラスのサポート数が極めて少なく，単純平均によるプロトタイプではサポート間の整合度差が反映されない．近年のコンテキストゲート研究では，入力ごとに動的なゲート値を算出し，動的に制御する方式が採用されてきた．しかし，各クラスあたりのサポート数が少ないとき，学習可能なゲートパラメータが十分に安定せず，タスク間で再現性が低下する．そこで本研究では，クエリとサポート例の意味的類似度および文脈的整合度の双方に基づき，静的に重みを付与する文脈重み付きプロトタイプを構築する．

まず，Prototypical Networks に倣ってクエリ q と文脈化ペアと拡張文脈化ペアのそれぞれのサポート例 $s_i^* \in \{s_i, s_i^{\mathrm{aug}}\}$ の意味的類似度をコサイン類似度で求める．次に，サポート例がクエリに近いほど指数的に大きくなるソフトマックス型の重み付け

$$\mathrm{sim}(q, s_i^*) = \exp\left(\frac{q \cdot s_i^*}{\tau |q| |s_i^*|}\right)$$

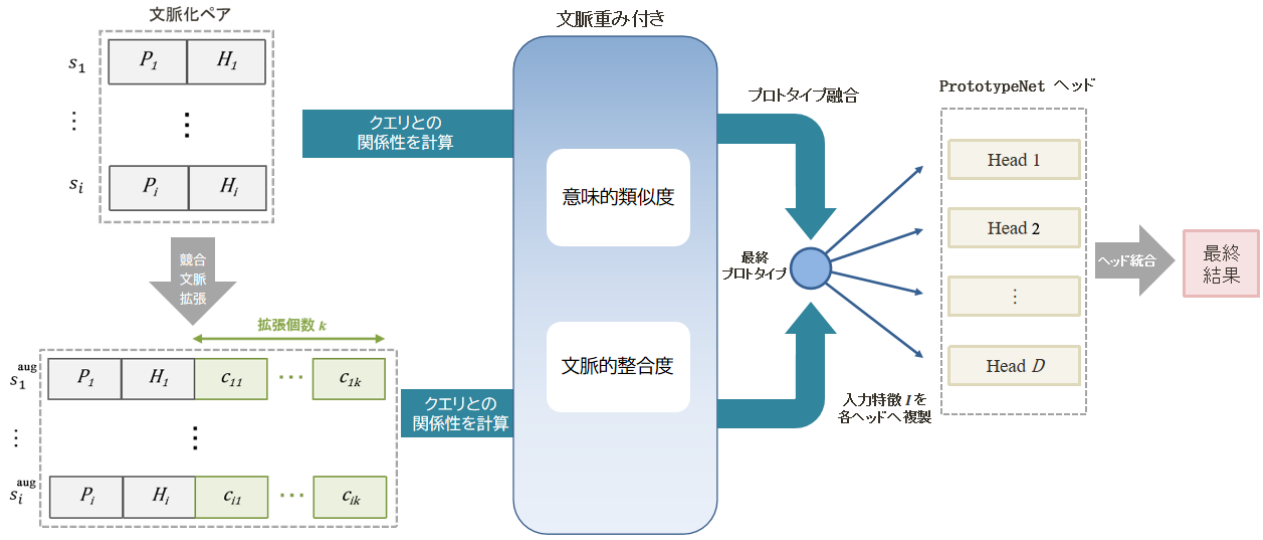

図 1: 提案手法のフローチャート

を導入する．ここで τ は類似度分布の鋭さを制御する温度パラメータである．一方，文脈的整合度を評価するため，軽量な MLP を導入する．クエリ q とサポート例 $s_i^* \in \{s_i, s_i^{\text{aug}}\}$ の相互作用ベクトル

$$[q, s_i^*, q \circ s_i^*, q - s_i^*]$$

を入力とし，次式によりスコア r_i^* を算出する．

$$h_i^* = \text{ReLU}(W_1[q, s_i^*, q \circ s_i^*, q - s_i^*] + b_1),$$
$$r_i^* = W_2 h_i^* + b_2$$

ここで r_i^* は s_i^* が q にどれほど文脈的に整合しているか文脈的整合度を表すスコアである．この類似度項 $\text{sim}(q, s_i^*)$ と整合度項 $\sigma(ar_i^* + b)$ を掛け合わせ，最終的なサポート重み $w_i^* \in \{w_i, w_i^{\text{aug}}\}$ を次のように定義する．

$$w_i^* = (\text{sim}q, s_i^*) \cdot \sigma(ar_i^* + b)$$

ここで a, b はスケーリング定数，σ はシグモイド関数である．重みの正規化を適用したプロトタイプ $p_t^{*,j} \in \{p_t^j, p_t^{\text{aug},j}\}$ 構築は以下の通りである．

$$p_t^{*,j} = \frac{1}{n} \sum_{i=1}^{n} \text{RoBERTa}(\frac{w_i^*}{\sum_j w_j^*})$$

これにより，各クラスのプロトタイプ $p_t^{*,j}$ は，意味的距離と文脈的な一致度による補正因子の両者を掛け合わせることで極端なケースを抑制し，少数ショット環境でも安定した重み付けされた表現となる．

3.3 プロトタイプ融合とヘッド統合

少数ショット設定では，データ点が少ないため局所的・相対的情報のいずれか一方に偏ると不安定化する．そのため，少数ショット下で過度な結合関数学習を避け，p_t^j と $p_t^{\text{aug},j}$ 両者の情報を静的に補完する線形混合により，汎化的な表現の最終プロトタイプ $p_t^{\text{final},j}$ を得る．

$$p_t^{\text{final},j} = \alpha\, p_t^{\text{aug},j} + (1 - \alpha)\, p_t^j, \quad \alpha \in [0, 1]$$

$p_t^{\text{final},j}$ と q の類似度に基づき，クラス確率を計算する部分はベースラインの UFO-ENTAIL と同様である．

競合文脈拡張は相対情報を付与する一方で，他候補の文脈品質が均質でない場合にノイズを混入しやすい．これは特に少数ショット条件下で顕著であり，相対文脈情報が有効に働く例と逆効果をもたらす例が混在するためである．このような競合文脈由来の表現分散を抑制するため，ベースラインの単一のマッチングヘッドを用いる代わりに，複数の PrototypeNet を並列に配置したマルチヘッドの出力を統合する．各ヘッド $d = 1, \ldots, D$ は同一の結合特徴

$$I = [p^{\text{final},j},\ q,\ p^{\text{final},j} \circ q,\ p^{\text{final},j} - q]$$

を入力として共有し，それぞれが独立した PrototypeNet によってクラススコア $z_j^{(d)} = \text{PrototypeNet}(I)$ を出力する．推論時は，各ヘッドのクラススコアを平均して最終スコアを得る．

$$\bar{z}_j = \frac{1}{D} \sum_{d=1}^{D} z_j^{(d)}, \qquad P(y = j|q) = \frac{\exp(\bar{z}_j)}{\sum_{j'} \exp(\bar{z}_{j'})}$$

訓練時は各ヘッドごとに独立に損失を計算し，その平均を最終損失として最適化する．このように，単一ヘッ

ド構造を多ヘッドの平均によって統合することで，競合文脈ノイズに対する分散低減と安定性を実現している．

4 評価実験

4.1 実験設定

本実験では，MCTest500 データセットをターゲットタスク T として学習・評価に使用して，提案手法の有効性を検証した．MCTest500 は複数選択式読解問題から構成され，各設問は 1 つの前提文と 4 つの仮説文を含むデータセットである．このデータセットは，150〜300 語程度の段落と，それに基づく 4 肢選択式の問題を 500 問含む読解ベンチマークであり，各文脈を正例と負例の二値 $j \in \{\text{entailment}, \text{non-entailment}\}$ に変換して扱う．学習，テスト，検証データへの分割は UFO-ENTAIL と同様とする．ソースタスク S としては MNLI を用い，事前学習済みチェックポイントの学習に使用する．このデータセットは，文章ペアに対して {Entailment, Neutral, Contradiction} の関係を型付けした大規模な自然言語含意タスクの汎用的なコーパスである．{Entailment} $\mapsto$ {entailment}，{Neutral, Contradiction} $\mapsto$ {non-entailment} の二型判定問題に再定式化して学習する．評価指標は MCTest500 の各設問に型の偏りが小さいことから平均正解率のみを用いる．

RoBERTa エンコーダの最終 6 層のみをファインチューニングの対象とした．ハイパーパラメータは，学習率は 6×10^{-5}，融合率 $\alpha = 0.6$ に固定し，それ以外は各ショット数ごとに探索する．学習は検証セットでの性能を基準に最良チェックポイントを保存し，最終的な評価はテストセット上で行う．

さらに，提案手法の各構成要素の効果を検証するため，以下のアブレーション設定を実施した．

- +競合文脈: 拡張文脈化ペアで構成されるプロトタイプを追加し，オリジナルプロトタイプと線形融合することを適用する．文脈重み付けを適用しないときでも単独に機能できる．
- +文脈重み：意味的類似度および文脈整合度に基づく重み付けのみ適用する．
- +類似度／+整合度：文脈重み付きプロトタイプの構成要素のうち一方のみを適用する．
- +ヘッド統合：プロトタイプヘッドを構築し，ヘッドごとに独立な損失計算を適用する．

各設定においてショット数 $n = 1, 3, 5, 10$ で実験を行い，5 つの乱数シード $30, 33, 36, 39, 42$ それぞれについて 6 回の試行を実施し，平均正解率を確認した．

4.2 実験結果

表 1 に各アブレーション設定の平均正解率を示す．

競合文脈のみを導入した場合，UFO-ENTAIL に対して平均で+3.2%の精度向上が見られた．さらに，UFO-ENTAIL にヘッド統合を追加した場合，$n = 3$ のとき+0.47%の改善が見られた．競合仮説を用いた文脈拡張により，少数ショット下でも相対的な文脈差の表現能力が高まり，加えて複数ヘッドの導入により，プロトタイプ形成時のノイズや偏りを平均化し，出力分布の安定化が達成されたと考えられる．文脈重み付きを適用する場合においても，同様な現象が確認された．効果の大きさは限定的ではあるが，一定の改善傾向が見られた．

文脈重み付きプロトタイプの導入では，特に $n = 3$ のとき顕著な向上が見られた．これは，意味的類似度と文脈整合度に基づく静的重み付けが，データ不足下でも信頼度の高いサポート例に重みを集中させた効果と考えられる．一方で，類似度または整合度の一方のみを使用した場合はそれぞれ+4.4%に対して+3.9%，+3.3%と低下し，両者の併用が必要であることが確認された．

最後に，すべての統合モデルは全体として最も安定した性能を示し，$n = 1 \sim 10$ すべてでベースラインを上回った．特に $n = 3$ のとき+4.4%を達成し，少数ショット条件下での相対文脈強調が有効であった．また，高ショット条件 $n = 10$ でも高い性能が維持されることが確認された．

5 まとめ

本研究では，少数ショットテキスト含意タスクにおける文脈情報の不足に対処するため，UFO-ENTAIL をベースラインとして，競合文脈拡張と文脈重み付けを統合したプロトタイプ融合モデルを提案した．さらにヘッド統合による安定化を加えることで，出力の分散を抑制することに成功した．その結果，ベースラインより最大約 4.4%の精度向上を達成し，特に少数ショット条件下で頑健な推論性能を確認した．

今後の課題として，候補群の中から品質の高い競合文脈を個別に評価し，精度高く抽出する手法を検討することが考えられる．

参考文献

[1] Dagan, I., Glickman, O., Magnini, B.: The PASCAL Recognizing Textual Entailment Challenge, *Machine Learning Challenges*, Springer, Vol. 3944, pp. 177–190 (2006)

表 1: n-shot における各アブレーション設定での MCTest500 テストデータの正解率.

ショット数 n	1	3	5	10
UFO-ENTAIL	0.7170	0.7306	0.7330	0.7423
UFO+競合文脈	0.7511	0.7577	0.7556	0.7701
UFO+競合文脈+文脈重み	0.7521	**0.7747**	0.7532	0.7723
UFO+競合文脈+類似度	0.7466	0.7694	0.7484	0.7588
UFO+競合文脈+整合度	0.7421	0.7637	0.7462	0.7540
UFO+競合文脈+ヘッド統合	0.7539	0.7589	0.7619	**0.7748**
UFO+競合文脈+文脈重み+ヘッド統合	**0.7570**	0.7745	**0.7583**	0.7742

[2] Snell, J., Swersky, K., Zemel, R.: Prototypical Networks for Few-shot Learning, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, pp. 4080–4090 (2017)

[3] Lu, W., Chen, J., Lu, J., Zhou, J.: Attentive Prototype Few-Shot Learning with Capsule Network-based Embedding, *European Conference on Computer Vision (ECCV) Workshops*, pp. 1–17 (2020)

[4] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: RoBERTa: A Robustly Optimized BERT Pretraining Approach, *arXiv preprint arXiv:1907.11692* (2019)

[5] Brown, T. et al.: Language Models are Few-Shot Learners, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33, pp. 1877–1901 (2020)

[6] Yin, W., Rajani, N. F., Radev, D., Socher, R., Xiong, C.: Universal Natural Language Processing with Limited Annotations: Try Few-shot Textual Entailment as a Start, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 8229–8239 (2020)

[7] Du, J., Yin, W., Xia, C., Yu, P. S.: Learning to Select from Multiple Options, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 4600–4614 (2022)

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I.: Attention is All You Need, *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, pp. 5998–6008 (2017)

[9] Richardson, M., Burges, C. J. C.: MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 193–203 (2013)

[10] Hao, Y., et al.: NoVO: A Voting-based Multi-Head Attention Framework for Robust Classification, *Proceedings of the 2021 Conference on Neural Information Processing Systems (NeurIPS) Workshops*, pp. 1–10 (2021)

[11] Arevalo, J., Solorio, T., Montes-y-Gómez, M., González, F. A.: Gated Multimodal Units for Information Fusion, *International Conference on Learning Representations (ICLR) Workshop Track* (2017)

[12] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R., Bengio, Y.: Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pp. 2048–2057 (2015)