Focal Loss を用いた Transformer による灯謎問題解答システムの構築

1 はじめに

近年深層学習の発展により, 人工知能による漫画や小説などの人間の創作物への理解といった分野の研究が盛んである. 本研究では創作物の一種「クイズ」に注目し, 中国の伝統的クイズゲーム「灯謎 (トウメイ)」を深層学習の手法で解く方法について提案する.

2 灯謎 (トウメイ)

灯謎は中国の伝統的クイズである. 質問者は問題を詩や熟語の形で出し, 回答者はそれに回答する. 答えは常に字または単語になる. 灯謎は質問に答えるための問題文以外の文書や知識など必要がないものが多く, 質問の文中から答えの情報を得ることが容易である. つまり, 質問を理解すれば回答できると考えられる. 灯謎を解くためには, 問題に隠された情報をもとに, 問われている内容を理解して抽出しなければならないため, 灯謎の研究は一種の情報抽出として考えることもできる.

灯謎のパターンは主に謎とヒントと答えで構成される. 謎は詩や熟語や普通の話し言葉で記述された文である. ヒントは答えの形を説明する文である. ヒントは 1 つ以上与えられ, 答えは字か単語である, 問題に隠された字の構成, 発音, 意味などの情報から解くことができる. 図 1 に灯謎の例を示す.

本研究では灯謎問題のうち,「字謎」と呼ばれる答えが一つの漢字のみとなる種類のみについて考える. 字謎の答えは, 単語の意味に加えて漢字の形も強く関わるので, 単純に大量の問題の文の情報のみをニューラルネットワークで学習しても効果が薄いと予想される. そこで本研究では漢字の形の情報に着目し, 漢字を構成する「SUB 漢字」成分と「漢字の画」成分を利用した Transformer モデルで灯謎問題解答システムを構築した.

3 要素技術

3.1 Transformer

Transformer[1] は 2017 年に Google が発表した Encoder Decoder 構造のモデルである. RNN や CNN などを使わず, Self Attention mechanism のみ使用して,

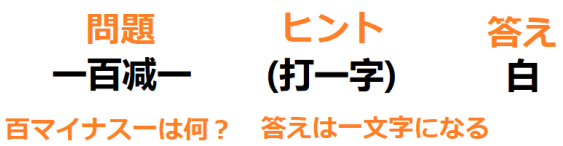

図 1: 灯謎の例

位置埋め込みの情報から Token(単語や漢字等) の重要性を計算することで, 高速な並列計算の実現が可能になる. 図 2 に Transformer のモデル構造を示す.

3.2 Focal Loss

Focal Loss[2] は 2017 年に Facebook が発表した物体検出を対象とする損失関数である. 分類クラス間のデータ不均衡である問題を解決するため, Focal Loss は分類が容易なサンプルの重みを下げることで, 分類が困難なサンプルにより焦点をあてる. この方法により, サンプル数が少ないクラスや分類が難しいサンプルに対して学習しやすくなる特徴がある.

次に Focal Loss の式を示す.

$$FL(p_t) = -(1-p_t)^{\gamma} log(p_t) \tag{1}$$

3.3 漢字の分解手法

従来の研究により, 漢字は常に中国語の最小単位として扱われている. そこで生じる Out-Of-Vocabulary 問題を解決するため,Cao らは 2020 年に木構造に基づいた漢字から SUB 漢字の分解手法 Hierarchical decomposition embedding (HDE)[3] を提案した. この手法では, 漢字は「構造」を親ノード,「SUB 漢字」を子ノードとし, 先行順で漢字を分解する.

そして 2021 年には Chen らがその手法を基に, 漢字を漢字の画の書き順で分解する手法を提案した [4].

図 3 に漢字の分解手法を示す.

3.4 Levenshtein 距離

Levenshtein 距離は,2 つの文字列がどの程度異なっているかを示す距離の一種である. 具体的には,1 つの

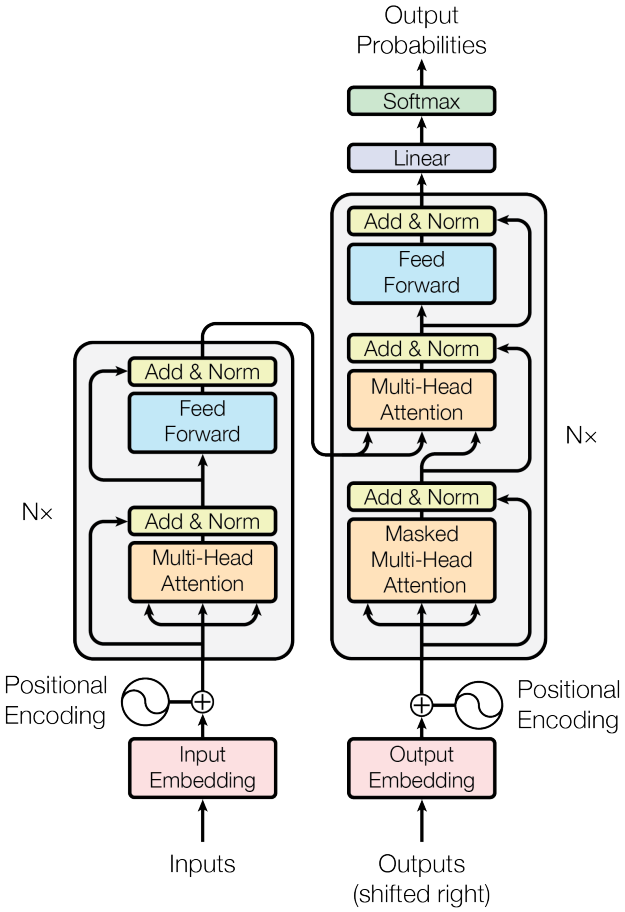

図 2: Transformer[1]

表 1: 字謎の種類

問題の総数	文中の字の情報のみで解ける問題	文中の字の情報のみで解けない問題
79725	72937	6788

文字列からもう一方の文字列に変形するのに必要な手順 (1 文字の挿入, 削除, 置換) の最小回数を計算する.

4 データセット

4.1 中華灯謎データベース

中華灯謎ベース[1]は, 中国各地の灯謎ファン達が集めた灯謎問題 1,362,911 件を収録したデータベースである.

本研究では灯謎のヒントの文を使わないため, 答えが 1 文字である問題 79,725 件のみ利用し, 研究用の灯謎のデータセットを構築した.

研究用灯謎データセットについて, 文中の字の情報のみで解ける問題 72,937 件のみを扱った.

表 1 に問題の種類を示す.

[1]http://www.zhgc.com/mk/

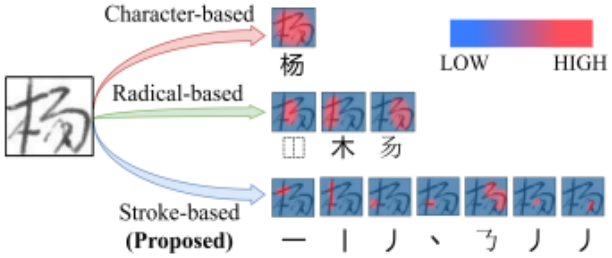

図 3: 漢字の分解手法 [4]

表 2: 漢字成分データセット

入力種数	出力種数 (SUB 漢字)	出力種数 (漢字の画)
447	420	9

4.2 漢字成分データセットの構築

Ideographic Description Sequence (IDS)[2] は, 中国語, 日本語, 韓国語の漢字データを「unicode」,「漢字」,「SUB 漢字」の形で集めたデータセットである.

「SUB 漢字」は漢字の構造を表示する表意文字記述文字 (Ideographic Description Characters) と漢字の部首などの成分で構成している, しかしその分け方は大まかである「SUB 漢字」となっている. そのため今回の実験は IDS データセットと HDE 法を利用し, 木構造で IDS データセットの「SUB 漢字」を更に細かく分解し漢字の「SUB 漢字」成分データセットを構築した. そのうち, 分けられない「独体字 (「人」や「水」等の漢字)」には特定なトークン「O」を補足した.

漢字の「画」データについて,Chen らの手法により「横棒」,「縦棒」,「左払い」,「点」,「鉤」の 5 種類に分類できる. この手法と TorchText ライブラリの「unk」,「pad」,「sos」,「eos」トークンを利用し漢字の画成分データセットを構築した. 表 2 に漢字成分データセットの構成を示す.

5 提案手法

今回利用した灯謎問題は「答えの上下構造と左右構造の漢字は圧倒的多い」というデータ不均衡問題がある. そのため問題を入力する時, モデルは「上下構造」と「左右構造」の予測を優先する傾向がある.

その問題を解決するため, 本研究は Focal Loss を導入した Transformer の手法を提案する.

図 4 にモデルの構造を示す. 灯謎問題を「SUB 漢字」に分け分散表現を生成し,Transformer Encoder に入力

[2]http://github.com/cjkvi/cjkvi-ids

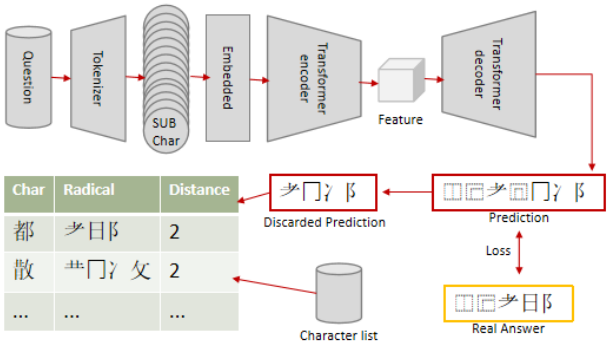

図 4: 提案モデルの構造

表 3: 資料のデータ数

Train Data	Valid Data	Test Data
58350	7293	7294

する.

続いて Transformer Decoder で答えを「SUB 漢字」あるいは漢字の「画」で出力する. 最後は生成した「SUB 漢字」や漢字の「画」を全漢字の「SUB 漢字」や漢字の「画」情報との Levenshtein 距離を計算し, 最も近い漢字 10 個を候補とする.

6 実験

6.0.1 データ処理

本実験では問題文中の字の情報のみで解ける問題 (字謎) で作成した研究用の灯謎データ 72,937 件を利用した. その中からランダムに Train Data 58,350 件, Valid Data 7,293 件, Test Data 7,294 件を抽出し, 実験に使用した.

表 3 にデータ数を示す.

6.1 実験内容

実験では Train Data と Valid Data に対し,「SUB 漢字入力 + SUB 漢字出力」と「SUB 漢字入力 + 漢字の画出力」の条件でそれぞれ「Cross Entropy」と「Focal Loss」による実験を実施し Train Loss と Valid Loss を計算する. そして Valid Data のみ「Precision」,「Recall」,「F1 値」を利用して生成したものの正解率を評価する.

そして Test Data に対し,「候補の中に正解があるがどうか」を表示する「Top1」,「Top3」,「Top5」,「Top10」の hit 率と「正解までどれほど改善が必要か」を表示する「平均 step」で Test Data の実験結果を評価する. 表 4 に実験モデルのパラメータを示す.

表 4: 実験用パラメータ (通用)

パラメータ	数値
分散表現の次元数	512
隠れ層の次元数	512
マルチヘッド	4
バッチサイズ	36
Dropout	0.1
最適化手法	Adam
学習率	0.00001
Epoch	200

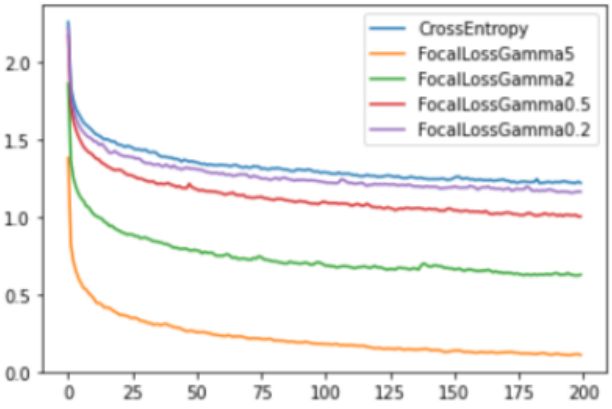

図 5: 訓練誤差曲線 (SUB 漢字)

6.2 実験結果

200 Epoch を経た訓練結果として,「Focal Loss (gamma = 5)」の条件で「SUB 漢字入力 + SUB 漢字出力」実験の Train Loss と Valid Loss はそれぞれ 0.107 と 0.229 に収束し,「SUB 漢字入力 + 漢字の画出力」実験の Train Loss と Valid Loss はそれぞれ 0.050 と 0.066 に収束した. 故に Focal Loss は訓練誤差の減少に良い効果があることが確認した.

図 5 と図 6 に実験の訓練誤差変化曲線を示す.

表 5 に訓練結果を示す.

テスト結果について,「SUB 漢字入力 + 漢字の画出力」実験で生成した漢字「漢字の画」の正解率は「SUB 漢字入力 + SUB 漢字出力」実験で生成した「SUB 漢字」の正解率より高いが,「正解に復元するステップ数」と「ヒット率」は「SUB 漢字入力 + SUB 漢字出力」実験の方が高いことを確認した. 表 6 にテスト結果を示す.

表 5: 訓練結果

	SUB 漢字					漢字の画				
Loss function	Train Loss	Valid Loss	Precision	Recall	F1	Train Loss	Valid Loss	Precision	Recall	F1
Cross Entropy	1.217	1.279	0.311	0.303	0.298	0.931	0.922	0.519	0.326	0.378
Focal Loss (gamma = 5)	0.107	0.229	0.188	0.275	0.215	0.050	0.066	0.418	0.377	0.385
Focal Loss (gamma = 2)	0.624	0.680	0.214	0.242	0.215	0.336	0.330	0.735	0.499	0.547
Focal Loss (gamma = 0.5)	1.001	1.097	0.361	0.266	0.290	0.856	0.848	0.527	0.326	0.353
Focal Loss (gamma = 0.2)	1.160	1.216	0.366	0.253	0.287	0.940	0.912	0.720	0.593	0.624

表 6: テスト結果

	SUB 漢字								漢字の画							
Loss function	Precision	Recall	F1	Top1	Top3	Top5	top10	Avg Step	Precision	Recall	F1	Top1	Top3	Top5	top10	Avg Step
Cross Entropy	0.344	0.320	0.318	0.010	0.027	0.028	0.062	4.415	0.645	0.620	0.602	0.005	0.009	0.011	0.016	7.259
Focal Loss (gamma = 5)	0.205	0.274	0.223	0.005	0.023	0.025	0.028	3.829	0.560	0.496	0.502	0.002	0.006	0.008	0.012	8.142
Focal Loss (gamma = 2)	0.281	0.299	0.276	0.007	0.027	0.031	0.041	4.015	0.729	0.571	0.601	0.005	0.012	0.015	0.021	9.621
Focal Loss (gamma = 0.5)	0.371	0.293	0.310	0.020	0.044	0.047	0.061	5.457	0.609	0.472	0.484	0.003	0.006	0.013	0.013	24.298
Focal Loss (gamma = 0.2)	0.375	0.284	0.309	0.009	0.022	0.026	0.040	5.132	0.673	0.614	0.609	0.005	0.010	0.013	0.018	7.493

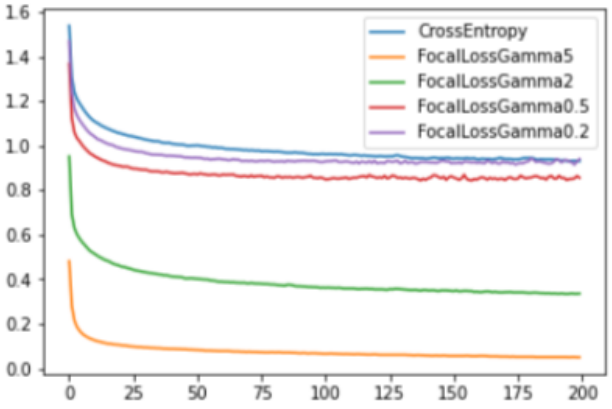

図 6: 訓練誤差曲線 (漢字の画)

7　まとめと今後の課題

本研究は,Focal Loss を用いた Transformer による灯謎問題解答モデルを構築し,「SUB 漢字入力 ＋ SUB 漢字出力」と「SUB 漢字入力 ＋ 漢字の画出力」の条件で Focal Loss の有効性を確認した. 結果として,Focal Loss は訓練誤差の減少に良い効果があることが確認した.

今後の課題として, 灯謎の「ヒント」部分で答えの漢字の画数や構造を予測するモデルと Beam Search デコーダーの導入でモデルの精度向上のを目指し, 問題の正解率の向上についてもふれておく.

参考文献

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc., 2017.

[2] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *CoRR*, Vol. abs/1708.02002, , 2017.

[3] Zhong Cao, Jiang Lu, Sen Cui, and Changshui Zhang. Zero-shot handwritten chinese character recognition with hierarchical decomposition embedding. *Pattern Recognit.*, Vol. 107, p. 107488, 2020.

[4] Jingye Chen, Bin Li, and Xiangyang Xue. Zero-shot chinese character recognition with stroke-level decomposition. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 615–621. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Main Track.