DualCSE: 文の明示的・暗黙的意味の埋め込み学習

小田 康平 [1] 荘 博閔 [2] 白井 清昭 [1] Natthawut Kertkeidkachorn[1]
[1] 北陸先端科学技術大学院大学 先端科学技術研究科
[2] 株式会社東芝 総合研究所
[1]{s2420017,kshirai,natt}@jaist.ac.jp
[2]pomin.chuang.x51@mail.toshiba

概要

文埋め込み手法は近年目覚ましい発展を遂げているものの，文の暗黙的な意味を処理することは依然として困難な課題である．この要因のひとつとして，これまでの文埋め込み手法が 1 つの文に対し 1 つの埋め込み（ベクトル）しか与えないことが挙げられる．これに対し，本研究では 1 つの文に対し明示的な意味と暗黙的な意味を表す 2 つのベクトルを付与する手法である DualCSE を提案する．これらのベクトルは同一空間内で共存しているため，情報検索やテキスト分類など，目的に応じて必要な意味を選択して利用することができる．実験により，DualCSE は文の明示的および暗黙的な意味を適切に符号化できることが示された．[1)]

1 はじめに

文埋め込みは自然言語処理分野において広く研究されてきた [2, 3, 4]．しかしながら，ほとんどの既存手法は文の暗黙的な意味[2)]を捉える能力に課題を抱えている．Sun らによると，最先端の文埋め込み手法 [5, 6, 7] であっても，MTEB の分類タスク [8] における明示的な意味と暗黙的な意味に対する精度の間に約 0.2 ポイントもの差が存在する [9]．この要因のひとつとして，既存手法は文に複数の解釈があることを想定せず，1 つの文に 1 つのベクトルしか割り当てないことが考えられる．

この課題に対処するため，本研究では 1 つの文に対しその文の明示的な意味と暗黙的な意味を表す 2 つの埋め込みを割り当てる手法を提案する．以下，本手法をdual-semantic constrastive sentence embedding framework (DualCSE) と呼ぶ．図 1 に示すように，DualCSE では文の明示的な意味と暗黙的な意味は同一空間上の異なるベクトルで表現される．例えば，“She conquered his heart.” (s_2) の明示的な意味は “She defeated his heart in battle.” (s_3) の明示的な意味に近く，s_2 の暗黙的な意味は “She won his affection and love.” (s_1) の明示的な意味に近い．さらに，s_1 および s_3 は明確に暗黙的な意味を持たないため，それぞれにおける明示的な意味と暗黙的な意味の類似度は s_2 のそれより高い．DualCSE により得られた文の明示的および暗黙的な意味の埋め込みは，従来の単一のベクトルを付与する文埋め込み手法では困難であった言外の意味の表現を可能にし，情報検索 [10] やテキスト分類 [11] といった基本的なタスクにおいて有用な文の抽象表現を提供するだけでなく，入力文の暗黙性の度合いを推定する指標 [12] の計算にも活用できる．DualCSE は，従来の教師あり文埋め込み手法 [13, 14, 15] に基づき，自然言語推論 (Natural Language Inference; NLI) データセットを用いた対照学習 [16] により文のエンコーダを訓練する．具体的には，明示的な意味と暗黙的な意味の双方を考慮した NLI データセット [17] を学習データとして利用し，本研究で新たに導入する対照損失によりパラメータを更新する．

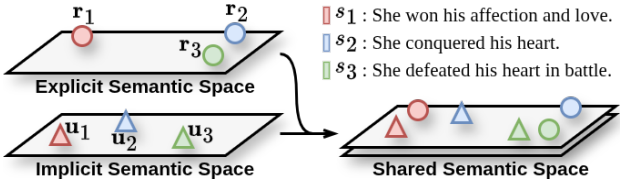

図 1 DualCSE の概要．明示的意味と暗黙的意味は共通の意味空間で表現される．

DualCSE の文間および文内の関係を捉える能力を，含意関係認識 (Recognizing Textual Entailment; RTE) および暗黙度推定 (Estimating Implicitness Score; EIS) の 2 タスクで評価する．実験により，DualCSE は従来手法よりも文間および文内の関係をより正確

1) 本論文の完全版は [1] を参照されたい．
2) 本論文では，「明示的な意味」を字義的な意味，「暗黙的な意味」を比喩的あるいは語用的な用法に由来する非字義的な意味として用いる．

表 1　INLI データセット内のデータサンプルの例

Premise Diane says, “Would you like to go a party tonight?” Sophie responds, “I am too tired.”
Implied Entailment Sophie would prefer not to attend the party this evening.
Explicit Entailment Sophie claims to be too tired.
Neutral The party will take place outside.
Contradiction Sophie is excited to attend the party this evening.

に捉えることを示す．

2　Implied NLI (INLI) データセット

本節では，DualCSE の訓練データとして使用される INLI データセット [17] について説明する．INLI データセットは，SNLI [18] や MNLI [19] といった標準的な NLI データセットとは異なり，表 1 に示すように，1 つの前提に対し implied-entailment，explicit-entailment，neutral，contradiction のラベルが付与された 4 つの異なる仮説が提供されている．ラベル implied-entailment と explicit-entailment は，それぞれ前提の暗黙的な意味と明示的な意味に対する含意を表し，言語モデルの文の暗黙的な意味を捉える能力を強化したり評価したりすることに役立つ．[3)]

3　DualCSE

本節では，DualCSE の詳細を述べる．DualCSE は入力文を明示的な意味と暗黙的な意味を表す 2 つの埋め込みに符号化する．まずそれらの埋め込みを学習するための損失関数を説明し，続いて埋め込みを生成するエンコーダについて述べる．

3.1　対照損失

INLI データセットの i 番目のサンプルについて，前提を s_i，explicit-entailment，implied-entailment，contradiction の仮説をそれぞれ s_{i1}^+，s_{i2}^+，s_i^- とし，それらの明示的な意味の埋め込みをそれぞれ $\mathbf{r}_i$，$\mathbf{r}_{i1}^+$，$\mathbf{r}_{i2}^+$，$\mathbf{r}_i^-$，暗黙的な意味の埋め込みをそれぞれ $\mathbf{u}_i$，$\mathbf{u}_{i1}^+$，$\mathbf{u}_{i2}^+$，$\mathbf{u}_i^-$ とすると，サイズが N のミニバッチにおける i 番目のサンプルに対する対照損失は以下のように計算される．

$$v(\mathbf{h}_1, \mathbf{h}_2) = e^{\mathrm{sim}(\mathbf{h}_1, \mathbf{h}_2)/\tau}. \tag{1}$$

3)　INLI データセットの統計は付録 A に記載されている．

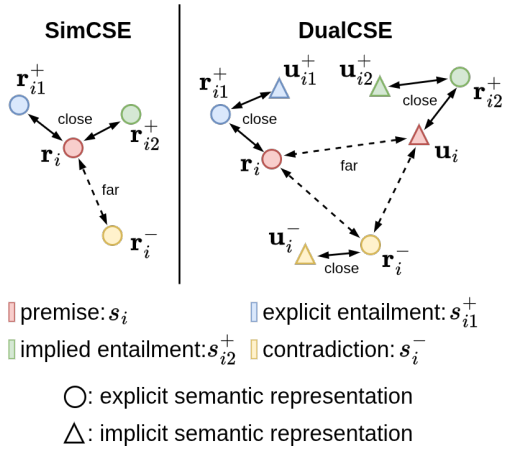

図 2　SimCSE[13] と DualCSE の対照損失の概念図

$$\begin{aligned} l_i = &- \log \frac{v(\mathbf{r}_i, \mathbf{r}_{i1}^+)}{\sum_{j=1}^N (v(\mathbf{r}_i, \mathbf{r}_{j1}^+) + v(\mathbf{r}_i, \mathbf{r}_j^-) + v(\mathbf{r}_i, \mathbf{u}_j))} \\ &- \log \frac{v(\mathbf{u}_i, \mathbf{r}_{i2}^+)}{\sum_{j=1}^N (v(\mathbf{u}_i, \mathbf{r}_{j2}^+) + v(\mathbf{u}_i, \mathbf{r}_j^-) + v(\mathbf{u}_i, \mathbf{r}_j))} \\ &- \log \frac{v(\mathbf{r}_{i1}^+, \mathbf{u}_{i1}^+)}{\sum_{j=1}^N v(\mathbf{r}_{i1}^+, \mathbf{u}_{j1}^+)} - \log \frac{v(\mathbf{r}_{i2}^+, \mathbf{u}_{i2}^+)}{\sum_{j=1}^N v(\mathbf{r}_{i2}^+, \mathbf{u}_{j2}^+)} \\ &- \log \frac{v(\mathbf{r}_i^-, \mathbf{u}_i^-)}{\sum_{j=1}^N v(\mathbf{r}_i^-, \mathbf{u}_j^-)}. \end{aligned} \tag{2}$$

式 (1) における v は 2 つのベクトルの類似度であり，$\mathrm{sim}(\mathbf{h}_1, \mathbf{h}_2)$ は 2 つのベクトル $\mathbf{h}_1$ と $\mathbf{h}_2$ のコサイン類似度，τ は温度パラメータである．一方，式 (2) は対照損失の定義である．ここで，既存手法の SimCSE[13] と DualCSE における対照損失の直感的な説明を図 2 に示す．ペア $(\mathbf{r}_i, \mathbf{r}_{i1}^+)$ と $(\mathbf{u}_i, \mathbf{r}_{i2}^+)$ は互いに近づき，ペア $(\mathbf{r}_i, \mathbf{r}_i^-)$ と $(\mathbf{u}_i, \mathbf{r}_i^-)$ は互いに離れるようにエンコーダが学習される．[4)] すなわち，前提と含意の仮説の埋め込みの類似度は高く，前提と矛盾の仮説の埋め込みの類似度は低くなる．さらに，ペア $(\mathbf{r}_{i1}^+, \mathbf{u}_{i1}^+)$，$(\mathbf{r}_{i2}^+, \mathbf{u}_{i2}^+)$，$(\mathbf{r}_i^-, \mathbf{u}_i^-)$ は互いに近づくよう促され，[5)] ペア $(\mathbf{r}_i, \mathbf{u}_i)$ は互いに離れるよう促される．[6)] これらは，INLI データセット内の仮説は前提よりも暗黙性の度合いが低く，仮説の明示的な意味と暗黙的な意味の埋め込みは類似するが，前提の明示的な意味と暗黙的な意味の埋め込みは類似しないと仮定して設計されている．

3.2　エンコーダ

本研究では，明示的・暗黙的意味の埋め込みを生成するエンコーダとして以下の 2 種類のモデルを用

4)　式 (2) の右辺の第 1 項と第 2 項に該当する．
5)　式 (2) の右辺の第 3，4，5 項に該当する．
6)　式 (2) の右辺の第 1 項と第 2 項の分母の $v(\mathbf{r}_i, \mathbf{u}_j)$ と $v(\mathbf{u}_i, \mathbf{r}_j)$ に該当する．

いる．

Cross-Encoder 単一の BERT[20] もしくは RoBERTa[21] モデルである．明示的な意味の表現 $\mathbf{r}$ を取得するときは入力 “[CLS] s [SEP] explicit” を与え，暗黙的な意味の表現 $\mathbf{u}$ を取得するときは入力 “[CLS] s [SEP] implicit” を与える．

Bi-Encoder 2 つの独立した BERT もしくは RoBERTa モデルである．一方のモデルが $\mathbf{r}$ を，もう一方のモデルが $\mathbf{u}$ を出力する．

両エンコーダにおいて，トークン [CLS] の最終層の隠れ状態が文埋め込みとして使用される．

4 評価実験

文間および文内の関係を表現できているかという観点から，DualCSE の有効性を 2 つのタスクにより検証する．1 つ目のタスクは含意関係認識 (RTE) であり，モデルが文間の含意関係を正しく捕捉できているかを評価する．2 つ目のタスクは暗黙度推定 (EIS) であり，入力文の暗黙的な意味が文字通りの意味からどの程度逸脱しているかを推定する．

4.1 実験設定

Cross-Encoder，Bi-Encoder の両方で，事前学習済み $\text{BERT}_{\text{base}}$ と $\text{RoBERTa}_{\text{base}}$ をエンコーダモデルとして使用した．本節では，開発データにおいて BERT より高い性能を示した RoBERTa モデルの設定と結果のみを報告する．バッチサイズと学習率は開発データを用いて最適化され，Cross-Encoder ではそれぞれ 64 と 5e-5，Bi-Encoder ではそれぞれ 32 と 3e-5 となった．また，温度パラメータ τ は先行研究 [13, 22] を参考に 0.05 に設定した．

4.2 RTE タスク

タスク定義 RTE は，与えられた前提 p と仮説 h 間の関係を entailment または non-entailment のいずれかに分類するタスクである．実験には INLI データセット [17] を使用し，元のラベル neutral と contradiction は non-entailment に，explicit-entailment と implied-entailment は entailment に変換した．

ラベル予測 前提 p と仮説 h の明示的な意味の表現をそれぞれ $\mathbf{r}_1$ と $\mathbf{r}_2$，前提 p の暗黙的な意味の表現を $\mathbf{u}_1$ とすると，DualCSE は以下の条件を満たす場合に entailment と予測し，それ以外の場合には non-entailment と予測する．

$$\max\big(\cos(\mathbf{r}_1, \mathbf{r}_2), \cos(\mathbf{u}_1, \mathbf{r}_2)\big) > \gamma. \tag{3}$$

表 2 RTE タスクの正解率 (%)．**Exp.**，**Imp.**，**Neu.**，および **Con.** はそれぞれ元のラベルが explicit-entailment，implied-entailment，neutral，および contradiction のサンプルに対する正解率を意味する．

Model	Exp.	Imp.	Neu.	Con.	All
SimCSE (SNLI+MNLI)	79.80	49.00	74.30	67.60	67.68
SimCSE (INLI)	90.60	69.10	66.90	91.00	79.40
DualCSE-Cross (ours)	90.20	73.40	68.40	88.70	80.18
DualCSE-Bi (ours)	91.90	69.90	72.10	87.60	**80.38**
Gemini-1.5-Pro	97.90	80.30	92.00	95.40	91.40

ここで，閾値 γ は開発データを用いて調整される．

ベースライン DualCSE と比較するベースラインとして SimCSE (SNLI+MNLI)[13] と SimCSE (INLI) の 2 つを用意した．前者はオリジナルの SimCSE モデルであり，後者は INLI データセットで訓練された SimCSE モデルである．これらのベースラインは，DualCSE と同様に，前提と仮説をそれぞれ単一のベクトルに符号化し，それらのコサイン類似度が閾値を超えるかどうかでラベルを予測する．さらに参考として，few-shot learning による大規模言語モデル (Large Language Model; LLM) の結果も提供する．

実験結果 RTE タスクの正解率を表 2 に示す．まず，提案手法である DualCSE は両エンコーダとも SimCSE (INLI) を上回り，文の明示的な意味と暗黙的な意味を区別することの有効性を実証している．次に，SimCSE (SNLI+MNLI) は明示的な意味に対する正解率 (Exp.) と暗黙的な意味に対する正解率 (Imp.) の差が最も大きい．これは，先行研究 [17] で報告されているように，SNLI と MNLI に暗黙的な意味を含む文が比較的少ないことが原因であると考えられる．最後に，LLM (Gemini-1.5-Pro) は一貫してエンコーダモデルより優れた性能を示している．しかしながら，エンコーダモデルと同様に，LLM も Exp. と比べ Imp. が低下する傾向が見られる．[7)]

4.3 EIS タスク

タスク定義 2 つの文 s_1 と s_2 が与えられ，どちらの文がより高い暗黙性の度合い（暗黙度）を示すかを予測する．評価には Wang らのデータセット [12] と INLI データセット [17] を用いる．Wang らのデータセットは暗黙的な意味を持つ文と持たない文のペアからなり，前者を暗黙度が高い文として選んだときに正解となる．INLI では，前提が仮説より暗黙的であると仮定し，前提を選んだときに正解する．

7) 他の LLM の結果は付録 B に記載する．

表 3　テキスト検索の一例．**Explicit semantic** と **Implicit semantic** は，INLI データセットにおける explicit-entailment と implied-entailment にそれぞれ相当する．これらは文検索のクエリとして使用していない．

Query: Madeleine has just moved into a neighbourhood and meets her new neighbour Pierre. Pierre says, "Are you from this state?" Madeleine responds, "I'm from Oregon."	
Explicit semantic: Madeleine is from Oregon.	**Implicit semantic**: Madeleine was born in a different state.
#1 Laverne moved from Canada. #2 Angela and her family live in Portland now. #3 Alyce works in Portland.	#1 The place does not belong to Quincy. #2 Madeleine enjoys food with some spice, but not if it's overly hot. #3 Earlene is not originally from this area.

表 4　EIS タスクの正解率 (%)

Model	INLI	Wang ら [12]
Length	99.90	73.37
ImpScore (original)	80.55	**95.20**
ImpScore (INLI)	99.97	81.56
DualCSE-Cross (ours)	99.97	79.31
DualCSE-Bi (ours)	**100**	77.48

暗黙度推定　文 s の暗黙度のスコアを式 (4) で計算し，式 (5) のように s_1 と s_2 のうちそれが高い文を選択する．

$$\mathrm{imp}(s) = 1 - \cos(\mathbf{r}, \mathbf{u}), \tag{4}$$

$$\arg\max\big(\mathrm{imp}(s_1), \mathrm{imp}(s_2)\big). \tag{5}$$

ベースライン　本実験では，Length，ImpScore (original)[12]，および ImpScore (INLI) の 3 つのベースラインを用意した．Length は s_1 と s_2 の内，長い方の文をより暗黙度が高いと予測する．ImpScore (INLI) は RoBERTa をエンコーダモデルとし，INLI データセットで訓練した ImpScore モデルである．

実験結果　EIS タスクの正解率を表 4 に示す．まず，DualCSE は INLI データセット，すなわち in-domain 設定においてほぼ 100%の正解率を達成した．ただし，これは入力文の長さが暗黙度の高い文を特定するための有効な特徴量として機能したためと考えられる．実際，Length もほぼ 100%の正解率を達成していることから，この仮説が裏付けられる．次に，Wang らのデータセット，すなわち out-of-domain 設定では，DualCSE と ImpScore（INLI）の正解率は共に約 80%まで低下する．また，DualCSE-Cross の正解率は ImpScore (INLI) とほぼ同等である．特筆すべきは，ImpScore が暗黙度を推定するために開発された手法であるのに対し，DualCSE は明示的および暗黙的な意味に対する表現を生成する汎用的なモデルであり，他の下流タスクにも応用できる点である．[8)]

8）　他のモデルの結果は付録 C に記載する．

表 5　アブレーション分析の結果

Loss function	RTE	EIS
DualCSE-Cross	**80.18**	**99.97**
w/o contradiction	64.57	99.88
w/o intra sentence	80.10	92.25
w/o contradiction & intra sentence	64.68	32.75

5　分析

アブレーション分析　提案された対照損失の構成要素の寄与を調査するためにアブレーション分析を実施した．具体的には，矛盾の仮説に関する損失を除外した場合，文内の関係に関する損失を除外した場合，およびそれら両方を除外した場合の 3 つの損失関数でモデルを訓練した．表 5 に示すように，矛盾の仮説に関する損失は RTE タスクにおいてより効果的であり，文内の関係に関する損失は EIS タスクにおいてより効果的であった．[9)]

テキスト検索　明示的および暗黙的な意味の表現の定性的な評価として，単純なテキスト検索を実施した．具体的には，INLI の開発データからいくつかの前提をクエリとして選択し，各クエリの明示的な意味と暗黙的な意味のそれぞれに対し，訓練データから類似度の高い仮説 3 件を検索する．表 3 に示す通り，DualCSE は明示的な意味と暗黙的な意味のそれぞれで類似する文を検索できる．

6　おわりに

本論文では，1 つの文に明示的な意味と暗黙的な意味の抽象表現を得るための文埋め込み手法 DualCSE を提案した．RTE および EIS タスクによる実験の結果から，DualCSE が明示的な意味と暗黙的な意味を別々の埋め込みに適切に符号化できることが実証された．本研究で使用した学習データは INLI のみと小規模なため，今後は大規模な学習データを構築しモデルを訓練したい．

9）　アブレーション分析の詳細および他のモデルの結果は付録 D に記載する．

参考文献

[1] Kohei Oda, Po-Min Chuang, Kiyoaki Shirai, and Natthawut Kertkeidkachorn. One sentence, two embeddings: Contrastive learning of explicit and implicit semantic representations. arXiv preprint arXiv:2510.09293, 2025. https://arxiv.org/abs/2510.09293.

[2] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 3982–3992, Hong Kong, China, November 2019. Association for Computational Linguistics.

[3] Ting Jiang, Jian Jiao, Shaohan Huang, Zihan Zhang, Deqing Wang, Fuzhen Zhuang, Furu Wei, Haizhen Huang, Denvy Deng, and Qi Zhang. PromptBERT: Improving BERT sentence embeddings with prompts. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 8826–8837, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

[4] Xianming LI, Zongxi Li, Jing Li, Haoran Xie, and Qing Li. ESE: Espresso sentence embeddings. In **The Thirteenth International Conference on Learning Representations**, 2025.

[5] Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pre-training. arXiv preprint arXiv:2212.03533, 2024.

[6] Xin Zhang, Yanzhao Zhang, Dingkun Long, Wen Xie, Ziqi Dai, Jialong Tang, Huan Lin, Baosong Yang, Pengjun Xie, Fei Huang, Meishan Zhang, Wenjie Li, and Min Zhang. mGTE: Generalized long-context text representation and reranking models for multilingual text retrieval. In Franck Dernoncourt, Daniel Preoţiuc-Pietro, and Anastasia Shimorina, editors, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track**, pp. 1393–1412, Miami, Florida, US, November 2024. Association for Computational Linguistics.

[7] Dun Zhang, Jiacheng Li, Ziyang Zeng, and Fulong Wang. Jasper and stella: distillation of sota embedding models. arXiv preprint arXiv:2412.19048, 2025.

[8] Niklas Muennighoff, Nouamane Tazi, Loic Magne, and Nils Reimers. MTEB: Massive text embedding benchmark. In Andreas Vlachos and Isabelle Augenstein, editors, **Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics**, pp. 2014–2037, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.

[9] Yiqun Sun, Qiang Huang, Anthony K. H. Tung, and Jun Yu. Text embeddings should capture implicit semantics, not just surface meaning. arXiv preprint arXiv:2506.08354, 2025.

[10] Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models. arXiv preprint arXiv:2104.08663, 2021.

[11] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In Dekang Lin, Yuji Matsumoto, and Rada Mihalcea, editors, **Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies**, pp. 142–150, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.

[12] Yuxin Wang, Xiaomeng Zhu, Weimin Lyu, Saeed Hassanpour, and Soroush Vosoughi. Impscore: A learnable metric for quantifying the implicitness level of sentences. In **The Thirteenth International Conference on Learning Representations**, 2025.

[13] Tianyu Gao, Xingcheng Yao, and Danqi Chen. SimCSE: Simple contrastive learning of sentence embeddings. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 6894–6910, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.

[14] Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith Hall, Daniel Cer, and Yinfei Yang. Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, **Findings of the Association for Computational Linguistics: ACL 2022**, pp. 1864–1874, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[15] Xianming Li and Jing Li. AoE: Angle-optimized embeddings for semantic textual similarity. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1825–1839, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[16] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé III and Aarti Singh, editors, **Proceedings of the 37th International Conference on Machine Learning**, Vol. 119 of **Proceedings of Machine Learning Research**, pp. 1597–1607. PMLR, 13–18 Jul 2020.

[17] Shreya Havaldar, Hamidreza Alvari, John Palowitch, Mohammad Javad Hosseini, Senaka Buthpitiya, and Alex Fabrikant. Entailed between the lines: Incorporating implication into NLI. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 32274–32290, Vienna, Austria, July 2025. Association for Computational Linguistics.

[18] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. In Lluís Màrquez, Chris Callison-Burch, and Jian Su, editors, **Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing**, pp. 632–642, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[19] Adina Williams, Nikita Nangia, and Samuel Bowman. A broad-coverage challenge corpus for sentence understanding through inference. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**, pp. 1112–1122, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[20] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[21] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019.

[22] Shohei Yoda, Hayato Tsukagoshi, Ryohei Sasano, and Koichi Takeda. Sentence representations via gaussian embedding. arXiv preprint arXiv:2305.12990, 2024.

A　データセットの統計

INLI データセット [17] と Wang らのデータセット [12] の統計を表 6 に示す．INLI データセットでは前提と仮説のペアの数を，Wang らのデータセットでは暗黙的な意味を持つ文と持たない文のペアの数を示す．

表 6　データセットの統計

Dataset	train	development	test
INLI	32,000	4,000	4,000
Wang ら [12]	101,320	5,630	5,630

B　RTE タスクの全モデルの結果

RTE タスクの全モデルの結果を表 7 に示す．

表 7　RTE タスクの全モデルの正解率 (%)

Model	Exp.	Imp.	Neu.	Con.	All
LLMs					
GPT-4	98.40	83.10	88.90	94.10	91.12
GPT-4o	98.30	84.50	87.20	94.30	91.08
GPT-4o-mini	97.30	74.30	90.30	94.40	89.08
Gemini-1.5-Pro	97.90	80.30	92.00	95.40	**91.40**
Gemini-2.0-Flash	98.20	85.50	85.40	93.40	90.62
Claude-3.7-Sonnet	97.10	75.90	93.00	95.90	90.47
DeepSeek-v3	99.10	85.20	87.40	93.30	91.25
Mistral Large	98.10	81.30	88.70	94.60	90.68
BERT-based					
SimCSE (SNLI+MNLI)	78.50	41.00	77.40	67.50	66.10
SimCSE (INLI)	89.80	67.60	65.70	83.90	76.75
ImpScore (INLI)	59.20	26.30	75.30	81.50	60.58
DualCSE-Cross (ours)	86.80	64.30	72.40	87.50	77.75
DualCSE-Bi (ours)	91.30	63.30	73.60	85.10	**78.32**
RoBERTa-based					
SimCSE (SNLI+MNLI)	79.80	49.00	74.30	67.60	67.68
SimCSE (INLI)	90.60	69.10	66.90	91.00	79.40
ImpScore (INLI)	81.60	56.80	47.70	61.60	61.92
DualCSE-Cross (ours)	90.20	73.40	68.40	88.70	80.18
DualCSE-Bi (ours)	91.90	69.90	72.10	87.60	**80.38**

C　EIS タスクの全モデルの結果

EIS タスクの全モデルの結果を表 8 に示す．

D　アブレーション分析の詳細

アブレーション分析の詳細を以下に記す．

w/o contradiction　損失を以下の式で計算する．

$$l_i = -\log \frac{v(\mathbf{r}_i, \mathbf{r}_{i1}^+)}{\sum_{j=1}^N (v(\mathbf{r}_i, \mathbf{r}_{j1}^+) + v(\mathbf{r}_i, \mathbf{u}_j))} - \log \frac{v(\mathbf{u}_i, \mathbf{r}_{i2}^+)}{\sum_{j=1}^N (v(\mathbf{u}_i, \mathbf{r}_{j2}^+) + v(\mathbf{u}_i, \mathbf{r}_j))} - \log \frac{v(\mathbf{r}_{i1}^+, \mathbf{u}_{i1}^+)}{\sum_{j=1}^N v(\mathbf{r}_{i1}^+, \mathbf{u}_{j1}^+)} - \log \frac{v(\mathbf{r}_{i2}^+, \mathbf{u}_{i2}^+)}{\sum_{j=1}^N v(\mathbf{r}_{i2}^+, \mathbf{u}_{j2}^+)}. \quad (6)$$

表 8　EIS タスクの全モデルの正解率 (%)

Model	INLI	Wang ら [12]
Length	99.90	73.37
ImpScore (original)	80.55	**95.20**
BERT-based		
ImpScore (INLI)	99.97	76.91
DualCSE-Cross (ours)	**100**	80.46
DualCSE-Bi (ours)	99.97	79.88
RoBERTa-based		
ImpScore (INLI)	99.97	81.56
DualCSE-Cross (ours)	99.97	79.31
DualCSE-Bi (ours)	**100**	77.48

w/o intra-sentence　損失を以下の式で計算する．

$$l_i = -\log \frac{v(\mathbf{r}_i, \mathbf{r}_{i1}^+)}{\sum_{j=1}^N (v(\mathbf{r}_i, \mathbf{r}_{j1}^+) + v(\mathbf{r}_i, \mathbf{r}_j^-))} - \log \frac{v(\mathbf{u}_i, \mathbf{r}_{i2}^+)}{\sum_{j=1}^N (v(\mathbf{u}_i, \mathbf{r}_{j2}^+) + v(\mathbf{u}_i, \mathbf{r}_j^-))}. \quad (7)$$

w/o contradiction & intra-sentence　損失を以下の式で計算する．

$$l_i = -\log \frac{v(\mathbf{r}_i, \mathbf{r}_{i1}^+)}{\sum_{j=1}^N (v(\mathbf{r}_i, \mathbf{r}_{j1}^+))} - \log \frac{v(\mathbf{u}_i, \mathbf{r}_{i2}^+)}{\sum_{j=1}^N (v(\mathbf{u}_i, \mathbf{r}_{j2}^+))}. \quad (8)$$

アブレーション分析の全モデルの結果を表 9 に示す．

表 9　アブレーション分析の全モデルの結果

Loss function	RTE	EIS
DualCSE-Cross-BERT	**77.75**	**100**
w/o contradiction	64.13	99.90
w/o intra sentence	77.50	47.13
w/o contradiction & intra sentence	64.38	31.83
DualCSE-Bi-BERT	**78.32**	99.97
w/o contradiction	65.97	**100**
w/o intra sentence	77.30	63.42
w/o contradiction & intra sentence	65.47	81.35
DualCSE-Cross-RoBERTa	**80.18**	**99.97**
w/o contradiction	64.57	99.88
w/o intra sentence	80.10	92.25
w/o contradiction & intra sentence	64.68	32.75
DualCSE-Bi-RoBERTa	80.38	**100**
w/o contradiction	66.13	99.95
w/o intra sentence	**80.57**	60.35
w/o contradiction & intra sentence	65.07	76.15