

ニコニコ大百科を用いた単語分散表現の学習と その弹幕コメント分類への応用

長尾 浩良[†] 田村宏士郎^{††} 桂井麻里衣[†]

[†] 同志社大学理工学部 〒 610-0394 京田辺市多々羅都谷 1-3

^{††} 同志社大学大学院理工学研究科 〒 610-0394 京田辺市多々羅都谷 1-3

E-mail: †{nagao21,koshiro.tamura,katsurai}@mm.doshisha.ac.jp

あらまし ニコニコ動画などの動画共有プラットフォームにおいて、弹幕コメントが普及している。しかし、動画の内容に無関係なコメントの書き込みによって、動画から得られる情報の質を低下させてしまうことがある。この問題に対し、本研究では単語分散表現に基づくコメントカテゴリ分類手法を提案する。まず、対象ドメインの大規模テキストデータであるニコニコ大百科を用いて単語分散表現を獲得する。次に、棄権オプションを持つ損失関数を用いることで、事前に定義されたカテゴリに該当しないコメントの検出を可能にする。実際の動画データを用いた実験では、ベースライン手法に比べて高い分類性能を発揮した。

キーワード クラス分類, 短文テキスト, 弹幕, BERT, ニコニコ動画

1 はじめに

ユーザー数の増加とともに影響力を拡大しているソーシャルメディアサービスによって、情報発信のあり方が大きく変化した。YouTube¹やニコニコ動画²などの動画共有プラットフォームでは、動画投稿やライブ配信など、不特定多数のユーザーに対する情報共有が可能である。これらのサイトでは、コメント機能を通じ、動画に対して様々な意見・感想をユーザーが書き込むことで投稿者・視聴者間のコミュニケーションが生まれる。さらに、コメントのやり取りにより、視聴者同士も自由に会話を楽しむことができる。コメント表示には様々な方法があり、例えばニコニコ動画やBilibili³などのサービスには、図1に示すような、動画の再生画面上を埋め尽くすようにリアルタイムにコメントを流す機能が搭載されている。この弹幕コメントと呼ばれる機能によって、ユーザーは動画コンテンツのみならず、他のユーザーとの一体感を楽しめる。

弹幕コメントはユーザー間コミュニケーションの促進や視聴体験の向上をもたらすが、一方で動画の内容に無関係なコメントや攻撃的なコメント、ユーザー間の争いも頻繁に発生する。これらは動画から得られる情報の質を低下させてしまい、結果としてユーザーの視聴体験を損なってしまう場合がある[1]。動画コメントのフィルタリングに関する従来研究では、文の感情極性を判定し、ネガティブなコメントを排除する手法などが存在する[2]。この手法では、差別や誹謗中傷を行うコメントを減らすことは可能でも、動画の内容に無関係なコメントをフィルタリングできないという問題がある。

そこで本稿では、弹幕コメントのカテゴリ分類手法を提案す

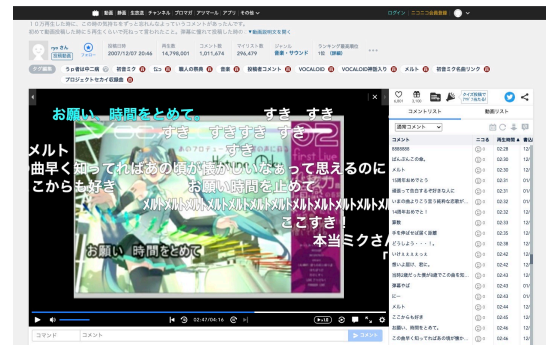


図1 ニコニコ動画における弹幕コメントの例。

る。動画コメントの中には、カテゴリを定めることが困難、あるいは曖昧なものが存在する。我々の直近の研究[3]では、それらの分類を棄権する枠組みを分類モデル学習時の損失関数に導入することで、判定を回避する方法を提案した。この分類モデルでは、日本語 Wikipedia で事前訓練したテキストエンコーダをコメント特徴抽出に用いていた。本稿では、より強力な意味特徴をとらえるために、対象ドメインの大規模なテキストデータから分散表現を獲得する。具体的には、ニコニコ動画特有の表現を多く含むニコニコ大百科データ⁴を用いてテキストエンコーダを事前学習し、動画コメント分類タスクで使用する。実際の動画コメントデータセットを用いた実験では、「政治」、「料理」、「動物」、「鉄道」、「スポーツ」、「ゲーム」の6クラスカテゴリ分類において、Wikipedia に基づく従来手法と比較して高い性能を発揮した。

本稿の構成は以下の通りである。2章では、大規模なテキストデータによる単語分散表現の学習および動画コメント分類に関連する従来研究を紹介する。3章では、本研究で使用するデータセットの詳細について述べる。4章では、単語分散表現

1 : <https://www.youtube.com/>

2 : <https://www.nicovideo.jp/>

3 : <https://www.bilibili.com/>

4 : <https://www.nii.ac.jp/dsc/idr/nico/nicopedia-apply.html>

の学習方法とそれに基づくコメント分類手法を提案する。5章では、単語分散表現の可視化およびコメントカテゴリ分類実験により提案手法の有用性を検証する。最後に6章では、まとめと今後の課題について述べる。

2 関連研究

2.1 単語分散表現に関する研究

単語分散表現とは、文章に現れる単語の意味を数値のベクトルとして表現する方法である。この表現によって、単語間の類似度計算などをベクトル空間で行うことが可能となる。古典的には、分布仮説に基づき、文書内の単語共起頻度をカウントする方法がある。この場合、学習用コーパス全体を見る必要があり、巨大なコーパスに対しては計算コストが非常に大きくなってしまおうという問題がある。一方、word2vec [4] に代表されるニューラルネットワークを用いた手法では、逐次的な推論処理により、巨大なコーパスに対しても効率的な計算が可能であり、さらに単語の文脈情報を捉えることができる。近年は、このような推論ベースの手法が主流となっており、特に Transformer [5] の発表以降、そのモデル構造に基づいて数多くの深層学習モデルが提案されている。Bidirectional Encoder Representations from Transformers (BERT) [6] はその代表例であり、一度大規模なコーパスで事前学習し、その後目的のタスクに応じて比較的少量のデータでファインチューニングするという方法をとることにより、種々の自然言語処理タスクで state-of-the-art となる結果を示した。本研究では、このような背景に基づき、学習モデルアーキテクチャとして BERT を採用する。

日本語テキストを用いた BERT の学習例はいくつか存在し、例として柴田ら [7] は日本語版 Wikipedia⁵ を用いて BERT を学習し、構文解析への応用例を示した。また、ホットリンク社 [8] は、自社で構築した大規模な Twitter データセットを用いて BERT を学習し、ツイートの評判分析タスクへの応用例を示した。本研究では、ニコニコ大百科のテキストを用いて BERT を事前学習することで、動画共有サイト特有の意味表現を獲得する。

2.2 動画コメントのフィルタリングに関する研究

動画コンテンツの情報の質を維持する、あるいは動画を視聴する際のユーザの体験をより良いものにするという観点から、動画コメントのフィルタリング手法が研究されてきた。Chen ら [9] は、YouTube のコメントから、TF-IDF 特徴量を抽出し、感情分類器を構築した。また、Bai ら [10] は、BERT のモデルアーキテクチャを利用し、動画の字幕コメントの感情分析を行った。これらの研究で提案された手法に基づき、ネガティブな内容の書き込みが行われた場合に、それらを排除することができる。しかし、無関係なコメントに対しては有効であるとは限らない。そこで本研究ではコメント内容のカテゴリ分類問題を考え、分類が困難なコメントについては判定を棄権すること



図2 ニコニコ大百科の記事ページの例。

で、柔軟なフィルタリング機能を実現する。

3 データセット

本章では、単語分散表現の学習に用いるニコニコ大百科データおよび動画コメント分類に用いるニコニコ動画コメント等データの詳細を説明する。

3.1 ニコニコ大百科データ

本研究で用いるニコニコ大百科データは、ニコニコ大百科⁶ に2008年5月から2014年2月までに投稿された記事全てについての記事ヘッダ、記事本文、それに付随する掲示板のデータから構成される。図2にニコニコ大百科の記事ページの例を示す。記事ヘッダは、記事ID、記事タイトル、記事ヨミ、記事種類、記事作成日時などが記載されたデータである。記事本文データは、記事ID、記事本文、記事更新日時などが記載されたデータである。掲示板データは、記事ID、レス番号、レス投稿日時、レス本文などが記載されたデータである。以上のメタデータは、記事IDによって紐付いている。本稿では、記事本文データおよび掲示板データのテキストそれぞれに前処理を適用し、それらを結合してコーパスを作成した。具体的には、以下の前処理を適用した。

- 共通の前処理：
 - URL を除去。
 - NFKC 正規化。
 - 文頭、文末の空白を除去。
 - 3回以上同じ文字が連続する場合は2文字に置換（例：wwwww → ww）。
- 記事本文に対する前処理：
 - HTML タグを除去。

5 : <https://ja.wikipedia.org/>

6 : <https://dic.nicovideo.jp/>

- 関連項目や類似項目、脚注などのリストを除去。
- 掲示板に対する前処理：
 - 文頭の「>>」などのメンション記号およびメンション先の番号を除去。
 - 前処理後の文字数が 10 文字以下の文章を除去。

記事本文のテキストは HTML ソースコードとして与えられるため、記事本文から HTML タグを除去した。記事本文および掲示板のデータにおいて記事 ID が重複するものを取り除いた。前処理適用後のテキストは読点あるいは改行記号で一文ずつに分割した。

以上の手順により、本研究で用いるコーパスは記事本文が 334 万文、掲示板が約 1018 万文、合計 1352 万文となった。コメント文を多く含んでいるため、口語表現やスラングなどの表現を学習できるほか、記事本文にもインターネット上で用いられる特有の語彙が多く見られる。大規模テキストコーパスの代表例である日本語版 Wikipedia では丁寧な表現が使われることが多く、掲示板機能なども搭載されていないため、口語表現やスラングなどの表現はあまり見られない。このような背景から、ニコニコ動画のコメント分類に用いる単語分散表現の学習データとしては、ニコニコ大百科が適していると考えられる。

3.2 ニコニコ動画コメント等データ

ニコニコ動画コメント等データ⁷は、ニコニコ動画に 2007 年 3 月から 2018 年 11 月までに投稿された動画のメタデータとそれに対するコメントデータから構成される。動画メタデータには、タイトル、タグ、投稿日時などの情報が記載されており、コメントデータには、コメント本文、投稿日時などの情報が記載されている。

本稿では、このデータセットから、2018 年 11 月までの 1 年間に投稿された、「政治」、「料理」、「動物」、「鉄道」、「スポーツ」、「ゲーム」の 6 つのカテゴリに属する動画のコメントデータを 1 万件ずつ、合計 6 万件抽出した。またこれらのテキストデータに対し、以下の前処理を適用した。

- 英単語を小文字に置換。
- 全角アルファベットを半角に置換。
- 記号やピクトグラムを除去。
- URL を除去。
- 文字数が 8 文字以下の文章を除去。
- すべての数字を 0 に置換。
- 3 回以上同じ文字が連続する場合は 2 文字に置換。

4 提案手法

4.1 BERT の事前学習

動画コメント分類に有用な単語分散表現を獲得するため、BERT のモデルアーキテクチャを利用し、事前学習を行う。BERT の事前学習には、Masked Language Model (Masked LM) と Next Sentence Prediction (NSP) という 2 つのタスクが存在する。Masked LM ではまず、文章中の 15% の単語をランダムに

表 1 2 名のアノテータによるラベルが一致したコメントの数。

カテゴリ	コメント数
政治	198
料理	213
動物	101
鉄道	149
スポーツ	193
ゲーム	203
棄権	398

選択し、それを [MASK] などの特殊なトークンに置換する。そして [MASK] に対して、置換される前のトークンが何であったかを予測する。NSP では、連続する 2 つの文章 A と B について、B を 50% の確率でコーパス中のランダムな文章で置換し、A と B が連続であるか否かを予測する。

本稿で用いるニコニコ大百科データセットには、コメント文が多く含まれているという特徴がある。NSP の学習により、文同士の意味の関係を捉えることができるようになるが、コメント文同士には意味の関係性があまり見られない。さらに、BERT の改良モデルである RoBERTa [11] などの論文では、NSP の有効性が明確でないことが指摘されている。そこで本稿では、事前学習タスクとして Masked LM のみを採用する。

4.2 コメント分類への応用

我々は 3.2 節において、あるカテゴリラベルが付与された動画に属する全てのコメントに、その動画のカテゴリラベルを擬似的に付与した。提案手法ではこの疑似ラベルを用いて、コメントの埋め込み表現を入力とした動画カテゴリ分類器を学習することで、両者の関係をモデル化する。

クラス分類タスクでは、損失関数として交差エントロピー損失が広く使用されているが、本稿では、Deep Gamblers (DG) [12] と呼ばれる損失を新たに導入し、予測の信頼性についても考慮できるように学習する。具体的には、動画のカテゴリ数を m としたとき、 m クラス分類問題に「予測の棄権」というクラスを新たに追加することで、次式の損失関数を用いて $m+1$ クラス分類問題として定式化する。

$$\max_w \sum_{x \in B} \log[f_w(x)_{j(x)} o + f_w(x)_{m+1}], \quad (1)$$

ここで、 f はモデルを表し、 x はその入力、 w は f のパラメータ、 $j(x)$ は x の正解クラスラベル、 B はバッチサイズ、 o は棄権の度合いを表すハイパパラメータである。 o は、 $1 < o < m$ の範囲で設定でき、 o を大きくするほど棄権をしやすくなる。本研究では、実験的に o を 3.5 とした。

モデル f は前節で事前学習した BERT モデルに全結合層を追加したものである。具体的には、BERT から出力される 768 次元の埋め込みベクトルを 7 次元のベクトルへ変換したあと、予測クラスを出力する。

提案手法によるコメントフィルタリングにはカバー率という概念を導入する。そして、棄権信頼度の昇順でデータを並べ替え、カバー率に対応する分のデータにカテゴリ分類結果を与え

7 : <https://www.nii.ac.jp/dsc/ldr/nico/nicocomm-apply.html>

表 2 各評価指標での比較結果.

モデル	カバー率	正解率		適合率		再現率		F 値	
		DG	SR	DG	SR	DG	SR	DG	SR
Wikipedia BERT	1.0	0.7067	0.6395	0.6961	0.6823	0.7160	0.5533	0.6995	0.6026
	0.95	0.7254	0.6726	0.7137	0.6823	0.7306	0.5780	0.7163	0.6185
	0.9	0.7468	0.7101	0.7333	0.6823	0.7496	0.6071	0.7364	0.6359
	0.85	0.7664	0.7519	0.7489	0.6823	0.7637	0.6406	0.7516	0.6551
	0.80	0.7920	0.7991	0.7738	0.6823	0.7869	0.6789	0.7769	0.6758
	0.75	0.8197	0.8235	0.8026	0.8104	0.8096	0.8107	0.8022	0.8045
	0.70	0.8351	0.8432	0.8202	0.8318	0.8216	0.8265	0.8170	0.8219
ニコニコ大百科 BERT	1.0	0.8325	0.7928	0.8242	0.7142	0.8358	0.6834	0.8236	0.6930
	0.95	0.8517	0.8338	0.8427	0.7142	0.8503	0.7139	0.8416	0.7102
	0.90	0.8729	0.8613	0.8638	0.8526	0.8699	0.8562	0.8625	0.8504
	0.85	0.8910	0.8810	0.8819	0.8717	0.8853	0.8713	0.8802	0.8689
	0.80	0.9066	0.8983	0.8969	0.8871	0.8996	0.8863	0.8958	0.8845
	0.75	0.9180	0.9193	0.9036	0.9103	0.9094	0.9073	0.9045	0.9061
	0.70	0.9230	0.9324	0.9087	0.9225	0.9156	0.9191	0.9103	0.9185

る。例えばテストデータに対する分類実験においてカバー率が 0.7 である場合、棄権信頼度が低い上位 70% のデータの分類結果のみで性能評価指標を算出する。

5 実 験

5.1 実験設定

BERT の事前学習には、HuggingFace の実装⁸を利用した。コーパス中の文章の分かち書きには SentencePiece [13] を採用し、語彙数は 32,000 とした。1 ノードあたり 8 つの GPU (メモリ計 360GB) を、4 つのノードで並行して学習させた。学習ステップ数を 1,200,000、エポック数に換算して約 23 epoch で学習させると、合計で 4 日程度を要した。

また、ニコニコ動画のコメント 60,000 件を、48,000 件の訓練データと 12,000 件の評価候補データに分割した。カテゴリ分類タスクでの最適化には Adam [14] を用い、バッチサイズは 32、エポック数は 12 とした。学習を安定させるため、Liu ら [12] の研究に従って、最初の 3 epoch は交差エントロピー損失、残りの 9 epoch はジャンプラー損失を使用した。

コメント分類性能の定量評価のために、評価候補データから各動画カテゴリに属するコメントを 350 件ずつランダムに抽出した。動画カテゴリという疑似ラベルを再評価するため、各コメントに対し 2 名のアノテータが手動でクラスラベルを付与した。2,100 件のコメントに対するアノテータ間のラベル一致率は 69.3% であった。カテゴリごとの一致コメント数を表 1 に示す。最終的に、2 名の付与ラベルが一致したデータをカテゴリ分類評価データとした。

5.2 コメントカテゴリ分類の評価

提案手法は棄却オプションのために DG を損失関数に用いたが、同様の目的の手法である Softmax Response (SR) [15] についても性能を報告する。SR では、モデル出力の Softmax 値の

最大値があらかじめ定めた閾値以下ならば、判定を棄権する。

学習コーパスの違いによる性能検証のため、Wikipedia による BERT の事前学習済みモデル⁹に基づくカテゴリ分類器をベースラインとして用いる。また、分類性能の評価指標として、正解率、適合率、再現率、F 値をカバー率を考慮して算出する。各評価指標におけるベースライン手法と提案手法の比較結果を表 2 に示す。なお DG、SR はそれぞれコメント棄却に用いた手法を表す。表 2 より、異なるコーパスのモデルで比較した場合、すべての評価指標とカバー率で、ニコニコ大百科で学習させたモデルが高い数値を示した。ゆえに、対象タスクのドメインと親和性の高いコーパスで言語モデルを学習することは有効といえる。

5.3 二次元空間における単語分散表現の可視化

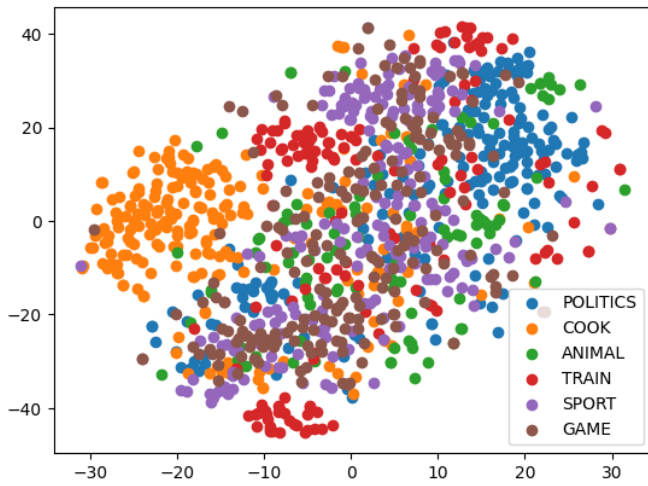
2 名のアノテータによるカテゴリラベル付きコメントデータを埋め込み対象とし、日本語版 Wikipedia およびニコニコ大百科で学習した分散表現の可視化結果を図 3 に示す。ノードの色は各カテゴリに対応する。これらの図より、ニコニコ大百科による単語分散表現の方が動画カテゴリによるクラスタが形成されやすいということがわかった。この結果からも、ニコニコ大百科で BERT を事前学習することが、弹幕コメントのカテゴリ分類器を学習するうえで有効であったといえる。

6 ま と め

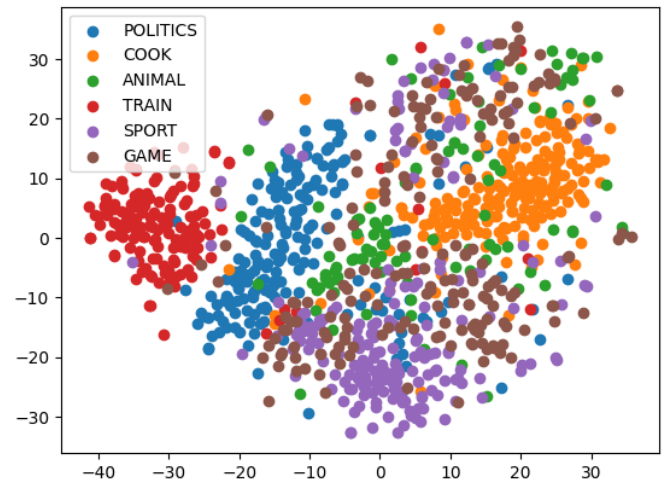
本研究では、ニコニコ大百科を用いた単語分散表現の学習および動画の弹幕コメント分類手法を提案した。提案手法では、予測の棄権が可能な損失関数を導入し、動画の内容に無関係なコメントを検出可能な分類モデルとして定式化した。実際の弹幕コメントデータを用いた実験では、日本語版 Wikipedia で獲得した単語分散表現と比較して、ニコニコ大百科を用いた単語分散表現がコメントカテゴリ分類タスクに適していることがわ

⁸ : <https://github.com/huggingface/transformers>

⁹ : <https://github.com/yoheikikuta/bert-japanese>



(a) 日本語版 Wikipedia による結果。



(b) ニコニコ大百科による結果。

図3 各コーパスで学習した単語分散表現の可視化。

かった。

大規模データによる単語分散表現の学習は上記のような有効な結果をもたらしたが、計算コストが非常に大きいという問題がある。これに対し例えば、ELECTRA [16] などの、大規模なデータを効率的に学習可能な最新の BERT の改良モデルを取り入れることで、この問題に対処できると考えられる。また、今後、本研究で提案した分類手法を用いて、実際にユーザが利用可能なフィルタリングシステムを開発する予定である。

7 謝 辞

本研究の一部は、JSPS 科研費 (20H04484) の助成によって行われた。本研究では、国立情報学研究所の情報学研究データリポジトリを通じて株式会社ドワンゴより提供されている「ニコニコ大百科データ」および「ニコニコ動画コメント等データ」を利用した。

文 献

- [1] Yue Chen, Qin Gao, and Pei-Luen Patrick Rau. Watching a movie alone yet together: Understanding reasons for watching danmaku videos. *International Journal of Human-Computer Interaction*, Vol. 33, No. 9, pp. 731–743, 2017.
- [2] Julio Savigny and Ayu Purwarianti. Emotion classification on youtube comments using word embedding. In *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, pp. 1–5. IEEE, 2017.
- [3] Koshiro Tamura and Marie Katsurai. Selective classification of danmaku comments using distributed representations. In *The 23rd International Conference on Information Integration and Web Intelligence (iiWAS)*, pp. 130–136, 2021.
- [4] Tomáš Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *International Conference on Learning Representations (ICLR)*, 2013.
- [5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [7] 柴田知秀, 河原大輔, 黒橋禎夫. BERT による日本語構文解析の精度向上. 言語処理学会 第 25 回年次大会, pp. 205–208, 2019.
- [8] Takeshi Sakaki, Sakae Mizuki, and Naoyuki Gunji. BERT pre-trained model trained on large-scale japanese social media corpus. <https://github.com/hottolink/hottoSNS-bert>, 2019. Last accessed: 14/02/2022.
- [9] Yen-Liang Chen, Chia-Ling Chang, and Chin-Sheng Yeh. Emotion classification of YouTube videos. *Decision Support Systems*, Vol. 101, pp. 40–50, 2017.
- [10] Qingchun Bai, Kai Wei, Jie Zhou, Chao Xiong, Yuanbin Wu, Xin Lin, and Liang He. Entity-level sentiment prediction in Danmaku video interaction. *The Journal of Supercomputing*, pp. 1–20, 2021.
- [11] 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.
- [12] Ziyin Liu, Zhikang Wang, Paul Pu Liang, Russ R Salakhutdinov, Louis-Philippe Morency, and Masahito Ueda. Deep Gamblers: Learning to abstain with portfolio theory. *Advances in Neural Information Processing Systems*, Vol. 32, pp. 10623–10633, 2019.
- [13] Taku Kudo and John Richardson. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv preprint arXiv:1808.06226*, 2018.
- [14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.
- [15] Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 4885–4894, 2017.
- [16] Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. ELECTRA: Pre-training text encoders as discriminators rather than generators. In *International Conference on Learning Representations (ICLR)*, 2020.