氏　　名	NOR SAIFUL AZAM BIN NOR AZMI
授 与 学 位	博士（工学）
学 位 記 番 号	博甲第238号
学位授与年月日	令和8年9月8日
学位授与の要件	学位規則第4条第1項
学位論文題目	Multifaceted Feature Fusion in Transformer Models with Application to Hateful Content Detection （トランスフォーマーモデルにおける多面的な特徴融合と有害コンテンツ検出への応用）
論文審査委員	主査　教　授　プタシンスキ　ミハウ　エドムンド 教　授　桝井　文人 教　授　前田　康成 教　授　升井　洋志 准教授　エロネン　ユーソ　カレビ　クリスティアン

学位論文内容の要旨

Transformer models, such as ELECTRA, RoBERTa, or Vision Transformer, have established state-of-the-art performance across natural language processing and computer vision domains. However, as these architectures scale in parameters, the literature exposes a fundamental limitation: they function primarily as sophisticated statistical pattern matchers relying on distributional semantic rather than comprehending explicit structural rules. Consequently, they inherently lack deep structural and contextual representation. At the input level, standard sub-word tokenizers optimize for vocabulary size but fragment morphology, degrading the grammatical data required to parse implicit meaning. At the output level, unimodal classifiers processing multi-layered data struggle to integrate diverse feature sets. They tend to over-rely on superficial signals from a single modality, ignoring the broader context and rendering them highly vulnerable to complex information.

To rigorously resolve these vulnerabilities, this research evaluates these architecture within the domain of hateful content detection, encompassing textbased cyberbullying and multimodal internet memes. Hateful content serves as the reasoning stress test, actively exploiting model limitation through ambiguity, sarcasm, and misleading contextual pairings such as where individual elements are harmless in isolation but highly toxic when combined. Three specific dataset hasbeen chosen to benchmark the model, an English cyberbullying dataset, The Facebook Hateful meme dataset, and the Innopolis Hateful Meme dataset. This thesis proposes a multifaceted feature fusion framework. This framework introduces advanced representational strategies at opposite ends of the processing pipeline: early fusion token modification at the input level and late-fusion integration at the decision level.

Phase 1 establishes an Early Fusion methodology designed to enforce structural and syntactic awareness in text-based cyberbullying detection. Traditional Part-of-Speech (POS) tagging encounters critical preprocessing issues, as standard tokenizers like WordPiece severely fragment appended tags and fail to properly segment tokens and grammatical labels within vocabulary files. To resolve this, a novel Symbolic Substitution method was developed, mapping POS tags to unique Greek character (e.g., verbs to $¥delta$, nouns to $¥mu$) fused directly to the word stem. Utilizing the SentencePiece Unigram tokenizer and pretraining an ELECTRA model from scratch on 32 gigabytes of customized data, tokens and grammatical functions were processed as unified embedding. This forces the model to process lexical meaning and structural rules simultaneously, yielding consistent improvement over ELECTRA baselines and increasing toxic-class recall from 0.5878 to 0.6209.

Phase 2 expands the framework to integrate contextual awareness in multimodal spaces using a Late Fusion architecture, RoBERViT, evaluated on the Facebook Hateful Meme and Innopolis Hateful Meme datasets. RoBERViT operates via two specialized streams: a ViT capturing global visual dependencies and a RoBERTa model processing textual semantics. To bridge the semantic gap, the textual stream is enriched using MoonDream, a generative vision-language model, which

appends detailed image caption to the native meme text. Independent feature vectors from both stream are concatenated at a dense classification head, enforcing joint decision-making. This late-stage integration mitigates “lexical dominance” and resolves “benign confounders” by ensuring textual predictions rely on corroborating visual evidence. On the highly imbalance Innopolis Hateful Meme dataset, RoBERViT-Large achieved a toxic-class F1-score of 0.6439, significantly outperforming the best text-only baseline of 0.5794.

To comprehensively evaluate decision-making processes, an Explainable AI (XAI) analysis utilized Local Interpretable Model-agnostic Explanation (LIME) alongside generative reasoning via LLaVA. This qualitative audit revealed critical failure; notably, baseline vision models routinely failed to parse visual intent, as focus patches were misaligned with primary subject in an image, instead utilizing the visual presence of text blocks as a shallow proxy for hate. Furthermore, T5-based paraphrasing disrupted the exact keywords standard models relied upon, resulting in severe performance degradation and proving that unimodal system suffers from intense lexical over-reliance.

Ultimately, scaling parameters along is insufficient for deep semantic comprehension. While the proposed multifaced framework demonstrates that structurally embedding data at the syntactic root and enforcing integration at the multimodal endpoint mitigates statistical shortcuts, fundamental challenges remain. This research serves as a foundational steppingstone toward true contextual understanding, highlighting that future advancement must refine these representational interventions to consistently and accurately detect complex hateful content.

審査結果の要旨

本学位論文では，モンゴル伝統文様の保存および再現を目的として，数学的モデリングに基づくデジタル化手法を考案した．まず，代表的な文様を幾何学的文様，植物文様，動物文様，象徴文様および自然文様の 5 種類に分類し，それらの構成，対称性および幾何学的複雑性を解析した．次に，円，円弧，対数らせん，アステロイドおよび直線等の基本的な幾何要素を，平行移動，回転，鏡映ならびに拡大・縮小等の幾何変換によって組み合わせることにより，複雑な文様を数学的に再構成した．考案した手法の有効性を検証するため，モンゴルの伝統的な嗅ぎたばこ入れ用ポーチ（ダーリン）の文様およびハート形文様を対象として，二つのケーススタディを実施した．再構成した文様をベクトル形式に変換した後，スクリプト言語に基づく CAD システムを用いて CAD モデルを生成し，FFF 方式の 3D プリンタにより造形した．比較検証の結果，考案した手法は，原画像のベクトルグラフィックスへの直接変換および手作業によるトレースと比較して，幾何形状の安定性，曲線接続の滑らかさおよび形状パラメータの制御性に優れていることが確認された．これにより，伝統工芸品に含まれる文様の分類，幾何要素への分解，数学的再構成，ベクトル化，CAD モデル化および 3D 造形に至る一連のデジタル・フィジカル化プロセスを構築した．考案した手法は，モンゴル伝統文様の保存，デジタルアーカイブ化，教育および博物館展示に寄与するとともに，文化的意匠を活用した製品や土産品の設計・製造にも応用可能である．さらに，熟練職人が減少する中，伝統的知識を継承し，文化的価値を有する文様を高精度かつ効率的に再現するための実用的な手段となることが期待される．これらの研究成果は，Mathematical and Computational Applications をはじめとする国際的に評価の高い学術誌ならびに国内外の学術会議において発表されている． よって，提出された学位論文の内容および該当する学術論文の質等を総合的に考慮した結果，申請者は北見工業大学博士（工学）の学位を授与される資格を有する者であると認める．