

ライトノベルにおける位置情報・意味情報・表紙要素に基づく 複合ネタバレ度指標に対する有効性評価

Effectiveness of a Composite Spoiler Metric Based on Positional, Semantic, and Cover Elements in Light Novels

井上大輝¹ 尾崎知伸^{1*}
Daiki Inoue¹ Tomonobu Ozaki¹

¹ 日本大学
¹ Nihon University

Abstract: This study evaluates the effectiveness of a composite metric focusing on three aspects of positional information, semantic information, and cover elements for the purpose of spoiler detection in light novels. Specifically, we used a manually annotated corpus with spoiler intensity as ground truth data to quantitatively analyze the relationship between each aspect and the spoiler degree. Furthermore, through application to summary generation by generative AI, we verified the utility of the composite metric in practical applications.

1 はじめに

ライトノベルは、個性的なキャラクターやテンポの良いストーリー展開を特徴とする物語小説の一形態であり、読者に深い没入感を伴う読書体験を提供している。その魅力の核心は、キャラクター間の関係性の変容や伏線の回収といったドラマチックな展開にあり、些細な描写が物語全体の方向性に重大な意味を持つことも少なくない。このような特性から、ライトノベルにおけるストーリー情報の取り扱いには細心の注意を要する。未読の読者がネタバレを含む情報に接触した場合、著者が本来意図していた驚きや感情的なインパクトが損なわれる懸念があるためである。その一方で、個々の読者の読書進捗に合わせ、物語の魅力を損なわずに要約や紹介文を生成する技術への需要は高い。ネタバレを適切に制御しつつ物語の核心を伝える要約生成システムが実現すれば、より効果的な情報共有や作品のプロモーションが期待できる。

これらのことを背景に、これまでネタバレ検出に関する研究が精力的に行われている。例えば Wan ら [1] は、書評データセットを用いた文単位のネタバレ検出タスクを対象に、ネタバレの傾向・性質を分析した上で、ユーザと作品のバイアスを考慮した階層的注意ネットワークに基づくモデルを提案した。また Chang ら [2] は、グラフとして表現される文の係り受け関係に GNN を適用することで、文脈に応じた意味識別を可能にする

手法を開発した。一方、前田ら [3] は、レビュー内の記述が本文のどの部分に対応するかを推定することで、ネタバレの可能性を評価する枠組みを示した。さらに Tran ら [4] は、ネタバレ検出をレビューと本文との意味的対応問題として再定義している。

本研究では、ライトノベルのネタバレを対象とした、位置情報・意味情報・表紙要素の3側面に着目した複合指標 [5] に着目する。また、この指標に基づく実応用に向けての初手として、指標が持つ性質・特徴を明らかにすることを試みる。具体的には、人手によりアノテーションを行ったネタバレ度付きコーパスを正解データとし、各側面とネタバレ度との関連を定量的に分析する。さらに、生成 AI による要約生成への適用を通じ、実応用における有用性を検証する。

本論文の構成は以下のとおりである。2章で対象とする複合指標を説明する。3章で指標に対する評価実験を行う。4章でまとめを行い、今後の課題を述べる。

2 複合ネタバレ度指標

本研究で対象とする複合ネタバレ度指標 [5] は、小説本文を固定長の文書(文書区間)に分け、位置情報・意味情報・表紙要素の3側面から各スコアを算出・総合することで最終的なネタバレ度を決定する。すなわち、小説 n における第 i 番目の文書区間 $c_i^{(n)}$ に対して位置スコア S_{pos} 、意味スコア S_{sem} 、表紙スコア S_{cov} を算

*連絡先: 日本大学 文理学部 情報科学科
〒156-8550 東京都 世田谷区 桜上水 3-25-40
E-mail: ozaki.tomonobu@nihon-u.ac.jp

出し, 重み w_{pos} , w_{sem} , w_{cov} による加重和である

$$S(c_i^{(n)}) = \frac{w_{\text{pos}}S_{\text{pos}} + w_{\text{sem}}S_{\text{sem}} + w_{\text{cov}}S_{\text{cov}}}{w_{\text{pos}} + w_{\text{sem}} + w_{\text{cov}}}$$

を, $c_i^{(n)}$ に対する複合ネタバレ度とする.

2.1 位置スコア

位置スコアは, 物語が進行するにつれてネタバレがより顕著になるという仮定に基づく. 小説 n における文書区間総数を $B^{(n)}$ としたとき, $B^{(n)} > 1$ を前提に, $c_i^{(n)}$ に対する位置スコア S_{pos} は次のように定義される.

$$S_{\text{pos}}(c_i^{(n)}) = i / (B^{(n)} - 1)$$

2.2 意味スコア

意味スコア S_{sem} は, w_{smy} と w_{tpl} を重みとする, 要約スコア S_{smy} とトリプルスコア S_{tpl} の加重和である.

$$S_{\text{sem}}(c_i^{(n)}) = \frac{w_{\text{smy}}S_{\text{smy}}(c_i^{(n)}) + w_{\text{tpl}}S_{\text{tpl}}(c_i^{(n)})}{w_{\text{smy}} + w_{\text{tpl}}}$$

要約スコア S_{smy} は, GPT-4o により推定される, 前後の文脈を考慮したネタバレ強度を表すスコアである. 以下のように定義され, その値域は $[0.0, 1.0]$ である.

$$S_{\text{smy}}(c_i^{(n)}) = \text{LLM} \left(D \left(\{c_j^{(n)}\}_{j=0}^{i-1}, c_i^{(n)}, D \left(\{c_k^{(n)}\}_{k=i+1}^{B^{(n)}-1} \right) \right) \right)$$

ここで LLM は, GPT-4o によるスコア推定を, また D は, GPT-4o mini によって得られる連続文書区間の要約テキストをそれぞれ表す.

一方, トリプルスコア S_{tpl} は, $c_i^{(n)}$ に含まれるトリプルの全体集合 \mathcal{T}_i^n から導出されるスコアであり, 以下のように定義される.

$$S_{\text{tpl}}(c_i^{(n)}) = \frac{1}{|\text{Top}_K(\mathcal{T}_i^n)|} \sum_{t' \in \text{Top}_K(\mathcal{T}_i^n)} S_{\text{tmp}}(t')$$

ここで Top_K は, \mathcal{T}_i^n に含まれる, 時間スコア S_{tmp} 上位 K 件からなる集合である.

2.3 時間スコア

トリプル t に対する時間スコア S_{tmp} は, w_{pst} と w_{fut} を重みとする, 過去スコア S_{pst} と未来スコア S_{fut} の加重和である.

$$S_{\text{tmp}}(t) = \frac{w_{\text{pst}}S_{\text{pst}}(t) + w_{\text{fut}}S_{\text{fut}}(t)}{w_{\text{pst}} + w_{\text{fut}}}$$

過去スコア トリプル $t \in \mathcal{T}_i^n$ に対する過去スコア S_{pst} は, t 単体のネタバレ危険性を表す内容スコア $S_{\text{cnt}}(t)$ と, 先行するトリプル集合 $\mathcal{PT}_i^n = \bigcup_{j=0}^{i-1} \mathcal{T}_j^n$ に対する新規性スコア S_{nov} との積, すなわち

$$S_{\text{pst}}(t) = S_{\text{cnt}}(t) \times S_{\text{nov}}(t)$$

と定義される. ここで内容スコア $S_{\text{cnt}}(t)$ は, 8種のネタバレを対象とした事前学習済み BERT モデルによる予測値 p_y^t ($y \in \{1, \dots, 8\}$) と閾値 θ_c を用い,

$$S_{\text{cnt}}(t) = 1 - \prod_{y \in \{1, \dots, 8\} \wedge p_y^t \geq \theta_c} (1 - p_y^t)$$

と定義される. 一方, 新規性スコア S_{nov} は, ベクトル間類似度を基礎とし, パラメタ θ_n と α_n を用いて

$$S_{\text{nov}}(t) = \sigma \left(\left(\theta_n - \max_{t' \in \mathcal{PT}_i^n} \cos(\text{enc}(t), \text{enc}(t')) \right) \times \alpha_n \right)$$

と定義される. なお, σ はシグモイド関数, enc はエンコーダによるトリプルのベクトル表現, \cos はベクトル間コサイン類似度である.

未来スコア トリプル $t \in \mathcal{T}_i^n$ に対する未来スコア S_{fut} は, 後続するトリプル集合 $\mathcal{FT}_i^n = \bigcup_{j=i+1}^{B^{(n)}-1} \mathcal{T}_j^n$ との平均類似度 avg と, 類似トリプル集合 sim_triples を用い

$$S_{\text{fut}}(t) = \text{avg}(t) \times \frac{\log(1 + |\text{sim_triples}(t)|)}{\log(1 + |\mathcal{FT}_i^n|)} \times \alpha_f$$

と定義される. なお α_f はパラメタである. また, パラメタ θ_f を用い, 平均類似度と類似トリプル集合は, それぞれ以下のように定義される.

$$\text{avg}(t) = \frac{1}{|\mathcal{FT}_i^n|} \sum_{t' \in \mathcal{FT}_i^n} \cos(\text{enc}(t), \text{enc}(t'))$$

$$\text{sim_triples}(t) = \{t' \in \mathcal{FT}_i^n \mid \cos(\text{enc}(t), \text{enc}(t')) \geq \theta_f\}$$

2.4 表紙スコア

表紙スコア S_{cov} は, 表紙単語群 (表紙に関する単語集合) \mathcal{C}^n を基準とする, 制御パラメタ α_s および閾値 θ_s を伴う飽和加重和であり,

$$S_{\text{cov}}(c_i^{(n)}) = 1 - \prod_{cw \in \mathcal{C}^n \wedge \cos(\text{enc}(c_i^{(n)}), \text{enc}(cw)) \geq \theta_s} \left(1 - \cos(\text{enc}(c_i^{(n)}), \text{enc}(cw)) \right)^{\alpha_s}$$

と定義される. なお \mathcal{C}^n の抽出は, 表紙イラストとタイトルを対象に, GPT-4o/4o-mini によって行われる.

3 評価実験

複合ネタバレ度指標の有効性と実用性を検証するため, 以下の3つの実験を行う¹. なお正解データは, 実

¹パラメタ設定: $w_{\text{smy}} = w_{\text{tpl}} = 1$, $K = 5$, $w_{\text{pst}} = w_{\text{fut}} = 1$, $\theta_c = 0.5$, $\theta_n = 0.8$, $\alpha_n = 1.2$, $\alpha_f = 1.0$, $\theta_f = 0.8$, $\alpha_s = 1.5$, $\theta_s = 0.5$

表 1: スコア間のピアソン相関行列

	正解	S_{pos}	S_{sem}	S_{cov}
S_{pos}	0.268	—	0.219	-0.157
S_{sem}	0.300	0.219	—	0.110
S_{cov}	0.147	-0.157	0.110	—

表 2: 最適重みパラメタ

	既知情報			重要情報		
	w_{pos}	w_{sem}	w_{cov}	w_{pos}	w_{sem}	w_{cov}
O_1	0.000	0.965	0.035	0.000	1.000	0.000
O_2	0.114	0.886	0.000	0.141	0.766	0.093
O_3	0.112	0.888	0.000	0.138	0.760	0.102

験協力者 6 名がそれぞれライトノベル 1 冊 (平均文字数約 13 万 ±3,000 字) を精読し, 各文書区間に対して 6 段階 (0 ~ 5) でネタバレ度を付与することで構築した。

実験 1: 位置, 意味, 表紙の各スコアが, 人間が感じるネタバレ度と整合しているかを検証する。

実験 2: 線形計画法を用い, 各スコアの最適な統合方法 (重み) を決定する。

実験 3: あらすじ生成タスクに複合ネタバレ度指標を適用し, 生成物の品質や安全性を評価する。

3.1 スコアに対する相関分析

各ネタバレ度スコアと正解データとの相関係数を表 1 に示す。得られた結果に基づき, 以下の 3 点について考察する。

第一に, すべてのスコアが正解データに対して正の相関を示した。特に意味スコア S_{sem} は 0.300 と最も高い相関を示しており, テキストの意味情報がネタバレ判定に強く関係していることが確認できる。

第二に, 表紙要素の役割についてである。表紙スコア S_{cov} と正解データとの相関係数は 0.147 と弱い。この結果は, 表紙イラストは読者にとって既知の情報でありネタバレには該当しない, という既知情報としての性質と, 表紙イラストはクライマックスの場面や重要人物を描写する傾向が強く, 重要なネタバレを含んでいる, という重要情報としての性質とが混在していることを示唆しており, 互いに相殺しあった結果として相関が現れなかったものと推察される。

第三に, スコア間の独立性である。特に S_{pos} と S_{cov} の相関係数は -0.157 と負の値を示した。これは, 表紙情報が物語の冒頭などの導入部に出現しやすいという直観的な傾向と一致する。このように, 互いに異なる時間的内容の傾向を持つ各指標を適切に組み合わせることで, 単一の指標では捉えきれないネタバレ箇所を相補的に検出できる可能性が高い。

3.2 統合重みの最適化

各スコアのネタバレに対する貢献を検証するため, 線形計画法を用い, 最適な統合重み ($w_{\text{pos}}, w_{\text{sem}}, w_{\text{cov}}$) を

導出する。なお表紙スコア S_{cov} に関しては, 既知情報・重要情報の両性質が考えられるため, 新たにスコア

$$\tilde{S}_{\text{cov}}(c_i^{(n)}) = \begin{cases} 1 - S_{\text{cov}}(c_i^{(n)}) & (\text{表紙を既知情報と見做す}) \\ S_{\text{cov}}(c_i^{(n)}) & (\text{表紙を重要情報と見做す}) \end{cases}$$

を導入し, 複合ネタバレ度を

$$S(c_i^{(n)}) = \frac{w_{\text{pos}}S_{\text{pos}} + w_{\text{sem}}S_{\text{sem}} + w_{\text{cov}}\tilde{S}_{\text{cov}}}{w_{\text{pos}} + w_{\text{sem}} + w_{\text{cov}}}$$

と再定義した上で最適化を行う。

具体的には, 文書の全体集合を N , 文書区間 $c_i^{(n)}$ に対する正解データを $G(c_i^{(n)})$ とし, 制約

$$w_{\text{pos}} \geq 0, w_{\text{sem}} \geq 0, w_{\text{cov}} \geq 0, w_{\text{pos}} + w_{\text{sem}} + w_{\text{cov}} = 1$$

のもとで, 以下の 3 つの目的関数をそれぞれ最小化する。またその際, 表紙を既知・重要などちらの情報と見做すのかを定め, 両者を比較する。

目的関数 O_1 : 文書区間の絶対誤差を直接最小化する。

$$\sum_{n \in N, 0 \leq i \leq B^{(n)}-1} |S(c_i^{(n)}) - G(c_i^{(n)})|$$

目的関数 O_2 : 全文書区間ペア $(c_i^{(n)}, c_m^{(m)})$ について, S の差と G の差の絶対誤差を最小化する。

$$\sum_{n \in N, 0 \leq i \leq B^{(n)}-1} \sum_{m \in N, 0 \leq j \leq B^{(m)}-1} \left| (S(c_i^{(n)}) - S(c_j^{(m)})) - (G(c_i^{(n)}) - G(c_j^{(m)})) \right|$$

目的関数 O_3 : 同一文書内における文書区間ペアについて, S の差と G の差の絶対誤差を最小化する。

$$\sum_{n \in N} \sum_{0 \leq i < j \leq B^{(n)}-1} \left| (S(c_i^{(n)}) - S(c_j^{(n)})) - (G(c_i^{(n)}) - G(c_j^{(n)})) \right|$$

得られた重みパラメタを表 2 にまとめる。結果から, 以下の 2 つの知見が伺える。

第一に, 単純な絶対誤差に基づく最適化における限界である。目的関数 O_1 では, 表紙スコア \tilde{S}_{cov} の設定 (既知 / 重要) に関わらず, 位置スコアの重み w_{pos} が 0.000 となり, 位置情報が評価に寄与しない結果となっ

た．特に設定（重要）では，意味スコアの重み w_{sem} が 1.000 に収束し，モデルが単一の指標と等価的になってしまっている．これは，スコアの絶対誤差を直接最小化する手法では，正解データと最も高い相関を持つ意味スコアのみ最適化が集中してしまい，物語の進行に伴う相対的な変化を示す位置情報の特性を十分に反映できないことを示している．

第二に，表紙スコアの設定による有効性の差異である． O_2 および O_3 において表紙要素を「既知情報」として扱った場合， w_{pos} は約 0.11 まで回復したものの， w_{cov} は 0.000 となり，表紙スコアは再度無効化された．これに対し，表紙要素を「重要情報」として扱った場合， w_{cov} に約 0.09 から 0.10 の有意な重みが割り当てられた．この結果は，表紙をネタバレを増進させる要素として定義することが，多角的なネタバレ度評価モデルを構築する上で不可欠であることを裏付けている．

3.3 要約生成

複合ネタバレ度指標の実用性を検証するため，大規模言語モデル（GPT-5）を用いたネタバレ配慮型あらすじ生成を題材とした評価実験を行った．具体的には，前節の実験結果に従い，設定 $\tilde{S}_{cov}(c_i^{(n)}) = S_{cov}(c_i^{(n)})$ ， $w_{pos} = 0.138$ ， $w_{sem} = 0.760$ ， $w_{cov} = 0.102$ を用い，以下の各方法で GPT-5 にあらすじを生成させ，それらを比較する．

方法 M_1 ： 小説全文を入力し，プロンプトでの言語的な指示に依存して生成する方法．ネタバレ指標は利用しない．

方法 M_2 ： 各文書区間の先頭に，メタデータとしてネタバレ度スコアを付与した上で，小説全文を入力する方法．プロンプトにおいて，高スコアの箇所は物語の核心であるため，あらすじには使用しないよう指示を与える．

方法 M_3 ： 予めネタバレ度スコアが低い文書区間を抽出・排除し，総文字数が約 3 万文字になるように調整したテキストのみを入力する方法．

あらすじに対する評価は，3 名の被験者によるブラインドテスト形式で実施した．各被験者に対し，各方法でそれぞれ 2 つずつ生成した 6 件のあらすじ M_i^j ($1 \leq i \leq 3, j \in \{a, b\}$) と，出版社による公式あらすじ P の計 7 つを提示し，ネタバレのなさ（安全性）と紹介文としての魅力（品質）の観点から総合的に優れていると感じた順に順位付けを求めた．

結果を表 3 に示す．得られた順位傾向より，以下の 2 つの知見が確認できる．

表 3: あらすじに対するランキング

評価者	1 位	2 位	3 位	4 位	5 位	6 位	7 位
A	P	M_2^b	M_3^b	M_2^a	M_1^a	M_3^a	M_1^b
B	P	M_2^b	M_2^a	M_1^b	M_1^a	M_3^b	M_3^a
C	M_1^a	M_2^a	P	M_2^b	M_1^b	M_3^b	M_3^a

第一に，機械的なフィルタリングに基づく方法 M_3 の評価が極めて低い点である．特に評価者 B および C において， M_3^a と M_3^b が最下位となっている．これは，ネタバレ度スコアが高い文書区間，すなわち物語の核心や山場を単純に削除した結果，あらすじとして読者の興味を惹くフックや重要な導入部分までもが欠落し，文脈が不自然となったためである．この結果は，ネタバレ度スコアが物語の魅力的な要素を適切に捉えていること，つまりネタバレ箇所の検出性能が高いことを逆説的に示している．

第二に，方法 M_2 の有効性である． M_2 によるあらすじは，すべての評価者において安定して上位を獲得した．方法 M_1 と比較しても，手法 M_2 の方が「物語の核心を保護しつつも具体的」という，安全面と品質面の高度なトレードオフを成立させている．これは，GPT-5 がネタバレ度スコアを情報の重要度や秘匿性に関するメタデータとして解釈し，全体の文脈を維持しながらも高スコア区間の情報漏洩のみを適応的に抑制した結果であると考えられる．

4 まとめ

本研究では，ライトノベル 6 冊を対象に，複合ネタバレ度指標に対する評価を行った．その結果，以下の成果が得られた．第一に，ネタバレ検出における表紙要素の重要性を明らかにした点である．従来，表紙イラストは読者が最初に接する情報であり，既知の要素としてネタバレには該当しないと解釈される傾向にあった．しかし本研究により，ライトノベルの表紙には物語のクライマックスや重要人物が描写されていることが多く，表紙との類似性が高いシーンほどネタバレ度も高いという相関が確認された．

第二に，複合ネタバレ度指標の最適重みを特定した点である．実験の結果，同一作品内における文書区間同士の相対的な差異を基準とする方法が最も有効であることが判明した．これは，作品ごとに評価基準やネタバレの性質が異なるため，同一作品内での相対的な盛り上がり捉えるアプローチが，より正確な特定に寄与するためである．

第三に，大規模言語モデルを用いた応用可能性を示した点である．算出されたネタバレ度をメタデータとしてモデルに提示することで，物語の核心を保護しつ

つ、読者の関心を惹く魅力的なあらすじが生成可能であることを確認した。単なる機械的な削除では文脈の不自然さを招くが、スコアを情報の重要度や秘匿性の指標として活用することで、生成物の品質維持とネタバレ防止の両立に成功した。

今後の課題として次の3点が挙げられる。第一に、評価規模の拡大と客観性の向上である。本研究における評価は限られた被験者数による試行的な検証にとどまっている。今後はより大規模な被験者を対象とした調査を行い、ネタバレ度指標の汎用性と実用性を統計的に検証する必要がある。

第二に、他ジャンルへの適応である。例えば、ミステリーやサスペンスといったジャンルでは、表紙イラストが読者を誘導するミスリードとして機能する場合も想定される。ジャンル固有の表現上の特性を考慮し、ネタバレ度算出のアルゴリズムを適応的に調整していくことが求められる。

最後に、スコア統合モデルの高度化である。各指標の線形和ではなく、指標間の相互作用や非線形な関係性を考慮した、より複雑なモデルを導入する余地がある。高度な機械学習手法を取り入れることで、人間の複雑なネタバレ感覚により近い判定モデルへの発展が期待される。

参考文献

- [1] M. Wan, R. Misra, N. Nakashole, and J. McAuley : Fine-grained spoiler detection from large-scale review corpora, In *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, pp.2605–2610, 2019.
- [2] B. Chang, I. Lee, H. Kim, and J. Kang : “Killing Me” is not a spoiler: Spoiler detection model using graph neural networks with dependency relation-aware attention mechanism, In *Proc. of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, pp.3613–3617, 2021.
- [3] 前田 恭佑, 土方 嘉徳, 中村 聡史, 酒田 信親 : ストーリー文書を用いたレビュー文書のネタバレ判定, システム制御情報学会論文誌, 32(3):87–100, 2019.
- [4] R. Tran, C. Xu, and J. McAuley : Spoiler detection as semantic text matching, In *Proc. of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp.6109–6113, 2023.
- [5] D. Inoue and T. Ozaki : A Composite Scoring Method for Spoiler Detection in Light Novels, *Proc. of the 2026 11th International Conference on Intelligent Information Technology*, 2026.