# 扇情的な記事判定に向けた定義作成とアノテーション

谷口祐太郎　上村真史　小澤俊介　関喜史
株式会社 Gunosy
{yutaro.taniguchi, masashi.uemura, shunsuke.kozawa, yoshifumi.seki}@gunosy.com

## 1　はじめに

推薦システムではクリック率が高いアイテムが推薦されやすい傾向にあり，その結果我々が提供するニュースアプリ「グノシー」では，推薦記事の中の性的な記事・攻撃的な記事・クリックベイト的な記事の割合が大きくなるという課題がある[1)]．このような記事を「**扇情的な記事**」と定義し，サービス内での表示を抑制しより良いユーザ体験の実現を目指すため，定義に基づいてデータセットを作成，さらに自動で判定するシステムを構築する．

類似の取り組みとして，ニュースの品質の判定 [1]，クリックベイトの判定 [2, 3]，ニュースの感情分析 [4, 5] などがある．しかしいずれも本研究の目的には不十分であり，独自のデータ定義とデータセットの作成が必要である．データセットの作成は自然言語処理の様々な分野で行われており [6, 7, 8]，文書・記事の品質に関する定義・アノテーションとしては [9, 10, 11] が挙げられる．[9] は類似した定義を提案しているが，本研究ではアプリでの表示抑制を前提とする，より詳細にカテゴリ分けされた定義を作成した．

扇情的な記事について，「定義の作成」，「ラベルデータの作成」，「モデルの作成と評価」の 3 つの工程を行った．我々の知る限り，扇情的な記事に関するデータセットで利用可能なものはないため，定義，データセットの構築から行い，構築したデータセットをもとにモデルの作成と評価を行う．以降の章でそれぞれの工程について詳しく述べる．

## 2　扇情的な記事の定義作成

扇情的な記事の定義を作成する．一般に「扇情的」といったとき，様々な定義・側面があるが，本研究では「サービス内でのユーザ満足度の低下を防ぐ」という目的に沿うように定義する．この定義作成は主に 3 人の作業者によって行われた．対象とする記事のカテゴリは芸能を扱う記事のカテゴリであるエンタメのみとする[2)]．今後，必要性・実現可能性に応じて対象カテゴリを拡張する．

### 2.1　定義作成の手順

以下に定義作成の手順を示す．

1. **記事分析**　一定数の記事に対して，各作業者が直感的に「扇情的だと思ったか否か」を判定し，それらを原因別に分類する．例えば，「表現が過激」・「タイトルが誤解を招く」・「画像がアダルト」・「中傷を含む」などである．
2. **仮定義作成**　仮分類をもとに作業者間で議論し，分類カテゴリと定義を作成する．
3. **仮ラベリング**　暫定の定義にしたがって記事に対してラベリングを行う．
4. **議論**　仮ラベリングの結果を照合し，作業者間で判定が不一致であった事例や，各人が判定に迷った事例について議論をする．議論内容の事例を付録 A に示す．
5. **定義作成 (修正)**　議論結果をもとに定義を修正する．
6. 3〜5 を繰り返し，より要件を満たすように定義を修正していく．

定義の修正は主に以下の 2 種類がある．

- 仮ラベリングを行うことで作業者間での意識の差が判明し，議論によって定義がより明瞭な言い回しに改善される．
- 仮ラベリングを行うことで新たな種類の事例が発見され，議論によって定義に追加される．また，必要に応じて削除・統合される．

1)　例えば「公共の場で開きにくいため，人に勧めにくい」，「過激な記事によりユーザに不快感を与える」，「サービスの信頼性を低下させる」などの原因でサービスに不利益を与える可能性がある．

2)　グノシーにおいて記事は分類システムによって予め定義されたカテゴリに振り分けられている．当初は他のカテゴリの記事も対象としていたが，「エンタメ以外のカテゴリでは扇情的な記事が少ない」・「政治カテゴリなどは扇情的かどうかの判断が難しい」などの理由からエンタメのみとした．

表 1　「扇情的な記事」の定義のクラス

| クラス名 | 説明 |
|---|---|
| SEXUAL | 過度に性的な表現または画像を含む記事. |
| VULGAR | 品位に欠ける・卑俗的な表現を含む記事．筆者が品位に欠ける書きぶりをしていたり，低俗な内容を取り上げていたりする記事などが含まれる. |
| OFFENSIVE | 攻撃的な表現を含む記事．特定の人物に対する中傷や，匿名による過度な批判を取り上げたものなどが含まれる. |
| CLICKBAIT | タイトルに本文内容に関する誤解を招くような表現を含む記事．あえて本文で取り上げる事象とは異なる捉え方に誘導したり，関係ない事実を並べたりしているものなどが含まれる. |
| VAGUE | タイトルが本文内容に対して不明瞭である記事．本文で取り上げる事象に対してタイトルの説明が不足しているものなどが含まれる. |

作業者間の議論だけでは判断しかねるケースでは，社内の他社員の意見や，社内や他社メディアのコンテンツポリシーなどを参考にした.

手順 3〜5 を 8 回繰り返し定義が完成した．仮ラベリングの記事数はサイクルによって異なるが，1 人 100〜200 記事程度である．また，3 人の作業者のうち 2 人のみが仮ラベリングを行う場合もあった.

### 2.2　定義

構築した「扇情的な記事」の定義は，5 つのクラス，各クラスに該当するための条件 (以下，**クラス条件**)，注意事項で構成されている．5 つのクラスラベルとその説明を表 1 に示す．定義の完全版は事業の関係上非公開としているが，今後の公開も視野に入れている．例えば SEXUAL クラスのクラス条件は，「画像がアダルトである」，「不適切な文言がタイトルに含まれる」などである．また，クラスラベルは背反ではなく，扇情的な記事を漏れがないように扱うためのものである．よって，1 記事に対して複数のクラスが当てはまる場合があり，実際にデータセットを作成するときは最もよくあてはまるクラスをラベリングする.

### 2.3　作業者間のラベル一致度の変化

作業者間での議論を重ね，認識合わせ・定義改善を行うにつれ，二値ラベルの作業者間の一致度は向上した．2 回目，6 回目，8 回目の仮ラベリングにおける Fleiss' kappa[12]，Krippendorff's alpha[13] を表 2 に示す．また，3 人中任意の 2 人の Cohen's kappa[14] の値も同じく増加する傾向であった．なお，8 回目の仮ラベリングの後，データセットのための本ラベリングが行われた.

表 2　作業者間でのラベル一致率の変化

| No. | n of samples | Fleiss' κ | Krippendorff's α |
|---|---|---|---|
| 2 | 200 | 0.394 | 0.395 |
| 6 | 100 | 0.479 | 0.480 |
| 8 | 100 | 0.773 | 0.774 |

表 3　ラベル・クラス内訳

| | |
|---|---|
| 扇情的である | 899 |
| SEXUAL | 573 |
| VULGAR | 64 |
| OFFENSIVE | 247 |
| CLICKBAIT | 12 |
| VAGUE | 1 |
| その他 | 2 |
| 扇情的でない | 1287 |
| 合計 | 2186 |

## 3　扇情的な記事のデータセット作成

データセットの概要を示す．対象は 2019 年 8 月〜2020 年 3 月のエンタメカテゴリの記事である．特定の期間の話題に偏ることを防ぐため半年以上の期間とした．記事数は 2,186 記事である．記事をランダムに抽出すると，正例が極めて少ない割合になる．そこで，社内の専門家によって扇情的な記事が含まれやすいと判断された特定メディアとそうでないメディアから同程度の記事数を取得する．メディア分類となることを避けるため，各メディアの正例・負例の割合が同程度になるようにする.

作業者は議論・仮ラベリングによって十分に教育されているとみなし，効率を重視して 1 記事について 1 人のみのラベリングとした．1 記事に対して，以下の情報をアノテーションする．以下の情報のうち判定を行う順番は任意でよいとした.

- 「扇情的」であるかどうか
- 「扇情的」ならば定義したクラスのうちどれに最もよく当てはまるか
- 「扇情的」ならば定義したクラス条件のうちどれに最もよく当てはまるか

ラベル・クラスの内訳を表 3 に示す.

## 4　扇情的な記事判定モデルの作成と評価

構築したデータセットを元にした，扇情的な記事を判定するモデルの作成と評価について述べる.

### 4.1 タスクの定義

SEXUAL，VULGAR ラベルを判定するタスクを Task1，OFFENSIVE，CLICKBAIT ラベルを判定するタスクを Task2 とする．これは，扇情的であるラベルの記事を分析した結果，SEXUAL，VULGAR ラベルが振られた記事，また OFFENSIVE，CLICKBAIT ラベルが振られた記事のそれぞれで書かれ方やサムネイルなどの性質が類似していたためである．

### 4.2 モデルの作成

Task1，2 それぞれについて，ルールベースと SVM の 2 種類のモデルを作成する．また，提供元が特定メディアであるかどうかを判定結果とする手法をベースラインとして比較する．

**ルールベース** Task1 については，記事のサムネイル画像による判定，記事のテキストによる判定の OR 結合を判定結果とする．サムネイルの画像判定には Amazon Rekognition[3] を用いてラベルを付与し，特定のラベル（例えば，Explicit Nudity や Suggestive）が付与されていた場合，正例と判定する．また，記事テキストによる判定では，あらかじめ選定した特徴語が 1 語以上記事タイトル・本文に含まれている場合，正例と判定する．特徴語は，SEXUAL，VULGAR ラベルの記事を分析し，出現しやすい単語 (「グラビア」,「セクシー」など) を人手で選定したものである．

Task2 については，記事のテキスト，発信元のメディア情報，本文の長さをもとにしたスコアリングによる判定を行う．記事テキストによる判定では，あらかじめ選定した特徴語ごとに割り当てたスコアをもとに，テキストに含まれる特徴語のスコア合計値を算出して判定を行う．さらに，提供元が特定メディアでない場合および記事本文の文字数が $t_w$ 以上の場合は減点を行う．これは記事が調査に基づいた内容で扇情的でないことが多いためである．最終的なスコアが $t$ 以上であれば正例と判定する．特徴語の例およびスコアリングのルールの詳細を付録 C に示す．

**SVM** SVM(Support Vector Machine)[15] を用いて扇情的な記事かどうかの分類を行う．Task1，2 ともに，記事のタイトルと本文から算出した TF-IDF を特徴量として用いる．形態素解析には MeCab[4]，辞書は NEologd[5] を用いる．Task1 についてのみ，SVM による判定と画像による判定の OR 結合を判定結果とするようにした．

3） https://aws.amazon.com/rekognition/，Amazon 社が提供する機械学習による画像や動画の分析を行う API

4） http://taku910.github.io/mecab/

**表 4** 各データセットの正例負例内訳

| | 検証用 | | 評価用 | |
|---|---|---|---|---|
| | 正例 | 負例 | 正例 | 負例 |
| Task1 | 356 | 766 | 281 | 519 |
| Task2 | 141 | 766 | 120 | 519 |

**表 5** 各タスクにおけるモデル毎の Precision，Recall，f1-score

| Task | モデル | Precision | Recall | f1-score |
|---|---|---|---|---|
| Task1 | ベースライン | 38.4% | 57.7% | 46.0% |
| | ルールベース | 70.5% | 88.3% | 78.4% |
| | SVM | 87.8% | 79.0% | 83.2% |
| Task2 | ベースライン | 24.1% | 69.2% | 35.7% |
| | ルールベース | 86.0% | 40.8% | 55.4% |
| | SVM | 77.6% | 43.3% | 55.6% |

### 4.3 実験条件

全 2,186 記事のうち数が少ない VAGUE，その他ラベルの 3 件を除いて，1,263 件を検証用，920 件を評価用とする．モデルの評価には，Precision，Recall，f1-score を用いる．検証用データを用いて，ルールベースモデルの特徴語の選択，また，SVM の学習を行い，評価用データを用いて性能を算出する．Task2 のルールベースモデルにおいて，$t = 0.20, t_w = 3000$ と定めた．これは検証用のデータを用いて Precision と Recall のバランスを考慮して決定した．

2 つのタスクではそれぞれ正例のデータとして異なる記事を利用する．Task1 では SEXUAL，VULGAR のいずれかのラベルが付与された記事，Task2 では OFFENSIVE，CLICKBAIT のいずれかのラベルが付与された記事を正例として用いる．負例として用いる記事はタスクによらず共通である．表 4 にデータの内訳を示す．

### 4.4 実験結果

表 5 に各タスクにおけるモデルの性能を示す．いずれのタスクにおいても，ルールベース手法（以下，RB）と SVM 手法（以下，SVM）は，ともにベースラインの性能を上回る結果となった．

Task1 においては，Amazon Rekognition の判定結果を用いることで，肌露出が多い画像が含まれる記事を正しく判定することができた．RB と SVM を比較すると，Precision は SVM が RB よりも 17.3 ポイント高いのに対し，Recall は RB が SVM より 9.3 ポイ

5） https://github.com/neologd/mecab-ipadic-neologd

表 6　誤判定傾向・事例

| Task | 手法 | FP | FN |
|---|---|---|---|
| 1 | RB | 事件内容に扇情的な語が含まれている場合や，肩書きとして「グラビア」などの語が出現する場合に正例と誤判定される． | 誤判定例のバリエーションは多く，正例の中で多くの割合を占める写真集紹介系の記事が多く見られる． |
| 1 | SVM | 肌の露出が少ない，健全な写真集・雑誌の紹介が誤判定される． | 下品な番組内容や卑俗的なゴシップなど，VULGAR ラベルの記事が誤判定される． |
| 2 | RB | 番組紹介記事におけるエンターテインメントとしての「激怒」や，中立的な立場から取り上げている炎上記事が「炎上」などの語により誤判定される． | 攻撃的な意見を含むが，「○○に似ている」といったような記事や，芸能人のゴシップ記事において攻撃的な語を含まないため誤判定される． |
| 2 | SVM | ポジティブなゴシップ，中立的な批判などが誤判定される． | False Negative 例は多く (Recall は低い)，頻出の攻撃的な記事も誤判定される． |

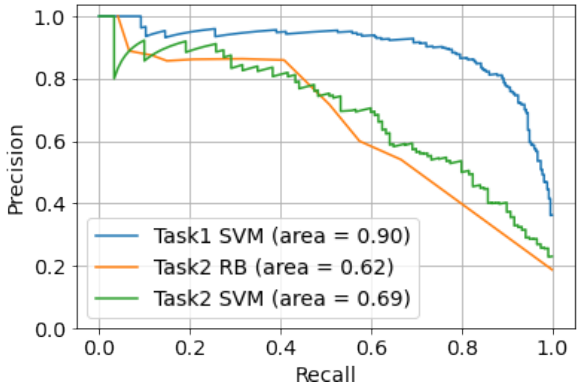


図 1　各タスクにおけるモデル毎の PR 曲線

ント高いという結果となった．

Task2 においては，RB が SVM より Precision が 8.4 ポイント，Recall が 2.5 ポイント高い結果になった．また，Task2 における f1-score は Task1 と比較して全体的に低い傾向になった．

図 1 に Precision-Recall 曲線を示す．例えば Recall が 0.80 のとき Precision は Task1 では 0.87，Task2 では 0.53 という結果であった．実サービスでは正例の取りこぼしを少なくすることが重要であるとして，Recall を優先することを想定している．付録 C に Task2-ルールベースモデルのルールごとの性能を示す．

### 4.5　考察

扇情的な記事判定での誤判定をもとに考察を行う．表 6 に各タスクの誤判定例を示す．

Task1 においては，正例に特定の語が出現しやすく，サムネイルもグラビアなど識別しやすいものが多いため，ルールベース，SVM 問わず一定の割合を正しく識別できた．RB の誤判定として目立つのは，非扇情的な文脈で特徴語を含む記事を，誤って扇情的な記事と判定する場合である．例えば，「グラビア」という特徴語はグラビアアイドルの肌の露出が多い画像を含むような記事を想定して追加したものだが，単なる肩書きとして「グラビアアイドル」という語を含む非扇情的な文章で誤判定してしまう．このため Recall は高くなるものの，Precision は低下したと考えられる．しかし FN 例にあるように，写真集を紹介する非扇情的な記事は，文体や出現語も正例とほとんど差がないため識別が難しい．

Task2 は手法を問わず Task1 よりも性能が低い．人手での判定も困難な記事が多いことが要因の一つである．具体的には，FP 例にあるような批判を表す文言について，過剰に一個人を中傷するものなのか，ただ注意喚起を行うものなのかの区別が難しく，どこからが扇情的に当てはまるのか人によっても判断が異なるため，モデルでも識別することが難しい．また，OFFENSIVE，CLICKBAIT ラベルの記事は表現が多様で，攻撃的な語は含まれないが攻撃的な意見を含む扇情的な記事も多く存在するため，文章の解釈が行えず手法を問わず性能が低い結果となった．

## 5　まとめ・今後の課題

ニュースサービスにおいて好ましくない影響を与える可能性がある記事を「扇情的な記事」として，複数の作業者が議論を行い定義を改善するというループを 8 回行い 5 クラスから構成される定義を構築した．定義に基づき，データセットを作成，自動判定するモデルを提案し，ルールベース手法・SVM 手法ともに，ベースラインよりも高い Precision，Recall を導いた．

今後は，ルールの改善 (特徴語の追加) や識別器の工夫により性能の向上を目指すとともに，定義の修正・対象カテゴリの拡張・データセットの拡張にも着手する．

# 参考文献


[1] Yen-Hao Huang, Ting-Wei Liu, Ssu-Rui Lee, Fernando Henrique Calderon Alvarado, and Yi-Shin Chen. Conquering cross-source failure for news credibility: Learning generalizable representations beyond content embedding. In *Proceedings of The Web Conference 2020*, WWW '20, p. 774–784, New York, NY, USA, 2020. Association for Computing Machinery.

[2] Ankesh Anand, Tanmoy Chakraborty, and Noseong Park. We used neural networks to detect clickbaits: You won't believe what happened next! In *European Conference on Information Retrieval*, pp. 541–547. Springer, 2017.

[3] Hai-Tao Zheng, Xin Yao, Yong Jiang, Shu-Tao Xia, and Xi Xiao. Boost clickbait detection based on user behavior analysis. In *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data*, pp. 73–80. Springer, 2017.

[4] Julio Rieis, Fabrício de Souza, Pedro Vaz de Melo, Raquel Prates, Haewoon Kwak, and Jisun An. Breaking the news: First impressions matter on online news. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 9, 2015.

[5] Antony Samuels and John Mcgonical. News sentiment analysis, 2020.

[6] 東中竜一郎, 光田航, 増村亮, 斉藤いつみ, 青野裕司. 雑談要約技術に向けた取り組み. 言語処理学会 第 26 回年次大会 発表論文集, pp. 1519–1522, 2020.

[7] 対話的議論の自動評価に向けたディベートデータセットの構築. 言語処理学会 第 26 回年次大会 発表論文集, pp. 708–711, 2020.

[8] 仁木裕太, 坂地泰紀, 松島裕康, 和泉潔. 因果判定データセットの構築と原因結果表現抽出への拡張. 言語処理学会 第 26 回年次大会 発表論文集, pp. 557–560, 2020.

[9] Hongyu Lu, Min Zhang, Weizhi Ma, Yunqiu Shao, Yiqun Liu, and Shaoping Ma. Quality effects on user preferences and behaviorsin mobile news streaming. In *The World Wide Web Conference*, WWW '19, p. 1187–1197, New York, NY, USA, 2019. Association for Computing Machinery.

[10] 渋木英潔, 中野正寛, 宮崎林太郎, 石下円香, 金子浩一, 永井隆広, 森辰則. 情報信憑性判断支援のための web 文書向け要約生成タスクにおけるアノテーション. 自然言語処理, Vol. 21, No. 2, pp. 157–212, 2014.

[11] Ioannis Arapakis, Filipa Peleja, Barla Berkant, and Joao Magalhaes. Linguistic benchmarks of online news article quality. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1893–1902, Berlin, Germany, August 2016. Association for Computational Linguistics.

[12] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. john wiley & sons, 2013.

[13] Andrew F Hayes and Klaus Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, Vol. 1, No. 1, pp. 77–89, 2007.

[14] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, Vol. 20, No. 1, pp. 37–46, 1960.

[15] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, p. 144–152, New York, NY, USA, 1992. Association for Computing Machinery.

表 7　議論例

| 議題 | 結論 |
|---|---|
| 「ある芸能人の結婚報告にファンたちが驚いた」という主旨のタイトルだが，本文の内容は「ドラマの中のウェディング姿を SNS にアップロードした」話だったというケースの場合，本定義に含めるか. | このようなタイトルによるひっかけはある種の娯楽とみなして本定義に含めない．ただし特定の人物などの悪いイメージの誤認を誘導するようなミスリードは本定義に含める. |
| ある芸能人の特定の行為に対する批判として SNS の書き込みを取り上げている記事は本定義に含めるか. | 中傷であったとしても正当な批判とみなせるとしても，根拠の薄い匿名のネガティブな書き込みを取り上げて作られている記事は記事リストやプッシュ通知で提供すべきでない．よって本定義に含める. |
| 「VAGUE」(不明瞭) クラスは「タイトルから内容が想像できない」場合に付与されるが，想像できるかどうかは読み手の知識によって変わるのではないか. | 読み手に背景知識があったとしても，何についての記事か理解できないような場合は「VAGUE」クラスとする．伝えたい事象に関する単語が含まれているなど，背景知識があれば理解できるようになっていれば「VAGUE」クラスには含めない. |

## A　議論例

表 7 に定義作成の過程の議論の例を示す.

## B　定義の変更例

1. ワイドショーでの芸能人のコメントを取り上げた記事などの事例を参考に「内容が薄い」というクラスを設けていたが，議論の結果，それだけで扇情的と判定するのは厳しすぎるという結論に至り，廃止された．ただし，匿名によるネガティブなバッシングを取り上げて作成された記事や，攻撃的なコメントを取り上げた記事は「OFFENSIVE」クラスとして扇情的であると判定することにした.
2. 「有名人による SNS の投稿画像の場合は肌の露出が多くても扇情的でないとする」という基準だったが，「プッシュ通知に適切かどうか」など本来の目的を考慮して議論した結果，肌の露出が多い記事は基本的に扇情的であると判定することにした.
3. 「卑俗なテレビ番組の内容を伝える記事は，事実をそのまま伝えているため扇情的でない」という基準だったが，変更例 2 同様，本来の目的を考慮して議論した結果，扇情的の定義に含めるという結論に至った.

表 8　Task2 特徴語の分類とその例

| | 詳細 | 例 |
|---|---|---|
| 強 | OFFENSIVE，CLICKBAIT ラベルの記事に多く含まれ，攻撃の程度が大きい語 | 「批判殺到」，「厳しい声」 |
| 中 | OFFENSIVE，CLICKBAIT ラベルの記事に多く含まれる語 | 「炎上」，「激怒」，「嘲笑」 |
| 弱 | 単体では攻撃的ではないが OFFENSIVE，CLICKBAIT ラベルの記事にしばしば出現する語 | 「ネット上」，「干される」 |
| 反 | OFFENSIVE，CLICKBAIT ラベルと誤判定されやすい記事に出現する語 | 「称賛」，「心配」，「反論」 |

表 9　Task2 ルールベースモデルのスコアリングルール

| ルール | スコア |
|---|---|
| Task2 特徴語 (強) が含まれる | +0.15 |
| Task2 特徴語 (中) が含まれる | +0.10 |
| Task2 特徴語 (弱) が含まれる | +0.05 |
| Task2 特徴語 (反) が含まれる | -0.05 |
| 特定メディアの記事でない | -0.20 |
| 本文の長さが一定以上である | -0.10 |

表 10　Task2 ルールベースモデルの各ルールでの性能比較．ルールにおいて，A は特徴語 (弱・中・強)，A' は特徴語 (弱・中・強・反)，B は特定メディア，C は本文の長さを示す.

| | 検証用データ | | | 評価用データ | | |
|---|---|---|---|---|---|---|
| | Prec | Recall | f1 | Prec | Recall | f1 |
| A | 70.6% | 73.1% | 71.8% | 72.8% | 55.8% | 63.2% |
| A‘ | 74.6% | 66.7% | 70.4% | 81.7% | 48.3% | 60.7% |
| A + B | 75.0% | 68.1% | 71.4% | 81.4% | 47.5% | 60.0% |
| A‘ + B | 79.1% | 61.7% | 69.3% | 86.2% | 41.7% | 56.2% |
| A + C | 72.3% | 72.3% | 72.3% | 74.7% | 54.2% | 62.8% |
| A‘ + C | 76.2% | 66.0% | 70.7% | 81.4% | 47.5% | 60.0% |
| A + B + C | 78.7% | 68.1% | 73.0% | 80.9% | 45.8% | 58.5% |
| A‘ + B + C | 81.3% | 61.7% | 70.2% | 86.0% | 40.8% | 55.4% |

## C　Task2 ルールベースモデルの詳細

Task2 に向けたルールベースモデルについて，特徴語の分類を表 8 に，スコアリングのルールを表 9 に示す.

また，Task2 に向けたルールベースモデルにおける，ルールごとの性能の比較を表 10 に示す.