

失敗経験の抽出に向けた QA サイトの分析 Analysis of Q&A Sites for Extracting Experience of Failure

下元 健吾†
Kengo Shimomoto

安藤 一秋‡
Kazuaki Ando

1. はじめに

インターネット上で、誰でも気軽に個人の意見を発信できるようになった。その結果、Web 上では個人の経験情報が様々な場所に記載されるようになった。こうした経験情報は、企業にとっては顧客のニーズを分析する情報として、また、一般ユーザにとっては意思決定や問題解決の際に利用されている。

このような経験情報の中でも「失敗」を経験した情報は特に有用であると考えられる。失敗経験情報は、その情報から失敗の原因を特定することで、失敗の予防策を獲得できる。また、失敗後の対処法を特定することで、失敗に対する解決策も獲得できる。このように失敗経験情報からは、多くの有用な情報を得ることができると考えられる。しかし、失敗経験情報は Web 上に散在しており、一般ユーザが自由に利用できる形で整理されたデータベースやサービスは現在存在しない。また、検索エンジンで失敗経験をキーワード検索しても、多くの情報に埋もれてしまい、欲しい情報が容易に獲得できない。さらに、キーワードが Web テキスト中に記載されていない失敗経験は検索できない問題もある。

そこで本研究で、Web 上から失敗経験情報を自動抽出することを目的とする。個人の失敗経験は Web 上のあらゆる場所に存在すると考えられるが、本研究では、個人の経験が書かれやすく、一定以上の文字数で記述されやすいブログメディアを対象に設定する。

本稿では、ブログから失敗経験情報を抽出する手法を検討するための事前調査として、QA サイトに記載される失敗情報について分析する。

2. 関連研究

浅沼らは、Twitter 上に書かれる個人の失敗情報を抽出する手法[1]を提案している。失敗について記述する際によく使用される「失敗ワード」を手がかり表現に利用し、ツイート内から失敗情報の抽出を行う。評価の結果、有用性のある失敗情報は低いもので 28%、高いもので 75% の適合度で抽出している。

小野らは、ブログから失敗情報を抽出する手[2]法を提案している。失敗談を「ある行動を起こした結果、思惑通りに物事が運ばなかったときの体験談」と定義し、失敗談を F 値 0.32 で抽出している。抽出された情報の有用性は評価されていない。

これらの手法の抽出精度は改善する余地が残っており、抽出した情報の有用性判定も行なわれていない。したがって、本研究では、高い精度で有用な失敗経験を抽出する手法の実現を目指す。

† 香川大学大学院工学研究科
Graduate School of Engineering, Kagawa University

‡ 香川大学工学部
Faculty of Engineering, Kagawa University

3. QA サイトに記載される失敗経験の分析

ブログから失敗経験を抽出する手法を検討する事前調査として、どのような失敗経験に対して、どのような情報を知りたいのかを確認するため、QA サイトを分析する。本分析では、QA サイトの中でも総記事数の多い Yahoo! 知恵袋を対象に、質問文の分析を行なう。

3.1. 失敗経験が記述されやすいカテゴリの調査

QA サイトはカテゴリ分けされている。失敗経験が記述されやすいカテゴリの有無を調査するため、ヒット数に基づいてカテゴリを分類する。Yahoo! 知恵袋は 3 階層にカテゴリ分類されているため、3 つの階層における各カテゴリに対して、キーワード「失敗」で検索し、ヒット件数を調査する。

図 1 に、第 3 階層の上位 18 件の各カテゴリに対するヒット件数を示す。

調査の結果、第 1 階層のカテゴリで 1 番ヒット件数が多かったものは「生き方と恋愛、人間関係の悩み」、第 2 階層では「恋愛相談、人間関係の悩み」、第 3 階層では「恋愛相談」であった。3 つの階層のどのカテゴリにおいても恋愛に関する失敗経験が多く投稿される傾向が確認できた。図 1 に第 3 階層の上位 18 件のカテゴリのヒット件数を示す。

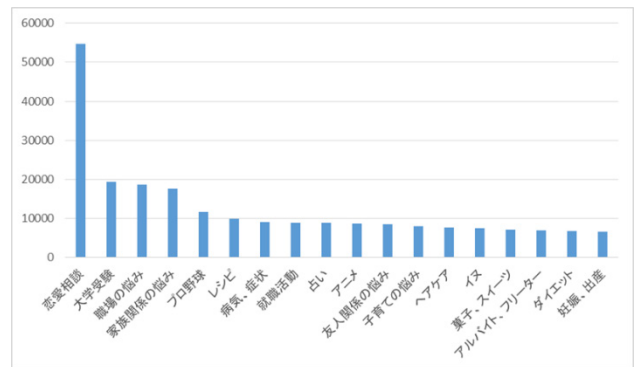


図 1: 第 3 階層のカテゴリ郡の「失敗」ヒット件数

図 1 から、恋愛相談のカテゴリでのヒット数が他のカテゴリより圧倒的に「失敗」を含んでいることがわかる。次の分析では、第 3 階層の上位カテゴリに注目して、失敗対象の分析を行なう。

3.2. 失敗対象の分析

QA サイトに記載されている失敗の対象を調査するために質問文の分析を行なう。先の分析において、ヒットした件数が多い第 3 階層の上位 5 カテゴリ (恋愛相談、大学受験、職場の悩み、家族関係の悩み、プロ野球) を分析対象とする。これらのカテゴリに対し、「で失敗」というキーワードで質問文を検索する。そして、ヒットした 1,000 件の質問文に対して形態素解析を行い、「で失敗」の直前に現れる単語の頻度をカウントする。

調査の結果、大学受験カテゴリでは「センター」「試験」という単語が「で失敗」の直前語として頻出し、二つを合わせた割合は 65.3%となった。カテゴリ内では特定の失敗経験についてのみ記載される可能性があると考えられる。また、職場の悩みカテゴリでは「仕事」という単語が 50.5%の割合で、家族関係の悩みカテゴリでは 11.2%の割合で、恋愛相談カテゴリでは 12.3%の割合で獲得できた。職場の悩み、家族関係の悩み、恋愛相談カテゴリなどの人間関係に関するカテゴリに関しては、類似した失敗経験が記載されている可能性が考えられる。

3.3. 質問文からの失敗経験の分析

質問者が失敗に関してどのような情報を欲しているかを確認するため、QA サイトの質問文をタイプに分類して分析する。分析対象は、QA サイトの第 3 階層の 6 つのカテゴリ「恋愛相談」「大学受験」「病気、症状」「友人関係の悩み」「Windows 全般」「法律相談」とする。

QA サイトのカテゴリ毎にキーワード「で失敗」で検索した 1,000 件の質問文の内、文末が「?」「か。」で終わる文を抽出し、QA サイトの主要質問文とする。その質問文の中から、質問者が実際に失敗を経験したことをふまえて質問している 50 文をそれぞれのカテゴリから抽出し、5 つの質問タイプに分類する。分類方法は渡邊らの QA サイトの分類研究[3]で用いられた方法を参考にする。渡邊らは質問のタイプを正解有と正解無の 2 タイプに分類した後、さらにそれらを複数のタイプに分類している。分類のタイプを表 1 に示す。また、6 カテゴリを表 1 のタイプに分類した結果を表 2 に示す。

表 1. 質問文の分類タイプ

質問タイプ	定義
事実(正解有)	事象の定義, 真実, 客観的な理由や手法を問う質問
根拠(正解有)	客観的な根拠, 理由を問う質問
経験(正解無)	回答者の経験や体験がなければ回答できない質問
提案(正解無)	問題の解決方法を問う, 情報提供を依頼する
意見(正解無)	推測, 嗜好など, 主観的に回答をしてよい質問

表 2. 各カテゴリの分類結果

カテゴリ名	事実	根拠	経験	提案	意見
恋愛相談	1	0	12	4	33
大学受験	1	0	2	3	44
病気, 症状	17	0	2	0	31
友人関係の悩み	3	1	3	19	24
Windows 全般	42	4	2	0	2
法律相談	36	0	1	1	12
カテゴリ合計	100	5	22	27	146

調査した 6 カテゴリ全体をみると「センターで失敗しても二次に自信あれば二次で挽回できますか?」のような意見を聞く質問タイプが最も多く、「禁煙外来で去年禁煙に失敗しましたが、一年たてばまた、保険適用になる」というのは

本当ですか?」のような事実を聞く質問タイプが 2 番目に多いことがわかった。

「恋愛相談」「大学受験」カテゴリでは、「男心が知りたいです付き合う前の段階でのデートで失敗したことありますか?」のような意見を聞く質問タイプが最も多かった。これは人の感情や進路等、解が一つに定まらないものについて相談するような質問が多く投稿されていると考えられる。

「Windows 全般」「法律相談」カテゴリでは、「手でインストールしても失敗になってしまうのですが、何か解決策はありますか?」のような事実を聞く質問が多い傾向が見られた。これはパソコンのエラーの対処方法や問題が起こった際の法律を基にした対処法等、解が一意に定まりやすい質問が多く投稿されていることが考えられる。

「友人関係の悩み」カテゴリは「恋愛相談」カテゴリと似たような質問傾向であったが、「友人関係の悩み」カテゴリでは、特に「みなさんは、失敗をし、自分のせいで他人に迷惑をかけてしまったとき、どのように立ち直っていますか?」のような問題解決の手法を問う提案タイプの質問が多く投稿される傾向が見られた。

「病気、症状」カテゴリは病気の判断を行うために、「アロマセラピーで失敗して、体調が悪くなった場合、病院に行った方がいいでしょうか?」のような意見を聞く質問と「カテーテルの手術で失敗すれば後遺症や死亡に繋がることってありますよね?」のような病気、症状の事実を聞く質問が多くなされる傾向が見られた。

今後は、ユーザが求める失敗対象を特定するため、質問タイプをさらに細分化し、回答文と共に分析する。

4. おわりに

本稿では、ブログから失敗経験情報を抽出する手法を検討するための事前調査として、QA サイトに記載される失敗情報について分析した。分析結果から QA サイトには恋愛関係の失敗経験が多く記載される傾向を確認した。失敗の対象分析から失敗を表す具体的な単語を抽出することができなかった。しかし、職場の悩みや家族関係の悩みなどの人間関係に関するカテゴリにおいては、失敗の対象が類似する可能性が確認できた。失敗を経験したことをふまえた質問文の調査からは、カテゴリタイプにより失敗を経験した質問者の質問傾向が異なることがわかった。

今後は、ユーザが求める失敗対象を特定するため、QA サイトの質問文の細分化と回答文の分析を行う。また、栗原らの研究[4]を参考に失敗経験情報の抽出方法を検討する。

参考文献

- [1] 浅沼裕輔, “Twitter からの失敗情報抽出手法の提案”, 甲南大学平成 26 年度卒業研究要旨, 2015.
- [2] 小野博紀, 内海彰, “Weblog からの失敗談抽出”, 2012 年度人工知能学会全国大会, pp.1-4, 2012.
- [3] 渡邊直人, 島田諭, 関洋平, 神門典子, 佐藤哲司, “QA コミュニティにおける質問者の期待に基づく質問分類に関する一検討”, 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM 2011), B5-1, 2011.
- [4] 栗原光平, 嶋田和孝, “ブートストラップ方を用いた Twitter からの不具合文抽出”, 言語処理学会第 21 回年次大会論文集, pp341-344, 2015.