

N-017

解決手段情報と効果情報について説明する明細書の重要段落抽出手法の提案 A method of extracting important paragraph that describes solution information and effect information

矢野裕和[†]

Hirokazu Yano

古田壮宏[‡]

Takehiro Furuta

東本崇仁[§]

Takahito Tomoto

赤倉貴子[§]

Takako Akakura

1. はじめに

特許公報は、年間 50 万件近く発行されている [1]。特許公報は技術文書としての役割があり、特許調査などに利用されている。しかし、大量の特許公報の中から必要な情報を探し出すには多くの時間を要する。また、特許公報 1 件当たりのページが多いため、1 件ずつ精読することは困難である。そこで本研究では、ある特許公報が必要なものかどうかを短時間で判断できる要約の作成方法を構築することを最終的目標とする。

特許出願に必要な書類の一つとして要約書がある。要約書の内容は請求項の内容をそのまま記載しているものが多い。しかし、一般に請求項は可読性が低い [2]、特許公報の概要を把握するための情報としては必ずしも適切ではない。

実際、様々な企業において、各企業独自の要約の作成や、専門の業者が作成した要約の購入などが行われている。A 社では、特許を理解するために必要な情報として「課題情報」「解決手段情報」「効果情報」の抽出を行い要約を作成している。ここで、課題情報は従来技術の問題点、解決手段情報はその問題点を解決するための手段、そして、効果情報は解決手段によってもたらされる発明の効果をそれぞれ意味している。ある特許公報において、これらの情報を把握することは、その特許の特徴の把握につながるが、A 社の現状としては、専門家が 1 件ずつ特許公報を読み込み、人手で抽出しているため、多大な時間と労力がかかっている。これらの自動抽出の実現は、A 社のみならず、他の企業においても有用であると考えられる。

本稿では、まず上記の 3 つの情報が特許公報のどこに存在しているのかを調査する。次に、その結果をもとに、解決手段情報を説明する段落の抽出手法の提案と評価を行う。さらに効果情報を説明する重要段落の抽出手法の提案と評価を行う。最後に今後の方針として、課題情報の抽出手法について述べる。

2. 各情報に関する基礎調査

明細書の構成は図 1 のようになっており [1]、発明の詳細な説明は【背景技術】節など複数の節から構成されている。また、それぞれの節は複数の段落で構成されており、各段落は複数の文から構成されている。

課題情報、解決手段情報、効果情報がどこから抽出されているかを調べるため、A 社の知的財産部で 10 年以上の経験を持つ専門家 1 名に特許公報 10 件を読んでもらい、それらの情報の抽出を行ってもらった。その

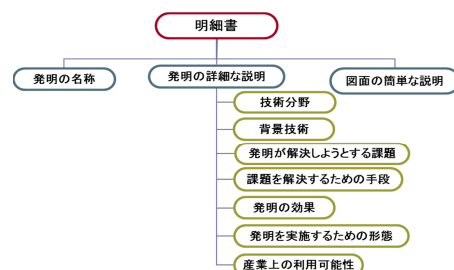


図 1: 明細書の構成

表 1: 各情報の調査結果

抽出情報	抽出元	件数
課題情報	【背景技術】	4
	【発明が解決しようとする課題】	6
解決手段情報	請求項	2
	【課題を解決するための手段】	8
効果情報	【課題を解決するための手段】	1
	【発明の実施形態】	3
	【発明の効果】	6

結果を表 1 に示す。この結果をもとに明細書からそれぞれの情報の抽出を目指す。

3. 解決手段情報を説明する段落の抽出手法と評価

表 1 より解決手段情報は、請求項または【課題を解決するための手段】節から抽出されていることがわかる。この【課題を解決するための手段】節から抽出している段落 8 件のうち、請求項と同じ内容のものは 5 件であった。よって、解決手段情報とは請求項の内容である可能性が高いと考えられる。しかし、請求項は可読性が低いので解決手段情報をそのまま要約の作成に利用することはできない。そこでより具体的な解決手段情報を説明する段落を抽出するために、先行研究 [3] の手法を用いて請求項の構成要素に対する明細書の重要段落の抽出を行う。

先行研究 [3] の手法により抽出された段落が、解決手段情報を説明する段落として適切かどうかを、前記の専門家に読んでもらい判断してもらった。対象は特許公報 10 件、構成要素は 24 個であり、精度と再現率を用いて評価した。

$$\text{精度} = \frac{\text{正しく抽出した段落数}}{\text{抽出した段落数}} = \frac{16}{21} = 0.76 \quad (1)$$

$$\text{再現率} = \frac{\text{正しく抽出した段落数}}{\text{抽出すべき段落数}} = \frac{16}{24} = 0.67 \quad (2)$$

請求項の全ての構成要素について正しく抽出できた特許公報は 7 件あり、本手法は実際の特許公報に適用できると考えられる。請求項や解決手段情報を読むときに、本手法で抽出した段落と一緒に読むことで、より理解しやすくなると考えられる。

[†]東京理科大学大学院工学研究科

[‡]奈良教育大学

[§]東京理科大学工学部

4. 効果情報を説明する段落の抽出手法の提案と評価

4.1. 抽出手法の提案

表 1 より、効果情報は【発明の効果】節から抽出されている場合が多いことがわかる。【発明の効果】節以外から効果情報を抽出している場合は、(1)【発明の効果】節の記述だけでは発明の効果を理解するのに不十分な場合と、(2)【発明の効果】節が明細書中に存在しないため、他の明細書の節から効果情報を抽出している場合がある。しかし【発明の効果】節は 89.5 % の明細書に存在しているため [4]、本稿では (1) を対象とする。

実務において効果情報を他の節の段落から抽出する場合【発明の効果】節に書かれている情報をもとにして抽出していることが専門家に対する調査によりわかった。そこで【発明の効果】節の形態素を用いて他の節の段落との類似度を計算することで、発明の効果について詳しく説明している重要段落の抽出を行う。

ここで、段落間の含意関係を表すスコアである段落間係数 e_{ij} について述べる。本稿では、高村ら [5] の研究と同様に、各段落同士の含意関係は、そこに含まれる単語がどの程度共通しているかによって表されるとみなす。よって、段落 i に対する段落 j の段落間係数 e_{ij} は、段落 j に含まれる単語のうち、段落 i に含まれる単語と共通するものがいくつあるかで表し、以下のように定義する。

$$e_{ij} = \frac{|W(i) \cap W(j)|}{|W(j)|} \quad (3)$$

ここで、 $W(i)$ は段落 i に含まれる単語の集合とする。単語は、名詞、動詞および形容詞を対象とする。段落間係数 e_{ij} を段落 i の段落 j に対する類似度とみなし、【発明の効果】節と類似度の高い段落を【発明の効果】節について詳しく説明している段落として抽出する。

4.2. 抽出手法の適用結果

図 2 は【発明の効果】節の段落である。この段落の情報をもとに類似度を計算し、重要段落として抽出したものが図 3 である。図 3 の段落は【発明を実施するための形態】節から抽出された。この段落は発明の効果についてさらに詳しく説明しており、どのようにして取出口をユーザに向けるかについて具体的に記述している。よって、図 2 の【発明の効果】節よりも【発明の効果】節の情報をを用いて抽出した図 3 の段落のほうが発明の効果についてわかりやすい内容であると考えられる。

【0010】
本発明によれば、載置台を傾斜位置に位置付けるとその載置台上に載せられている買物がこの取出口をユーザに向けることができ、これにより、車椅子ユーザの袋詰め作業をしやすくなるものことができる。

図 2: 【発明の効果】節 (特開 2008-54935)

【0033】
図 6 は、買物かご 109 を傾斜配置した状態を示す袋詰め作業台 101 の斜視図である。載置台 107 の自由端側が自由落下すると、載置台 107 は傾斜配置され、傾斜位置に位置付けられる。これに伴い、載置台 107 の載置面 107a に配置されている買物かご 109 も傾斜配置される。その結果、買物かご 109 は、図 6 に示すような状態、つまり、ユーザの方向に取出口が向けられた状態で保持される。したがって、車椅子ユーザや背が低いユーザにしてみると、買物かご 109 が自分の方向に傾けられた状態となるので、商品の取り出しが容易になる。これにより、車椅子ユーザや背が低いユーザに良好な袋詰め作業性を提供することができる。

図 3: 【発明の効果】節をもとに抽出 (特開 2008-54935)

4.3. 抽出手法の評価

提案手法により抽出された段落が、効果情報を説明する段落として適切かどうかを、前記の専門家に読んでもらい判断してもらった。評価対象は特許公報 9 件であり、精度を用いて評価を行う。

$$\text{精度} = \frac{\text{正しく抽出した段落数}}{\text{抽出した段落数}} = \frac{6}{9} = 0.67 \quad (4)$$

精度は 7 割程度であったが、明細書すべての節を対象に類似度を求め段落の抽出を行うのではなく、【実施形態】節に絞って段落の抽出を行うことでより精度を上げることができると考えられる。

5. 課題情報抽出手法の提案

10 件中 8 件の特許公報について、課題情報が記述されている段落の特徴を調査したところ、(1) 表 2 に示すようなネガティブ表現がある段落であり、なおかつ (2) その次の段落に発明の目的が記述されている段落であることがわかった。

表 1 の結果より、明細書の【背景技術】節と【発明が解決しようとする課題】節を対象に課題情報の抽出を行う。まず表 2 に示すような、発明の目的について述べられていると考えられる手がかり語を検索する。そして、目的の手がかり語が含まれる段落から、一つ前の段落を対象にネガティブ表現を検索する。この段落にネガティブ表現があれば、この段落を課題情報として抽出する。ない場合はさらに前の段落に対して、ネガティブ表現の検索を行う。ネガティブ表現が出現するまでこの作業を繰り返すことで、課題情報の抽出を行う。今後、この手法を用いて課題情報の抽出と評価を行っていく予定である。

表 2: 課題情報を抽出するための手がかり語 (一部抜粋)

目的表現	本発明は、目的、提供する
ネガティブ表現	問題があった、しかしながら、低下する、負担となる、できない、困難であった

6. まとめ

要約作成に必要な「課題情報」、「解決手段情報」、「効果情報」について各情報の抽出元の調査を行った。そのうえで、3 章の解決手段情報を説明する段落の抽出手法と、4 章の効果情報を説明する段落の抽出手法を用いて、段落の抽出を行った。専門家による評価を行い、実際の特許公報に対する提案手法の適用可能性を示した。

参考文献

- [1] 特許庁, <http://www.jpo.jp/indexj.htm>, 参照 June, 2012.
- [2] 新森昭宏, 奥村学, 丸川雄三, 岩山真, “手がかり句を用いた特許請求項の構造解析,” 情報処理学会論文誌, Vol.45, No.3, pp.891-905, 2004.
- [3] 矢野裕和, 古田壮宏, 赤倉貴子, “特許請求項の構成要素に対する重要段落抽出手法の検討,” 電子情報通信学会 2012 年総大会 ISS 学生ポスターセッション予稿集, pp.78, 2012.
- [4] 酒井浩之, 野中寿史, 増山繁, “特許明細書からの技術課題情報の抽出,” 人工知能学会論文誌, Vol.24, No.6, pp.531-540, 2009.
- [5] 高村大地, 奥村学, “施設配置問題による文書要約のモデル化,” 人工知能学会論文誌, Vol.25, No.1, pp.174-182, 2010.