

特許中のフローチャート画像からの説明文の自動生成

難波英嗣¹ 久保翔平¹ 福田悟志¹

概要: 我々は、幅広い分野の一連の手続きに関する知識をテキストから自動抽出し、それらを体系化することを目指している。その試みのひとつとして、本稿では、特許中のフローチャート画像からの説明文の自動生成に取り組む。説明文生成器の構築は、(1)フローチャート画像からの文字列の抽出、(2)機械学習用データの作成、(3)T5を用いた説明文生成器の構築、の3つの手順から構成される。本研究では、11,188件のフローチャート画像と説明文の対を用いて、画像からの説明文生成器を構築した。実験の結果、複雑な形状のフローチャートについては課題が残るものの、質の高い説明文が生成できることが確認できた。

キーワード: フローチャート, 文書生成, 文字認識, 特許

Automatic Generation of Explanatory Text from Flowchart Images in Patents

HIDETSUGU NANBA¹ SHOHEI KUBO¹ SATOSHI FUKUDA¹

Abstract: Our goal is to automatically extract knowledge about a set of procedures in a wide range of fields from text and systematize them. As one of our attempts, this paper addresses the automatic generation of explanatory text from flowchart images in patents. The construction of an explanatory text generator consists of three steps: (1) extraction of text strings from flowchart images, (2) creation of data for machine learning, and (3) construction of an explanatory text generator using T5. In this study, we constructed an explanatory text generator from images using 11,188 flowchart image-explanatory text pairs. Experimental results confirmed that the explanatory text generator can generate high-quality explanatory text, although some issues remain for flowcharts with complex shapes.

Keywords: flowchart, text generation, character recognition, patent

1. はじめに

ある特定の目的を達成するための一連の手続きを記したものを手順テキストと呼ぶ。我々は、幅広い分野の一連の手続きに関する知識をテキストから自動抽出し、それらを体系化することを目指している[1]。その試みのひとつとして、本稿では、特許中のフローチャート画像からの説明文の自動生成について述べる。

フローチャート画像の説明文を自動生成する上で、特許の要約と選択図に着目する。選択図とは、発明の概要を速やかにかつ的確に把握できる図のことであり、通常は、要約の内容の理解に必要と思われる図を出願人が特許中の図表の中から選択する。我々の過去の研究において、特許中の画像がフローチャートであるか否かを自動判別する分類器を構築しており、再現率0.9422、精度0.9625を達成している[1]。この分類器を用いて、選択図がフローチャートのものだけを抽出すれば、フローチャートとその説明文(要約)の対を自動的に大量に生成することができる。さらに、この対を用いれば、機械学習を用いてフローチャート画像から説明文を自動生成するシステムを構築できると考えられる。

本論文の貢献は以下のとおりである。

- フローチャート画像とその説明文の11,188対を自動構築した。
- この対を用いて機械学習により、フローチャート画像から説明文を自動生成するシステムを構築した。

2. 関連研究

2.1 フローチャートの解析

近年では、myExperiment や SHIWA など、フローチャートを共有するサービスがはじまっており、これに伴い、あるフローチャートと類似するものを検索する技術の需要が出てきている。Starlingerら[2]は、あるフローチャートと別のフローチャートがどの程度似ているのかを算出するため、2つのフローチャート間の対応関係を取る様々な手法について検討している。

フローチャート画像の解析における関連研究プロジェクトとして CLEF-IP がある[3]。CLEF(Conference and Labs of the Evaluation Forum)とは、ヨーロッパを中心に行われている情報検索に関するワークショップであり、CLEF-IP は特許を対象としたタスクのことを指す。このタスクは実験レベルだけではなく、現実の課題に即した検索タスクのた

¹ 中央大学
Chuo University

めのデータセットを提供することで、多言語及びマルチモーダル特許検索タスクの研究の促進を図っている。CLEF-IP では図形を認識し、フローチャートの要素となるテキスト、エッジ、ノードを検出しフローチャートの認識を行っている。

フローチャート画像を認識するこの他の研究として、Herrera-Cámara のものがある[4]。この研究では、手書きフローチャート画像を解析し、C 言語のソースとして出力する手法を提案している。Sethi らは、深層学習関連の論文中の図表画像からフローチャートを識別し、さらにフローチャートを解析することで、Keras と Caffe でソースを出力するシステムを構築している[5]。本研究では、フローチャート画像を入力とし、自然言語文としてその説明文を出力する点が異なる。

2.2 図表からのテキストの生成

Chart to Text とは図表から得られる重要な事柄を自然言語文として生成するタスクのことである。Zhu ら[6]はこの問題に対し、散布図、折れ線グラフ、横棒グラフ、縦棒グラフの4種類の図に対して、とそれに対応する分析的な記述を生成するシステム AutoChart を構築している。生成されたテキストと図表を人間と機械による評価を行ったところ、生成されたテキストは情報量が多く首尾一貫しており、対応する図表に関連していることを実証している。Obeid ら[7]は Chart to Text においてチャートの重要な個所の位置情報を埋め込むことで要約の質の向上を図っている。

Tan[8]らは科学系の学術論文に含まれる円グラフや棒グラフ、折れ線グラフ、Kantharaj ら[9]は主に経済や市場、社会問題をテーマにした棒グラフと折れ線グラフをもとに T5 や BART、GPT2 などの生成器を使用し図表から文を生成している。本研究ではグラフではなく、フローチャート画像を入力とし、そこからの説明文の自動生成を目指す。

3. フローチャートからの説明文の自動生成

3.1 説明文生成の手順

説明文生成器の構築は以下の3つの手順から構成される。

- (手順1) フローチャート画像からの文字列の抽出
 - (手順2) 機械学習用データの作成
 - (手順3) T5 を用いた説明文生成器の構築
- 各手順について、3.2~3.4節でそれぞれ説明する。

3.2 フローチャート画像からの文字列の抽出

Google Cloud Vision (<https://cloud.google.com/vision>)の文字認識機能を用いてフローチャート内の文字列を抽出する。フローチャートと文字認識結果の例を、図1および図2にそれぞれ示す。

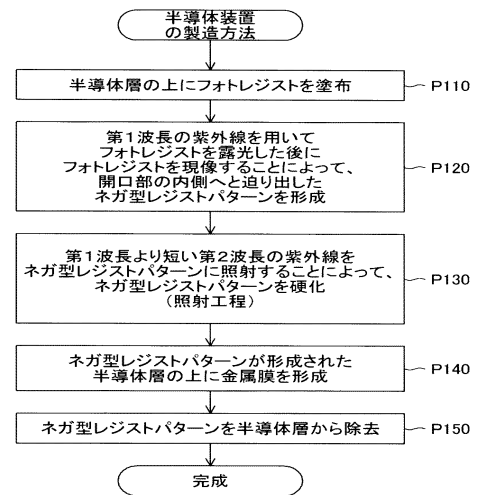


図1: 特許に含まれるフローチャートの例 (特開 2017-106955)

半導体装置¥n の製造方法¥n 半導体層の上にフォトレジストを塗布¥n 第1波長の紫外線を用いて¥n フォトレジストを露光した後に¥n フォトレジストを現像することによって、¥n 開口部の内側へと迫り出した¥n ネガ型レジストパターンを形成¥n 第1波長より短い第2波長の紫外線を¥n ネガ型レジストパターンに照射することによって、¥n ネガ型レジストパターンを硬化 ¥n(照射工程)¥n ネガ型レジストパターンが形成された¥n 半導体層の上に金属膜を形成¥n ネガ型レジストパターンを半導体層から除去 ¥n 完成 ¥n P110¥nP120¥nP130¥nP140¥nP150

図2: 図1の画像の文字認識結果

3.3 機械学習用データの作成

大量のフローチャート画像からの文字認識結果と説明文の対を用いて、機械学習により説明文生成器を構築する。その際、文字認識結果と説明文の内容が大きく異なるものは学習用データとして不適切であると考え除外する。具体的には、フローチャート画像の文字認識結果と説明文の類似度を、以下に示すゲシュタルトパターンマッチングで計算し、しきい値以上の対のみを学習に用いる。

$$D_{ro}(S_1, S_2) = \frac{2K_m}{|S_1| + |S_2|} \quad (0 \leq D_{ro} \leq 1)$$

K_m : マッチした文字数

S_1, S_2 : 比較する2つの文字列

$|S_1| + |S_2|$: 比較する文字列の長さの和

ここで、ゲシュタルトパターンマッチングは $D_{ro}(S_1, S_2) \neq D_{ro}(S_2, S_1)$ であるため、 $D_{ro}(S_1, S_2) \cdot D_{ro}(S_2, S_1)$ を計算した上でしきい値を決定する。なお、計算には Python の difflib ライブラリを用いる。また、特許画像がフローチャートで

表 1：抽出されたフローチャート画像と説明文対の数

| | 2008年 | 2009年 | 2010年 | 2011年 | 2013年 | 2014年 | 2016年 | 2017年 | 2019年 | 合計(件) |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| 0.1(類似度) | 2211 | 2005 | 1904 | 1776 | 1847 | 1600 | 1568 | 1831 | 2083 | 14742 |
| 0.05 | 6401 | 6035 | 5846 | 5175 | 5166 | 4619 | 4311 | 4817 | 4995 | 42370 |
| 0.03 | 10267 | 9797 | 9392 | 8314 | 8195 | 7454 | 6966 | 7656 | 7870 | 68041 |
| 0.01 | 16784 | 16290 | 15697 | 13745 | 13722 | 12491 | 11788 | 12717 | 13321 | 113234 |
| 全数 | 22170 | 21890 | 21021 | 18515 | 18771 | 17413 | 16563 | 17709 | 18169 | |

あるかどうかは、過去の成果[1]を用いて判定する。

3.4 T5 を用いた説明文生成器の構築

手順 2 で得られたデータと言語モデルの一種である T5 を用いて説明文生成器を構築する。T5 の入出力は、例えば、図 1 のフローチャート画像に関しては、入力が図 2、出力が図 3 となる。

半導体装置の製造方法は、半導体層の上にフォトレジストを塗布する工程と；第 1 波長の紫外線を用いてフォトレジストを露光した後にフォトレジストを現像することによって、開口部の内側へと迫り出したネガ型レジストパターンを、形成する工程と；第 1 波長より短い第 2 波長の紫外線をネガ型レジストパターンに照射することによって、ネガ型レジストパターンを硬化させる照射工程と；照射工程を行った後、ネガ型レジストパターンの開口部から露出する半導体層の上に、ニッケル (Ni) から主に成る金属膜を形成する工程と；ネガ型レジストパターンを半導体層から除去する工程とを備える。

図 3：図 1 の画像に対応する説明文

参考までに、しきい値が低いものの例を図 4～6 に示す。

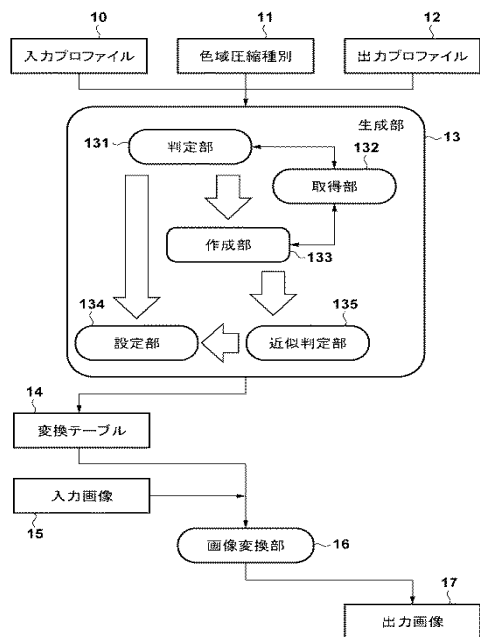


図 4：フローチャートの例 (特開2017-225095)

入力プロファイル 14、10、15、131、134、変換ル
設定部 11、入力画像 色域圧縮種別 13、判定部 13、作成
部 11、画像変換部 133、取得部 13、近似判定部
出力 77 化 生成部 132、135、16、12、13、17、出力画像

図 5：図 4 の画像の文字認識結果 (低類似度)

第 1 の色変換プロファイルにおけるチャンネル数と、第 2 の色変換プロファイルにおけるチャンネル数と、が一致するか否かを判断する (第 1 の判断)。第 1 の色変換プロファイルが規定する色再現特性と、第 2 の色変換プロファイルが規定する色再現特性と、が一致するか否かを判断する (第 2 の判断)。第 1 の判断及び第 2 の判断において一致すると判断した場合には、第 1 の色変換プロファイル及び第 2 の色変換プロファイルに基づく色変換用のテーブルの代わりに、色変換が省略可能であることを示す情報を出力する。

図 6：図 4 の画像に対応する説明文 (低類似度)

特許公開年およびしきい値ごとに抽出されたフローチャート画像と説明文数を表 1 に示す。実際に T5 を用いて学習すると、学習に非常に時間がかかることが判明した。そこで、次節で述べる実験では、6 年分の公開公報でしきい値 0.1 以上のものを利用する。

4. 実験

提案手法の有効性を調べるため、実験を行った。

4.1 実験条件

【使用データ】

2019、2016、2014、2013、2011、2010 年の特許のうち、選択図にフローチャートが含まれ、かつゲシュタルトパターンマッチングによる類似度が 0.1 以上の特許 11,188 件を実験に用いた。このうち 90% を訓練用データ、残りを検証用データと評価用データとした。

【ハイパーパラメータ】

T5 のハイパーパラメータは以下のとおりである。

```
# 学習に用いるハイパーパラメータを設定する
args_dict.update({
    "max_input_length": 280, # 入力文の最大トークン数
    "max_target_length": 256, # 出力文の最大トークン数
    "train_batch_size": 8, # 訓練時のバッチサイズ
    "eval_batch_size": 8, # テスト時のバッチサイズ
    "num_train_epochs": 6, # 訓練するエポック数
})
args = argparse.Namespace(**args_dict)

train_params = dict(
    accumulate_grad_batches=args.gradient_accumulation_steps,
    gpus=args.n_gpu,
    max_epochs=args.num_train_epochs,
    precision= 16 if args.fp_16 else 32,
    amp_level=args.opt_level,
    gradient_clip_val=args.max_grad_norm,
)
```

【評価方法】

システムの出力と特許要約との類似度を、以下の尺度を用いて評価する。

- ROUGE-N: 最もベーシックな指標で、N-gram 単位での一致度合いをとる手法である。今回は N=1, 2 で評価する。(https://github.com/pltrdy/rouge)
- ROUGE-L: 生成した要約と人出で作成した要約とで一致する最大のシーケンスを評価するものである。
- BERTScore [10]: 言語モデル BERT を利用した自動評価指標。事前学習された BERT から得られるベクトル表現を利用して、テキスト間の類似度を計算する。

4.2 実験結果

ROUGE-1、ROUGE-2、ROUGE-L、BERTScore のそれぞれに対し Recall、Precision、F1-score を求めた結果を表 2 に示す。また、図 7 のフローチャート画像に対する文字認識結果、T5 で生成された文、人手による特許要約(正解)を図 8~10 にそれぞれ示す。

表 2: T5 を用いた要約器の評価結果

| | Recall | Precision | F1-score |
|-----------|--------|-----------|----------|
| ROUGE-1 | 0.47 | 0.72 | 0.55 |
| ROUGE-2 | 0.26 | 0.46 | 0.32 |
| ROUGE-L | 0.41 | 0.64 | 0.49 |
| BERTScore | 0.74 | 0.77 | 0.75 |

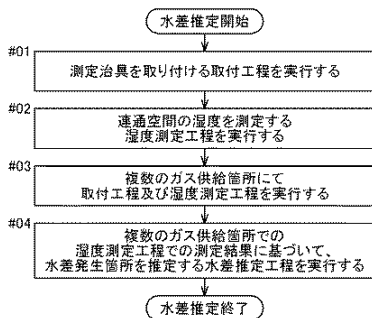


図 7: 説明文生成の対象画像の例 (特開 2019-178957)

#01#02#03#04 水差推定開始測定治具を取り付ける取付工程を実行する空間の湿度を測定する湿度測定工程を実行する複数のガス供給箇所にて取付工程及び湿度測定工程を実行する複数のガス供給箇所での湿度測定工程での測定結果に基づいて、水差発生箇所を推定する水差推定工程を実行する水推定終了

図 8: 図 7 の画像の文字認識結果

測定治具を取り付ける取付工程と、複数のガス供給箇所にて取付工程及び湿度測定工程を実行する湿度測定工程と、複数のガス供給箇所での湿度測定工程での測定結果に基づいて、水差発生箇所を推定する水差推定工程と、を含む。

図 9: 図 7 の画像から自動生成された説明文

ガス供給箇所へ供給されるガスの流量を測定するガスメータの一次側へ連通接続する連通空間を有すると共に当該連通空間の湿度を測定する湿度測定部を有する水差推定治具を取り付ける取付工程と、湿度測定部にて連通空間の湿度を測定する湿度測定工程と、湿度測定工程にて測定される湿度に基づいて、ガス管に水差しが発生しているか否かを推定する水差推定工程とを実行する。

図 10: 図 7 の画像に対する説明文(正解)

4.3 考察

図 7 に示すようなフローチャート内に分岐がない単純な形状の場合には、良好な解析結果が得られた。一方で、失敗例には大きく 2 つのパターンがあることが確認された。

【失敗パターン 1】

図 11 に示すような複雑な形状のフローチャートでは質の低い説明文が生成される傾向が見られた。なお、図の赤枠は、説明のために著者が加えたものである。

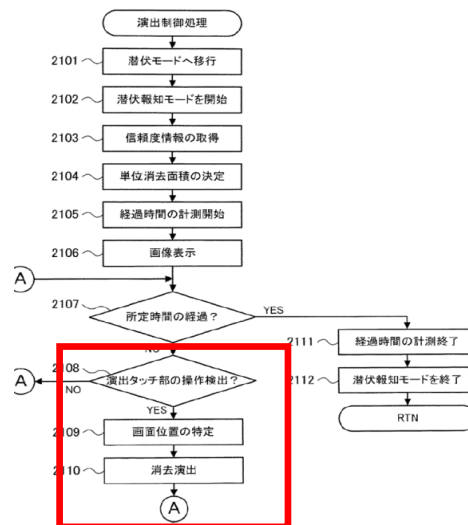


図 11: 分岐があるフローチャートの例

図 11 から生成された説明文を図 12 に示す。

潜伏モードへ移行し(s2110)、信頼度情報を取得し(s2110)、変倍率を決定し(s2110)、経過時間の計測を開始する(s2112)。そして、画像表示部が所定時間経過しているか否かを判定し(s2112)、所定時間が経過すると(s2112 にて yes)、潜伏報知モードを終了する(s2112)。潜伏報知モードを終了すると(s2112 にて yes)、潜伏報知モードを終了する。

図 12：図 11 の画像から自動生成された説明文

図 12 を見ると、全体的に s2110 などのステップ ID が説明文と対応していないが、これは、今回は、図中の各文字列の座標を全く考慮していないためである。1 つ目の条件分岐の個所は「所定時間が経過すると(s2112 にて yes)、潜伏報知モードを終了する。」のように、ステップ ID(s2112)以外は正しい文が生成されている。一方で、図 11 の赤枠内は説明文に含まれていない。現在は、Google Cloud Vision の文字認識結果で出力された文字列をそのまま T5 の入力としているが、今後は文字列の座標情報を考慮し、適切に並び替えるなどの前処理が必要になると考えられる。

【失敗パターン 2】

機械翻訳やテキスト要約などのニューラルベースの文生成で観測される無意味な文字列の繰り返しが出来た説明文にも見られた。図 13 の説明文の例を見てもわかるとおり、1 段落目に「停止保持ブレーキを解除した状態で自動緊急ブレーキを解除する」という文字列が繰り返し出ている。正しい説明は図 14 に示すとおりである。

自動緊急ブレーキ制御装置は、停止保持ブレーキを解除する停止保持ブレーキ解除部と、停止保持ブレーキを解除した状態で自動緊急ブレーキを解除する自動緊急ブレーキ解除部と、停止保持ブレーキを解除した状態で自動緊急ブレーキを解除する自動緊急ブレーキ解除部と、停止保持ブレーキを解除した状態で自動緊急ブレーキを解除する自動緊急ブレーキ解除部と、を備える。

方法は、仮想空間データを特定するステップと、仮想視点を特定するステップと、hmd の動きに応じて視野画像を更新するステップと、不整合事象が検出された場合には、仮想空間のコンテキストに応じた視覚効果を決するステップと、視野画像に視覚効果を付与するステップと、視野画像への視覚効果の付与をキャンセルするステップとを含む。

図 13：無意味な文字列の繰り返しを含む自動生成された説明文の例

自動緊急ブレーキによって交差点内で停車した場合、s305 で 2 次ブレーキの禁止、s310 でハザードランプの点滅、s320 でアイドリングストップの禁止を実行する。s325 で停車したと判定した後、s330 で発車しても安全であると判定すると、s340 で停止保持ブレーキを解除する。

図 14：図 13 に対応する説明文の正解例

5. おわりに

本研究では、特許から 11,188 件のフローチャート画像とその説明文の対を抽出し、このデータを用いて T5 でフローチャート画像の説明文を自動生成するシステムを構築した。実験の結果、複雑な形状のフローチャートについては課題が残るものの、質の高い説明文が生成できることが分かった。今後は、フローチャート内の文字列をそのまま使うのではなく、画像内の各文字列の位置情報を考慮することで、条件分岐を複数含むなど、複雑な形状のフローチャートに対しても適切な説明文を生成できるよう検討する。

謝辞 本研究は JSPS 科研費 22K12154 および 20H04210 の助成を受けたものである。

参考文献

- [1] 樊エイブン, 福田悟志, 難波英嗣, 特許中の画像とテキストを用いた手順オントロジーの構築, 情報処理学会研究報告第 145 回情報基礎とアクセス技術研究会, 2022.
- [2] Starlinger, J., Brancotte, B., Cohen-Boulakia, S., and Leser, S., Similarity Search for Scientific Workflows, Proceedings of the VLDB Endowment, Vol. 7, No. 12, pp.1143-1154, 2014.
- [3] Piroi, F., Lupu, M. and Hanbury, A., Overview of CLEF-IP 2013 Lab Information Retrieval in the Patent Domain, Information Access Evaluation. Multilinguality, Multimodality, and Visualization. CLEF 2013. Lecture Notes in Computer Science, Vol. 8138. Springer, Berlin, Heidelberg, 2013.
- [4] Herrera-Cámara J.I., FLOW2CODE - From Hand-drawn Flowchart to Code Execution, Master Thesis, Texas A&M University, 2017.
- [5] Sethi, A., Sankaran, A., Panwar, N., Khare, S., and Mani, S., DLPaper2Code: Auto-generation of Code from Deep Learning Research Papers, Proceedings of the 32th AAAI Conference on Artificial Intelligence, 2018.
- [6] Zhu, J., Ran, J., Lee, R.K., Li, Z., and Choo, K., AutoChart: A Dataset for Chart-to-Text Generation Task, Proceedings of the International Conference on Recent Advances in Natural Language Processing, pp. 1636-1644, 2021.

- [7] Obeid, J. and Hoque, E., Chart-to-Text: Generating Natural Language Descriptions for Charts by Adapting the Transformer Model, Proceedings of the 13th International Conference on Natural Language Generation, pp. 138-147, 2020.
- [8] Tan, H., Tsai, C., He, Y., and Bansal, M., Scientific Chart Summarization: Datasets and Improved Text Modeling, Proceedings of the AAAI-22 Workshop on Scientific Document Understanding, 2022.
- [9] Kantharaj, K., Leong, R. T., Lin, X., Masry, A., Thakkar, M., Hoque, E., and Joty, S., Chart-to-Text: A Large-Scale Benchmark for Chart Summarization, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, pp. 4005-4023, 2022.
- [10] Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., and Artzi, Y., BERTScore: Evaluating Text Generation with BERT, arXiv:1904.09675 [cs.CL], 2019.