NLP-Driven Automation of Business Process Diagrams

Dustin Hooks and **Bashayer Alharbi** and **Luna McBride**
University of Colorado Boulder
Boulder, CO, USA
dustin.hooks@colorado.edu
bashayer.alharbi@colorado.edu
luna.mcbride@colorado.edu

Abstract

This project looks to automate the notes to diagram workflow in business process discovery and definition by using natural language processing models to more effectively extract important items from provided textual dialogs. We looked at using a summarization pipeline to remove some of the noise while also making the data easier to process. This result was then sent further into another model to attempt to pull out the important process information. This was done in an attempt to not only make the overall workflow more efficient, but to also reduce bias and mitigate the risk of miscommunication during the discovery process. However, it was found that utilizing full, non-summarized dialogs with targeted prompt engineering provided the best results.

1 Introduction

The objective of this project is to take a verbal/textual description of a complex business process and generate a formal representation of that business process which could be used for diagramming and implementation of that business process. Addressing this task was intended to involve the following:

- Use of a pre-built summarization pipeline to generate a higher-level summary of a textual business process description.
- Use of named-entity recognition to realize the distinct process steps represented in the summary.
- Use of sequence recognition to realize the sequential dependency of the process steps.

After some trial and error, it was found that prompt engineering, use of a Flan Large model, and using long dialogs as input provided the best results.

2 Motivation

Business Process Management and Automation is a very impactful but time-consuming endeavor. One of the primary tasks performed by consultants which work in this area is interviewing their clients to describe and recognize their business processes which then inform the implementation and automation of those business processes via software systems. Much of the human effort required is spent, over multiple days and weeks, as follows:

1. Perform multiple client interviews
2. Take sufficient notes from the interviews
3. Review and recognize the business processes from the notes
4. Diagram business process flows representing the business processes
5. Review and revise the business process flow diagrams with the client
6. Implement business process flows into software systems(s)

We feel that the human effort required and risk of miscommunication could be significantly reduced by automating steps 2-4 and 6. To support that automation, our project will focus on using NLP techniques to automate step 3. We will highlight existing work that has been done on various aspects of this process in Related Work.

3 Related Work

3.1 Extracting Business Process Models using Natural Language Processing (NLP) Techniques

This paper about Extracting Business Process Models (Sintoris and Vergidis, 2017) discusses key issues that need to be addressed when analyzing the text data (the biggest being relevance and references) as well as the processes the authors intend to go through to build a model that constructs a business process model. It also points out different words that could help signal a change in the model, such as "in parallel" and "if." These are important considerations to keep in mind in our project as well, as seeking out these terms and eliminating irrelevant noise will make this process much easier.

3.2 From Text to Visual BPMN Process Models: Design and Evaluation

The Visual BPMN Process Models (Ivanchikj et al., 2020) paper introduces a Business Process Model and Notation (BPMN) Sketch tool that leverages NLP aiming to convert textual description to BPMN diagrams. It shows the potential of NLP in facilitating BPM automation. It highlights the ability to produce real-time BPMN diagrams per the stakeholders description of the business process. Our project, however, focuses on automating interview notes analysis to identify business processes.

3.3 Automated Business Process Discovery from Unstructured Natural-Language Documents

This Automated Business Discovery (Chambers et al., 2020) paper utilized distinguishing issues by topic and clustering to extract events from unstructured documents. It showed potential in pulling out entities and sorting them into an easier to digest format, using emails for IT topics as their example. This could be a useful idea to grab for this project to make the process easier on the end user. It also would make the information easier to diagram, but we may not want to distinguish topics in the same way as the paper shows depending on our data.

3.4 SummEval: Re-evaluating Summarization Evaluation

This Summarization Evaluation (Fabbri et al., 2021) paper shows various models and metrics, along with how they scored in different categories. It shows the value of many major metrics on the scene for this task in order to make a better decision. Our project intends to have a summarization pipeline before the main model to make the data easier to handle, so having a list of metrics and what they are good at showing will prove invaluable. There still will need to be some human intervention to see if the summarization portion is creating summaries we would like to use, but being able to guide these decisions by metrics based on factors we decide are more important for our task will make for a better overall outcome.

4 Methodology

Our experiments centered around a small set of conversations generated based on finished business flows provided by a representative. We had to work backward from those flows using ChatGPT to get our actual conversation data. We then experimented with both these conversations and summarized versions in a Flan T5 model, utilizing traditional model training methods and prompt engineering to test the model's accuracy. Human verification and Rouge were used to see how the model did for our given problem.

4.1 Synthetic Dataset Generation

The flows we obtained made it so we had an idea of how people in industry expect the output to be. This may have been the best way to verify what we were doing, but not having the original conversations that turned into these flows made data extraction difficult. This is where we decided to use ChatGPT, which was able to create conversations out of the flows given to us. This selection came out of necessity. Even if we could have found our own data to train on outside of these flows, we would have needed to generate the conversations that led to the flows either way in order to use them for any sort of verification or otherwise.

Figure 1 is an example of a flowchart that is drawn out of a conversation with a client. We used this flowchart to generate a synthetic conversation between a client and a company representative. For simplicity purposes, we further converted the flowchart to a text-based process flow, which can be found in appendix A along with an example of the generated conversation.

4.2 Models

To handle the ideas we had for this project, we needed both a model to summarize the conversations and a model to actually test for feasibility. The summarization factor itself was not the main issue we wanted to look into, so we grabbed a BART model from Huggingface's model library that was already fine-tuned on conversations. This model by user Mr-Vicky-01 can be found here (Mr-Vicky-01, 2023).

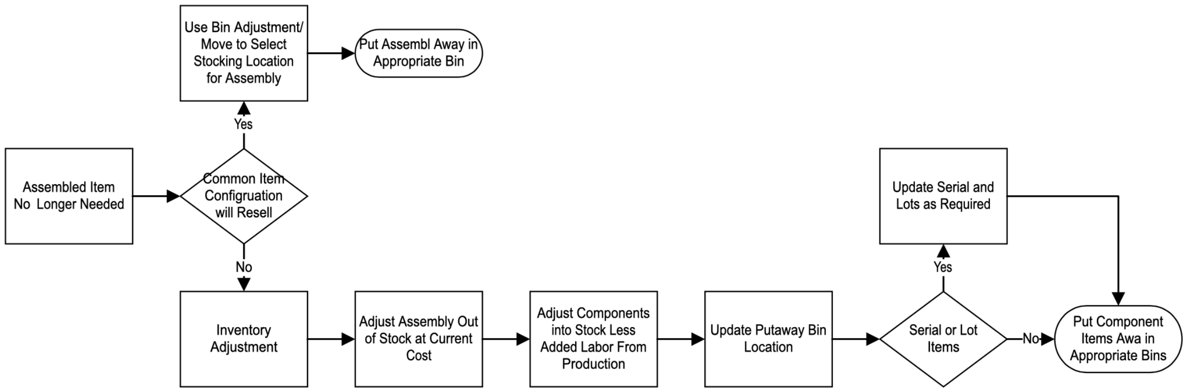

Figure 1: A Business Process Diagrams Example

As for the post-summarization model, we decided to use Google's Flan T5. This was due to its strong reputation for few-shot learning. It was used in both a traditional training context and a prompt engineering context to test the strengths and weaknesses of the model. This model would have normally been considered out of scope due to its size, but we had so little data that we could not just train a smaller GPT model and call it a day. This is where Google Colab's higher resource availability shone through, allowing for higher memory usage to fit the Flan T5 model and stronger GPUs to make these experiments actually feasible to run.

Flan T5 also had an advantage that really helped us when it came to our longer entries, being that its 512 token "limit" was not an actual barrier. The model would take in texts much longer than that and process them well. This made it a better choice than any longformer implementations we also tried looking into. Despite the large size of T5, it was still the smaller choice compared to the vast sizes that come with longformers.

5 Results

5.1 Training

The normal training run did not work well for this model, as expected. The primary reasons it did not work, however, actually were not caused by the reasons that would typically be expected. The training actually ran pretty fast on Colab, so we thought we could potentially train a model with our few examples and at least have the format included in its memory in hopes it would provide some benefit down the road. The testing step, however, showed off a huge problem with this method.

The model.generate() step on the T5 typically did not add much memory overhead to the runtime. The GPU memory would go up a bit, but not in any major way. The post-training model, however, took up a great deal of memory for its generate step. So much, in fact, the GPU with the highest memory in Google Colab (A100 at 40GB) would often overload if significant memory refactoring was not done before the test run. The test was only set to generate based on one example. This made the trained model unfeasible to use at all, as it locked us into the more expensive GPU for no good reason. This made our attempts at regular training methods a bust.

5.2 Prompt Engineering

The next path we went down was using prompt engineering to push the model into making the outputs we wanted. Our first attempts at this methodology started when only four of the thirteen data points were generated, so the original tests included three examples to prime the model on and a fourth to see the results. In terms of human readability, this actually worked quite well. The model did not pick up on newline characters or the arrows we had for human readability, but it quickly picked up on the business wording and the A: ... B: ... structure used to lay out steps in the flow. The next iteration of this idea with all thirteen data points, however, proved that giving the model the full conversations instead of just the summaries lead to better results.

5.3 Metrics

The previous iteration of the prompt engineering used human readability and our opinions to judge the model. We decided to step it up and put numbers to our observations. We decided to run four experiments, each being a combination of a Flan T5 model (Large or XL) and a version of the data (summaries or full conversations.) These were run with a split of 10 entries given to the model with the final 3 entries being used as tests. These test entries included one easy one to the next flow, a medium-difficulty flow with some branching paths, and a hard-difficulty flow with multiple branching paths. RougeL F1 scores were used to compare the different runs. These results are shown in table 1.

These numbers may seem low, but the comparative power they give show that full conversations in the Large model were best for this use case.

6 Discussion

The lower numbers on the Rouge prove it is not the best metric for this cause. The flows look good from a human standpoint. It is very possible that the metric is considering where the words would line up compared to one another, which would make it fail for these flows from a judgement perspective. It works just fine from a comparison perspective, however.

There were also plans to discuss the difference in time taken for each of those runs, but that proved to be GPU dependent. The two large Flan model tests ran in 2 minutes for summaries and 22 minutes for conversations on the strongest A100 GPUs, but took 14 and 17 minutes respectively on the relatively weaker L4 GPUs. We decided to err on the side of the cheaper GPU and not consider the time difference between the conversation data and the summary data too significant. Industry implementations would prefer the more cost-effective versions anyway, so it would fit the domain well to consider the lower strength GPUs as the better baseline. Further expansion on these ideas would need to consider this fundamental difference, however.

7 Conclusion and Future Work

The primary objective of our work is to solve a human-intensive business problem by completely automating the generation of accurate and complete business process flow diagrams directly from human audio dialog. This paper presents a portion of that work in which we address an intermediate portion of this problem by translating textual dialog to textual flows. We have demonstrated that we can break this into smaller steps and use appropriate NLP techniques at each of those steps to provide the components required for a pipeline that could successfully translate from human dialog to generated flow diagrams.

We have not covered here all the required transformations required to fully automate this process. Future work would include investigating contextual understanding and translation of human audio dialog of business process descriptions, the generation of process flow diagram images from textual flows, and ultimately wrapping this all into a software tool that would allow dynamic and interactive flow diagram generation based on human verbal conversation and commands.

References

Alexander J. Chambers, Amy M. Stringfellow, Ben B. Luo, Sophie J. Underwood, Tony G. Allard, Ian A. Johnston, Sarah Brockman, Leslie Shing, Allan Wollaber, and Courtland VanDam. 2020. Automated Business Process Discovery from Unstructured Natural-Language Documents. In Adela Del Río Ortega, Henrik Leopold, and Flávia Maria Santoro, editors, *Business Process Management Workshops*, volume 397, pages 232–243. Springer International Publishing, Cham. Series Title: Lecture Notes in Business Information Processing.

Alexander R. Fabbri, Wojciech Kryściński, Bryan McCann, Caiming Xiong, Richard Socher, and Dragomir Radev. 2021. SummEval: Re-evaluating Summarization Evaluation. In *Transactions of the Association for Computational Linguistics*, volume 9, pages 391–409. MIT Press Direct.

Ana Ivanchikj, Souhaila Serbout, and Cesare Pautasso. 2020. From text to visual BPMN process models: design and evaluation. In *Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*, pages 229–239, Virtual Event Canada. ACM.

Mr-Vicky-01. 2023. BART Fine-tuned for Conversational Summarization on Hugging Face. Hugging Face Model Hub.

Konstantinos Sintoris and Kostas Vergidis. 2017. Extracting Business Process Models Using Natural Language Processing (NLP) Techniques. In *2017 IEEE 19th Conference on Business Informatics (CBI)*, pages 135–139, Thessaloniki, Greece. IEEE.

Test Type	Easy Test	Medium Test	Hard Test
Flan Large + Summaries	0.0458	0.0283	0.0119
Flan XL + Summaries	0.0228	0.0174	0.0074
Flan Large + Conversations	0.0431	0.0558	0.1454
Flan XL + Conversations	0.0075	0.0102	0.0172

Table 1: Model Size and Conversation Type Comparison

A Dataset Generation

A.1 Generated Conversation

Below is a subset of the synthetic Conversation generated by ChatGPT based on figure 1.

```
"Client (C): Hello, I'd like to revisit
    our last conversation regarding the
    custom industrial printers. I've
    gathered more specific requirements
    from our teams.
 Company Representative (CR): Good
    morning! I'm glad we're
    reconnecting on this. Our Solutions
     Department is ready to cater to
    your detailed needs. L e t s
    discuss these specifics so we can
    update your order precisely.
 Client: The production department
    insists on printers that can
    sustain continuous 24/7 operation
    without overheating. They also need
     to be compatible with our existing
     manufacturing software.
 CR: For continuous operation, we can
    integrate enhanced cooling systems
    into your printers. Compatibility
    with your software will be
    addressed by updating the printers'
     firmware to sync with your
    manufacturing systems.  I ll  note
    these requirements for our order
    entry update.
 Client: Can we also explore more
    durable materials for the printer
    components to prevent wear and tear
    ?
 CR: Absolutely. We have access to
    advanced polymers that offer
    greater durability without
    compromising printer performance.
    These materials can be included in
    the component picking process by
    the Solutions Department.
 Client: That's essential for us. And
    with regard to environmental
    conditions, they'll be exposed to
    dust and debris.
 CR:  I ll  recommend printers with
    enclosed designs to protect against
     debris. They can be fitted with
    dust filters that are easy to clean
     and replace.
 Client:  W e d  need a detailed
    demonstration of these features.
    Can that be arranged before
    finalizing the order?
CR: Certainly. Once the component
    picking is complete, we can arrange
     a demonstration to showcase how
    these features work and the
    benefits they offer.
Client: I'm also thinking about the
    ease of maintenance. Can we ensure
    that the parts are easily
    accessible for our technicians?
CR:  T h a t s  a valid concern.  W e ll
    design the printers with
    maintenance in mind. Components
    will be modular for easy access and
     quick replacement, ensuring
    minimal downtime for your
    production.
Client: That's great. And about the
    additional hardware for Wi-Fi and
    Bluetooth capabilities, how will
    that affect our cost?
CR: When we add new hardware components
    , t h e r e s  usually an additional
    cost. However, if  w e re  replacing
     less efficient parts, we'll deduct
     their value. Should the value
    additions lead to a net negative,
     w e ll  adjust the process price
    accordingly.
Client: Could we get a cost-benefit
    analysis of adding these new
    components versus the potential
    increase in productivity?
CR: Absolutely. Part of our service is
    to provide a comprehensive cost-
    benefit analysis. This will help
    you make an informed decision about
     the additions.
Client: And suppose the final cost
    exceeds our budget?
CR: In that case, we will revisit the
    components and features to see
    where adjustments can be made
    without impacting performance. Our
    goal is to find the right balance
    between functionality and cost.
Client: Once we finalize the components
     and the cost, what is the
    confirmation process?
CR: After updating the production order
     with the new components and
    adjusted costs, our quality
    assurance team will review the
    changes.  T h e y ll  then change the
    'Confirm Pick' status to 'Confirmed
    ' to ensure that the order reflects
     all your requirements accurately.
Client: And if something is amiss?
```

```
CR: If there are discrepancies, we'll
    address them immediately. We can
    adjust the order and go through the
     confirmation process again to
    ensure everything is perfect.
Client: When can we expect to see the
    first batch?
CR: Once the order is saved, we
    typically have a lead time of six
    to eight weeks for the first batch.
     But rest assured, we will expedite
     where possible and keep you
    informed every step of the way.
Client: I appreciate the detail. Please
     prepare a proposal outlining the
    entire process, including all the
    steps we discussed today.
CR: I'll include all the specifics:
    from the Solutions Departments
    initial alerts to saving the final
    order, along with a complete
    timeline and cost breakdown.
Client: Thanks for your comprehensive
    support. We're looking forward to
    enhancing our production
    capabilities with these printers.
CR: We're here to support your goals.
    You'll have the proposal soon, and
    I'm available for any further
    clarification. Lets make this
    project a success together."
```

A.2 Text-Based Process Flow

We converted the flowchart to a text based process flow to easily train our model.

```
A: Solutions Department Alerts CSR of
    Required Change
→ Proceed to B
B: Order Entry
→ Proceed to C
C: Update Component Item ID and Quantity
    per Assembly as Required
→ Are additional components being added?
    If yes, proceed to D. If no, go to
    E.
D: Add Components
→ Deduct Value of Added Components from
    Value and Add Process Prices
→ Does Value Add Process Line result in
    Negative Value? If yes, proceed to G
    . If no, go to F.
E: Increase Value Add Process Price by
    the Value of Removed Components
→ Proceed to F
F: Update Value Add Process Order Cost
    to Remaining Balance not Deducted
    from Value Add Process Price
→ Proceed to H
G: Update Value Add Process Price to $0
→ Proceed to H
H: Save
```

A.3 Summary

```
Client has gathered more specific
    requirements for the custom
    industrial printers. The production
    department insists on printers that
    can sustain continuous 24/7
    operation without overheating and
    are compatible with the existing
    manufacturing software. They need to
     be compatible with your software to
     sync with your manufacturing
    systems and durable materials for
    the printer components to prevent
    wear and tear. The materials can be
    included in the component picking
    process by the Solutions Department.
     The quality assurance team will
    review the changes and adjust the
    production order with the new
    components and adjusted costs. They
    will change the 'Confirm Pick'
    status to 'Confirmed'. If something
    is amiss, they will adjust the order
     and go through the confirmation
    process again to ensure everything
```

B Example Results

B.1 Full Conversation

```
A: Solutions Department Alerts CSR of
    Required Change Proceed to B B:
    Order Entry Update Component Item ID
     and Qty per Assembly as Required
    Are additional components needed? (
    if yes, go to C; if no, go to D) C:
    Add Components Deduct Value of Added
     Components from Value Add Process
    Price Does Value Add Process Line
    result in Negative Value? (if yes,
    go to E; if no, go to F) E: Increase
     Value Add Process Price by the
    Value of Removed Components Proceed
    to F E: Update Value Add Process
    Price to 0 Proceed to F F: Update
    Value Add Process Order Cost to
    Remaining Balance not Deducted from
    Value Add Process Price Save
```

B.2 Summary

```
A: Client Initiates Service Request B:
    Customize Service Define Service
    Scope Calculate Service Cost Is the
    cost acceptable to the client? (If
    yes, go to C; if no, go to D) C:
    Client Approves Cost Schedule
    Service Delivery Provide Service
    Follow-up with Client Is the client
    satisfied? (If yes, go to E; if no,
    go to F) D: Adjust Service Scope &
    Recalculate Cost Return to step C E:
     End Process F: Address Client
    Concerns Return to step C after
    adjustments
```