Automated Document Extraction for Financial Services:

Enhancing Efficiency and Accuracy

Automated Document Extraction for Financial Services: Enhancing Efficiency and Accuracy

Javier Tordable, Technical Director, Office of the CTO, Google
Haochen Zhou, Software Developer, Developer X, Google
Vishnu Kumar, Chief Financial Engineer, The Interface Financial Group

Abstract— This white paper explores the challenges faced by commercial lenders in integrating their lending software with clients' accounting systems in the financial services ecosystem. It presents a comprehensive solution that leverages advanced technologies to automate the extraction of structured data from unstructured financial documents. The integration of Document Understanding AI by Google and the proprietary algorithms developed by The Interface Financial Group enables efficient document processing and accelerates decision-making in the trade finance industry. This white paper provides a detailed explanation of each model used, their capabilities, and the collaborative efforts between The Interface Financial Group and Google Cloud. It also discusses the implementation of ensemble learning algorithms and the impact of these technologies in various vertical markets.

I. INTRODUCTION

In the financial services market, the integration of lending software or platforms with clients' accounting systems is a common challenge faced by commercial lenders. This integration is essential for financial service providers to efficiently collect and analyze data from different financial reports, including banking transactions, balance sheets, income statements, accounts receivable, accounts payable reports, and more. These data sets are crucial for real-time analysis and decision-making processes, enabling qualitative and quantitative credit assessments and approval determinations.

However, the integration of these systems poses difficulties, as some borrowers either lack accounting software with integration capabilities or are hesitant to grant third-party access to their internal data. Consequently, clients often resort to sending unstructured financial documents, such as bank statements and financial statements, through user interfaces that support document uploads. While Optical Character Recognition (OCR) technology is utilized to extract data from these documents, the results are often unreliable.

Document Intelligence, is an evolving research area focused on automatically reading, comprehending, and analyzing business documents. These documents are essential for the day-to-day operations of a company and include purchase orders, financial reports, business emails, sales agreements, vendor contracts, letters, invoices, receipts, and more.

Business documents exhibit diverse formats, presenting information in natural language and organized in various ways, such as plain text, multi-column layouts, tables, forms, and figures. The complexity arises from the variability of layouts and formats, the suboptimal quality of scanned document images.

II. Collaborative Efforts and Model Capabilities:

The data science teams at The Interface Financial Group and Google Cloud developed a robust solution that leverages Document Understanding AI and proprietary algorithms. Document Understanding AI has become invaluable for processing unstructured documents, addressing the challenges faced by lenders. The teams created classification algorithms that capture the layout and textual attributes of each important field on a financial document, including tables, forms, and figures.

IFG Data Science team created a powerful document extraction tool that demonstrates exceptional accuracy and reliability. By utilizing Google's Document Understanding AI and incorporating IFG's proprietary algorithms, the tool achieves an accuracy rate of 99+% in recognizing, classifying, and extracting fields of interest from unstructured invoices.

Building upon the success of the invoice recognition tool, IFG extended its efforts to extract data from a wider range of scanned documents, including balance sheets, income statements, accounts receivable, and accounts payable reports. The IFG-Google document extraction tool stands apart from conventional solutions by employing machine learning techniques instead of rigid templates. It categorizes new document types, regardless of their format, and recognizes over 100 fields of interest.

III. Testing Methodology and Model Performance:

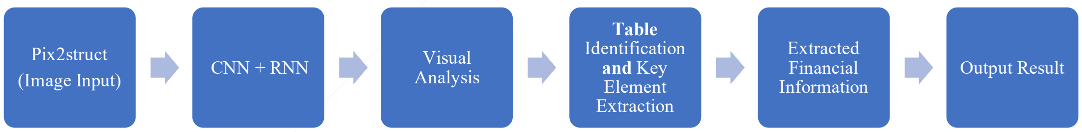
A meticulous testing approach was implemented to guarantee the precision and dependability of our machine learning models. Our process involved gathering extensive datasets that encompassed line-items extracted from balance sheets, profit & loss statements, accounts receivables, and accounts payables documents or images. To enhance the dataset's diversity, we augmented the initial data to create a diverse collection of 155,011 financial documents, incorporating various styles and formats.

To establish an effective model, we employed an 80/20 split strategy. This entailed allocating 20% of the data for training the machine learning models, while the remaining 80% was dedicated to model validation. This partitioning allowed us to effectively train our models on a substantial amount of data while ensuring adequate validation to assess their performance accurately.

IV. Ensemble Learning: Enhancing Accuracy and Predictive Models

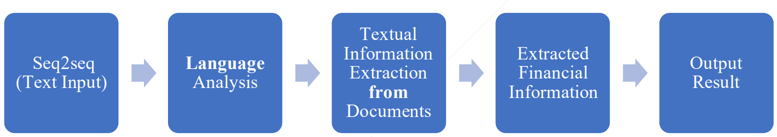
When applied to financial documents such as balance sheets, profit & loss statements, accounts receivables, and accounts payables, the mentioned models employ specific strategies to extract relevant information. Here's a detailed explanation of how each model works with these financial documents:

- Pix2struct utilizes a combination of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to process document images and extract structured representations. When applied to financial documents, the model first employs CNNs to analyze the visual content of the document images. It identifies key visual elements, such as tables, headings, numerical values, and textual information.



  Once the visual features are extracted, the model uses RNNs, often LSTM networks, to capture the sequential relationships within the document. This enables Pix2struct to understand the hierarchical structure of the document, including the arrangement of tables, sections, and their respective contents. The model segments and organizes the extracted information into a structured representation, making it easier to retrieve specific data points, such as values from balance sheets or profit & loss statements.

- Seq2seq models, applied to financial documents, focus on extracting textual information. These models take the document text as input and generate a structured representation or extract specific data points. Seq2seq models are trained on large amounts of financial text data, allowing them to learn the language patterns and conventions commonly found in balance sheets, profit & loss statements, and other financial documents.

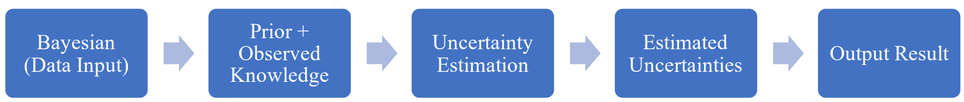


  The input sequence to the seq2seq model consists of the document text, while the output sequence can be a structured representation or specific extracted data points, such as net income or accounts receivables. The model's encoder part processes the input sequence and generates a fixed-length representation, capturing the important information. The decoder part then utilizes this representation to generate the desired output sequence.

  By learning from a diverse range of financial documents, seq2seq models can effectively extract relevant information, such as numerical values, financial ratios, or categorical data, from balance sheets, profit & loss statements, and other financial reports.

- Bayesian models, when applied to financial document extraction, provide probabilistic reasoning and uncertainty estimation. They consider prior knowledge and observe evidence to make accurate inferences and quantify uncertainty associated with the extracted data.

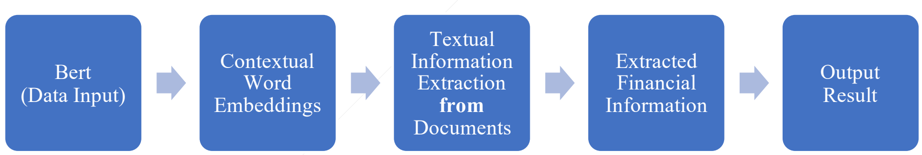
  In the context of financial documents, Bayesian models can be utilized to estimate uncertainties related to numerical values, ratios, or other extracted information. By modeling the distribution of these values and considering prior knowledge, the models can provide more comprehensive insights into the reliability and confidence of the extracted data points.



  For example, when extracting numerical values from financial documents, Bayesian models can estimate the uncertainty intervals or confidence intervals associated with those values. This information is valuable in financial analysis as it provides a range of possible values, allowing decision-makers to assess the reliability and make informed judgments.

- Large Language Models (LLMs) models, with their powerful contextual understanding of language, are applied to financial documents to extract relevant information and gain a deeper understanding of the textual content.

  When processing balance sheets, profit & loss statements, accounts receivables, and accounts payables, these models analyze the language used in these documents. They capture the relationships between words and sentences, considering the semantic context, and generate high-quality word embeddings.



  These embeddings can be used for various tasks, such as named entity recognition to identify key financial entities (e.g., company names, monetary values) or sentiment analysis to assess the overall sentiment expressed in the document.

By leveraging the contextual understanding provided by LLM models, financial documents can be analyzed more accurately, enabling the extraction of essential information, and supporting financial analysis processes.

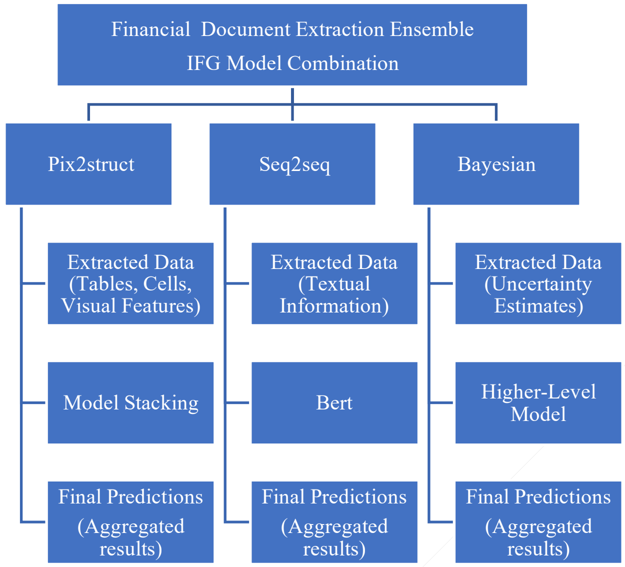
Each of these models employs specialized techniques to process financial documents, including images or textual data, and extract the desired information such as values, ratios, and structured representations. They leverage their unique strengths to enhance accuracy and reliability in financial document extraction and enable more effective financial analysis.

IFG's ensemble approach combines the strengths of these models to improve the accuracy and robustness of document extraction. Ensemble learning techniques involve training multiple models and aggregating their predictions to obtain a final result. In the context of financial document extraction, an ensemble approach can be employed as follows:

- The Model Combination: The output of each individual model (Pix2struct, seq2seq, Bayesian, LLM) is combined to create a consolidated result. For example, the extracted information from each model can be aggregated, and a voting mechanism or weighted averaging can be used to determine the final values. This approach leverages the complementary strengths of different models and increases the overall accuracy by considering multiple perspectives.

- Model Stacking: The outputs of different models can be used as features for a subsequent model to make the final prediction. For instance, the outputs of Pix2struct, seq2seq, Bayesian, and LLMs can be fed as inputs to a higher-level model, such as a decision tree or neural network. This higher-level model learns to make predictions based on the ensemble of outputs, combining the extracted information from multiple models in a more sophisticated manner. Model stacking allows for the capture of complex relationships between different features and can further enhance the accuracy and robustness of the document extraction process.

By leveraging an ensemble approach, the strengths of each individual model can be harnessed, mitigating the weaknesses of any single model. This enhances the accuracy, effectiveness, and generalizability of the document extraction process for financial analysis, leading to more reliable insights and decision-making.



The diagram shows the inclusion of model stacking, where the outputs of Pix2struct, Seq2seq, Bayesian, and LLM models are used as features for a higher-level model. The higher-level model learns from the ensemble of outputs and makes the final predictions, resulting in more accurate and reliable extracted from financial information, includes but not limiting to Client id, balance sheet total checking savings, total AR, total current assets, bs total assets, bs total AP, total current liabilities, total liabilities, retain earnings, bs net income, total equity, total liability & equity, total income, total expenses, net income, Profit & Loss Start Date, Profit & Loss End Date, Profit & Loss other income, Profit & Loss other expense.

Conclusion:

The collaboration between The Interface Financial Group and Google Cloud has resulted in an innovative solution for automating document extraction and structuring. By integrating Document Understanding AI with IFG's proprietary ensemble algorithms, they can accurately extract structured data from unstructured financial documents. Additionally, the implementation of ensemble learning algorithms and expanded NLP techniques further enhance accuracy and versatility.

These advanced technologies allow The Interface Financial Group to streamline their processes, accelerate decision-making, and reduce manual efforts. Thus, creating a significant advancement in document intelligence and setting a new standard for automated document processing in the financial services industry.