Enhancing Named Entity Recognition for Agricultural Commodity Monitoring with Large Language Models

Abir Chebbi*
abir.chebbi@hesge.ch
university of applied sciences and arts western Switzerland & University of Geneva
Geneva, Switzerland

Guido Kniesel
guido.kniesel@hslu.ch
Lucerne University of Applied Sciences and Arts
Lucerne, Switzerland

Nabil Abdennadher
nabil.abdennadher@hesge.ch
university of applied sciences and arts western switzerland
Geneva, Switzerland

Giovanna Dimarzo
giovanna.dimarzo@unige.ch
University of Geneva
Geneva, Switzerland

ABSTRACT

Agriculture, as one of humanity's most essential industries, faces the challenge of adapting to an increasingly data-driven world. Strategic decisions in this sector hinge on access to precise and actionable data.

Governments, major agriculture companies, and farmers have expressed a need for worldwide monitoring of crop commodity quantities and prices. However, the complex and diverse nature of agricultural data and crop commodities, often presented in unstructured formats, pose significant challenges in extracting meaningful insights.

This study delves into the effectiveness of Large Language Models, particularly in Named Entity Recognition, focusing on their ability to efficiently tag and categorize crucial information related to agriculture, vessel tracking, imports, and exports, thereby enhancing data accessibility.

Our results indicate that while fine-tuning a base model achieves high precision, Large Language Models, particularly GPT-4 and Claude v2, demonstrate comparable performance with the added benefit of requiring no additional training for new entity recognition.

This research highlights the promising role of Large Language Models in agricultural AI, suggesting their use as a scalable solution for real-time data analysis and decision support in agriculture.

KEYWORDS

Large Language Models, Named Entity Recognition, Agriculture, Commodities monitoring.

ACM Reference Format:
Abir Chebbi, Guido Kniesel, Nabil Abdennadher, and Giovanna Dimarzo. 2024. Enhancing Named Entity Recognition for Agricultural Commodity Monitoring with Large Language Models. In *4th Workshop on Machine Learning and Systems (EuroMLSys '24), April 22, 2024, Athens, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3642970.3655846

1 INTRODUCTION

Agriculture is one of the oldest and most essential industries, providing food and raw materials for humanity. This vital sector is increasingly driven by data, with myriad of decisions made daily that impact production, sustainability, and resource management.

Monitoring the quantities and prices of crop commodities on a global scale is becoming increasingly important. Many governments, major agriculture companies, and farmers are demanding access to this information in order to make informed decisions.

In this information-intensive environment, agricultural producers, and small agricultural commodities buyers/sellers require accurate, timely, and actionable insights to make informed decisions. For instance, Fluctuations in soybean prices can be influenced by a variety of factors including weather conditions in major soybean-producing regions, global trade policies, and changes in demand. By having access to real-time data on commodities prices, which can be extracted from financial news, market analysis reports, and

EuroMLSys '24, April 22nd 2024, Athens, Greece

ACM ISBN 979-8-4007-0541-0/24/04…$15.00
https://doi.org/10.1145/3642970.3655846

commodities exchanges, the farmer can make an informed decision on when to sell the crop to maximize profit.

However, the challenge lies in the nature of the data, Crucial information exists in an unstructured format, scattered across different sources such as market reports, news articles, and digital communications. Extracting and analyzing meaningful insights from this unstructured data is a complex and time-consuming task.

Natural Language Processing (NLP) emerges as a key player in this scenario. Its ability to process and understand human language, both in text and speech, makes it an ideal tool for dissecting the vast, unstructured datasets prevalent in agriculture. Among NLP technologies, Named Entity Recognition (NER) is particularly relevant. NER can categorize key information like crop types, geographical locations, and market trends, transforming raw data into valuable knowledge.

In this context, We aim to develop an automated solution that provides efficient analysis of various information sources. For example, a specialist in a specific crop would be able to quickly and effectively access relevant files and news about that crop. The solution will utilize NER to categorize and prioritize information, making it more accessible and useful for stakeholders in the agricultural industry.

Despite its potential, the application of NLP in agriculture, especially NER, has been less developed compared to other sectors like healthcare and finance. A significant obstacle is the scarcity of specialized datasets in the agricultural domain, which is crucial for developing robust systems. This paper addresses this gap by evaluating the efficacy of Large Language Models(LLMs) for NER tasks in the agricultural domain. Our focus is on identifying and classifying key entities, such as product types and locations, within agricultural texts. The lack of specialized agricultural datasets for NER raises a critical question: Can LLMs be effectively used and replace traditional model training approaches for NER tasks, particularly within the agricultural domain?

This paper is structured as follows: Section II reviews the evolution of NER methods and the use of NER. Section III details our methodology, outlining the datasets and models used. Section IV presents our results, demonstrating the promising capabilities of LLMs in accurately performing NER tasks. Through this research, we contribute to the field of agricultural AI and demonstrate the practical utility of LLMs in fostering data-driven and sustainable agricultural practices.

2 RELATED WORK

2.1 The evolution of NER techniques

NER is a specialized task within Natural Language Processing that focuses on identifying and classifying entities from unstructured text into predefined categories. These entities can vary depending on the domain of the application. For instance, in the agriculture field, we could have Commodity, Temperature, Precipitation, Price, etc.

In the early stages of NER, systems were largely rule-based including LaSIE-II [9], FACILE [3], and LTG [13] systems. As an example, Kim [8] proposed ProMiner, which leverages a pre-processed synonym dictionary to identify protein mentions and potential genes in biomedical text. These methods were somewhat effective but lacked the sophistication needed to understand context and complex language nuances.

The introduction of statistical methods, particularly Conditional Random Fields(CRFs), marked a significant improvement in NER systems by considering the context and dependencies of words. This algorithm has found application across multiple domains, including biomedical text [17], tweets [10], and chemical text [16]

However, the real breakthrough came with the advent of deep learning techniques, initially pioneered by Collobert and Weston[4]. The most used methods for NER with deep learning are Recurrent Neural Networks(RNNs) such as BiLSTM [12], convolutional Neural networks [7], etc. These networks enhanced the ability to capture contextual information over longer stretches of text, significantly improving the accuracy of NER systems.

The current state-of-the-art in NER is dominated by transformer-based like BERT [6], Roberta [11], and their variants. These models, pre-trained on vast text corpora in a self-supervised manner, have shown remarkable success in understanding the context and subtleties of language, providing a robust foundation for NER tasks. They can be fine-tuned with domain-specific datasets to perform well on specialized NER tasks.

2.2 NER in agriculture

Despite these advancements, NER's application in agriculture remains limited, partly due to the lack of large, domain-specific datasets and the extensive human labor required for labeling. Agriculture possesses unique jargon and terminology, necessitating tailored NER solutions.

Sayan De et al. [5] address the challenge of extracting relevant information from agricultural data. They propose an 'AgriNER' dataset that includes thirty-six types of entities and nine types of relations pertinent to the agricultural domain. This resource is aimed at facilitating the creation of Knowledge graphs.

Quoc Hung Ngo et al. [14] developed a new tagset for NER in the agricultural domain. Their study outlines a two-stage process: initially employing semantic-based approaches for entity detection and corpus building, followed by deep learning techniques for entity recognition from plain text.

Despite these developments, the NER models and datasets previously established do not fully align with our specific needs. The primary challenge lies in the discrepancy between the entities we seek to identify and those covered by the existing tagsets.

2.3 LLMs and NER

Recent advancements in LLMs such as GPT-3.5 and GPT-4 [15] have demonstrated an impressive ability to adapt to various textual domains, hinting at a promising direction for overcoming challenges in agricultural NER and other specialized fields. These models' flexibility and understanding potentially pave the way for more nuanced and accurate entity recognition across diverse domains.

Wang et al., as referenced in [18] effectively bridged the gap between traditional labeling tasks and the generative nature of LLMs with their GPT-NER. This system leverages the inherent strengths of LLMs, adapting them for precise entity recognition. GPT-NER achieves comparable performance to the supervised baseline.

Furthermore, Zhou et al. [19] explore the potential of LLMs in NER through targeted distillation. This methodology involves distilling a large, general-purpose LLM into a more focused NER model, resulting in remarkable accuracy across numerous domains and entity types. However, it is noted that these advancements have yet to be specifically addressed in the agriculture domain.

In summary, NER has undergone significant evolution, from rule-based systems to advanced LLMs. While considerable progress has been made, the journey of NER, especially in specialized domains like agriculture, is ongoing with continuous innovations and adaptations needed to meet specific challenges and requirements

3 METHODOLOGY

In this study, our primary goal is to assess the efficacy of LLMs in the NER for the efficient tagging of agricultural documents based on their key information. These documents, varying in format from text to tables, originate from diverse sources. They cover information about trading, vessel tracking, exports, imports, and crop status in different countries. Therefore, we aim to extract entities such as Currency, Exchange Month Reference, Location, Period Aggregation Type (year, month, etc.), Unit (Tones, Hectar, Hour, etc.), Vessel Status (Loaded, moved, discharged, etc.) Trade Flow Mode (Import, Export, etc.), etc.

3.1 Data collection

The dataset encompasses various commodity types, including Janice, Soybean, Vegetable Oil, and Sugar, among others. These documents, sourced in formats such as PDF and Word files, were obtained from a diverse array of contributors including brokers and international organizations, ensuring a rich representation of the global landscape of crop commodities.

For the intricate task of document annotation, we employed the AWS Ground Truth tool [2], hosted on Amazon Web Services (AWS). This tool was chosen for its advanced capabilities in facilitating a visual understanding of the content within PDF documents. Figure 1 illustrates the user-friendly interface of the AWS Ground Truth annotation tool, showcasing its utility in our research process.

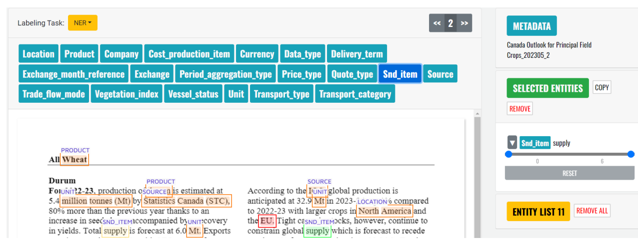

Figure 1: Interface of the AWS Ground Truth Annotation Tool

Our preliminary results are centered around 5 entities: 'Location', 'Product', Snd_item, Currency, and Unit, providing a foundational understanding of model performance in these key areas. We initiated our research by annotating a selection of documents pertinent to the agricultural market using AWS Ground Truth Sagemaker. This process allowed for the creation of a dataset comprising 395 samples, each representing a sentence from the agricultural domain. These samples include two crucial fields:

- Text: This field contains the original sentence that underwent the annotation process.
- Entities: Within this field, we maintain a structured list of annotated entities. Each entity is described by a sublist containing the following attributes:
  - start: The index of the first token of the entity within the sentence.
  - end: The index of the last token of the entity within the sentence.
  - label: The label or category assigned to the entity, indicating its specific type or nature within the agricultural context.

Table 1: Entity types used for annotation

Entity	Entity description	Examples
Product	Words that denote a specific product or item in the agriculture domain	Corn, Cotton Wheat, Cattle Mutton, Poultry, etc.
Location	Words that represent geographical locations, such as city names, countries, etc.	Decatur, IL , OHIO, Fort Dodge, etc.
Snd_item	reference for something in the inventory's lifecycle	inventory, livestock, exports, demand, etc.
Currency	Words that represent various types of currency	Brazil Cruzado Novo ,Iraqi Dinar, Egyptian Pound, US Dollar, etc.
Unit	Words that denote units of measurement in various contexts	USD per metric ton, hectare, tons, meter, Day, Week, etc.

3.2 NER models

In our pursuit to enhance NER within this specific sector, we selected a suite of models renowned for their linguistic capabilities and adaptability across various domains. The models chosen for evaluation include:

- GPT 3.5 turbo
- GPT 4 [15]
- Claude v2 [1]
- Multilingual BERT using spacy-transformers

To tailor our models for the specific NER task in agriculture, we pursued different approaches based on the model type:

- Large Language Models(LLMs): For the LLMs, such as GPT-3.5 Turbo, GPT-4, and Claude v2, we leveraged the few-shot prompting technique. This approach involves providing the models with a small set of examples to guide their understanding and generation capabilities towards agricultural entity recognition. By crafting prompts that include examples of the entities we aim to identify, we effectively steer the models to recognize similar patterns in new texts.
- Base Model: The Multilingual BERT model, facilitated by the spaCy-transformers library, bridges spaCy's efficient NLP tools with the powerful transformers from Hugging Face. Initially, Multilingual BERT's pretrained capabilities are geared towards general language understanding, which necessitates fine-tuning to align with our domain-specific objectives. We refined the model using our annotated dataset, which encapsulates the nuanced vocabulary and entity types related to agriculture, vessel tracking, imports, and exports. The fine-tuning process was meticulously planned, with the dataset segmented into training and evaluation subsets. Specifically, 60% of the samples (237) were allocated for training, enriching the model's learning phase with diverse examples. The remaining 40% (158 samples) constituted the evaluation set, serving as a benchmark to assess the model's proficiency in accurately identifying and categorizing entities within the agricultural context.

4 RESULTS+ DISCUSION

For NER evaluation metrics, we use the F1 score which represents the harmonic mean between precision and recall, providing a balance between them. It is calculated from precision and recall metrics (1).

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{1}$$

Where:

- Precision: is the number of correctly identified positive results (true positive) divided by the number of all identified positives, including those not identified correctly.
- Recall: is the number of correctly identified positive results divided by the number of actual positive results.

Figure 2 illustrates F1 scores attained by different models across five entity types. The models evaluated are GPT-3.5 Turbo, GPT-4, Claude v2, and Multilingual BERT. The entity types assessed are Product, Location, Snd_Item, Currency, and Unit. Key observations are as follows:

- The F1_score for 'Product' entity recognition is highest with Multilingual Bert, closely followed by GPT-4 , and lowest with Claude v2.
- 'Location' entity recognition scores are fairly consistent across all models, with GPT-4 and Multilingual Bert sharing the highest scores.
- The 'Snd_Item' entity has the most variation in scores, with the highest score achieved by Multilingual Bert and the lowest by GPT-3.5 Turbo.
- For 'Currency', Claude v2 outperforms the other models, with GPT-4 coming in a close second.
- The 'Unit' entity recognition is best performed by Multilingual BERT, followed by GPT-4.

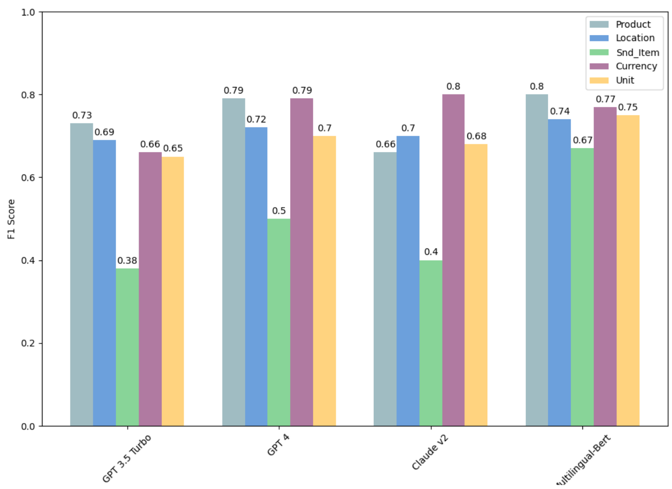

Figure 2: Comparing the F1 scores of various language models for NER tasks, specifically in identifying the five entities.

The 'Snd_Item' entity proves to be particularly challenging, as it is highly context-dependent. For instance, discerning 'Imports' as a 'Snd_Item' when 'Import' can also represent a 'Trade_Flow_Mode' entity requires nuanced understanding.

Among the LLMs, GPT-4 demonstrated superior performance compared to Claude v2 and GPT-3.5 turbo. This suggests that GPT-4 exhibits a higher level of proficiency in identifying the five entities.

Compared to all evaluated models, multilingual BERT showcased robust performance across all entity categories. This model achieved higher F1 scores, which can be attributed to their customized architectures tailored specifically for NER tasks.

Incorporating LLMs into the production environment for NER tasks within the agricultural domain presents a significant opportunity to overcome some of the limitations inherent in base models. The results of our study show that LLMs, particularly GPT-4 and Claude v2, perform commendably, with F1 scores approaching those of fine-tuned base models. This suggests that LLMs have the potential to be highly effective in accurately recognizing agricultural entities, even when they have not been explicitly trained on domain-specific datasets.

The advantage of using LLMs in production is their ability to handle the introduction of new entities. In this dynamic field, where new terms may be required, LLMs' adaptability can be leveraged to maintain accurate entity recognition without the need for continuous retraining. This contrasts with base models, which typically require an annotated corpus to learn new entities, which is a process that can be time-consuming and resource-intensive.

However, while LLMs show promising results, it's crucial to consider the trade-offs. The robust performance of fine-tuned base models on specific tasks must be balanced against the flexibility and lower maintenance requirements of LLMs. Furthermore, questions about the interpretability and predictability of LLMs in production scenarios persist, underscoring the need for ongoing research into their behavior and performance.

CONCLUSION AND FUTURE WORK

In this study, we embarked on an exploration of the capabilities of LLMs in performing NER tasks within the agricultural domain. Through meticulous annotation and evaluation using the F1 score metrics, we compared various models including GPT 3.5 turbo, GPT-4, Claude v2, and Multilingual BERT, specifically focusing on five entities: Product, Location, Snd_Item, Currency, and Unit. Our findings reveal that LLMs, particularly GPT-4 and Claude v2, hold their ground against base models, displaying a robust capability to identify entities. This is particularly noteworthy as it suggests the potential of LLMs to perform well even without extensive domain-specific training.

Nonetheless, the deployment of LLMs does not come without its considerations. The interpretability and predictability of LLMs within real-world production environments remain areas requiring further understanding.

Moving forward, our goal is to increase the application of LLMs in the agricultural domain while addressing the challenges and limitations observed in these models. Key areas of future research include:

- **Expanding Entity Types:** We aim to extend our research to include a wider array of entity types to cover our analytical needs. This will provide a more comprehensive understanding of the models' capabilities and the intricacies of agricultural data.
- **Assessing Model Trustworthiness:** Given the mixed trust in LLMs' performance, an important direction is developing methodologies to assess and ensure the reliability of these models in agricultural contexts. This includes exploring techniques for error analysis, interpretability, and robustness in LLM applications.
- **Continuous Improvement:** Implementing mechanisms for continuous learning and adaptation can help models stay relevant and improve over time.

REFERENCES

[1] [n. d.]. Model Card and Evaluations for Claude Models Anthropic.

[2] 2024. L'humain dans la boucle – Amazon SageMaker Ground Truth – AWS. https://aws.amazon.com/fr/sagemaker/groundtruth [Online; accessed 28. Mar. 2024].

[3] William J Black, Fabio Rinaldi, and David Mowatt. [n. d.]. FACILE: DESCRIPTION OF THE NE SYSTEM USED FOR MUC-7.

[4] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*. 160–167.

[5] Sayan De, Debarshi Kumar Sanyal, and Imon Mukherjee. [n. d.]. AgriNER: An NER Dataset of Agricultural Entities for the Semantic Web. https://brat.nlplab.org

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[7] Tao Gui, Ruotian Ma, Qi Zhang, Lujun Zhao, Yu-Gang Jiang, and Xuanjing Huang. 2019. CNN-Based Chinese NER with Lexicon Rethinking.. In *ijcai*, Vol. 2019.

[8] Daniel Hanisch, Katrin Fundel, Heinz Theodor Mevissen, Ralf Zimmer, and Juliane Fluck. 2005. ProMiner: Rule-based protein and gene entity recognition. *BMC Bioinformatics* 6 (5 2005). Issue SUPPL.1. https://doi.org/10.1186/1471-2105-6-S1-S14

[9] K Humphreys, R Gaizauskas, S Azzam, C Huyck, B Mitchell, H Cunningham, and Y Wilks. [n. d.]. UNIVERSITY OF SHEFFIELD: DESCRIPTION OF THE LaSIE-II SYSTEM AS USED FOR MUC-7. www.dcs.shef.ac.ukkresearchhgroupssnlpgate

[10] Xiaohua Liu, Shaodian Zhang, Furu Wei, and Ming Zhou. [n. d.]. Recognizing Named Entities in Tweets. , 359-367 pages. http://sourceforge.net/projects/opennlp/

[11] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[12] Ling Luo, Zhihao Yang, Pei Yang, Yin Zhang, Lei Wang, Hongfei Lin, and Jian Wang. 2018. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. *Bioinformatics* 34, 8 (2018), 1381–1388.

[13] Andrei Mikheev, Claire Grover, and Marc Moens. [n. d.]. DESCRIPTION OF THE LTG SYSTEM USED FOR MUC-7. www.ltg.ed.ac.uksoftware.

[14] Quoc Hung Ngo, Tahar Kechadi, and Nhien An Le-Khac. 2021. Domain specific entity recognition with semantic-based deep learning approach. *IEEE Access* 9 (2021), 152892–152902. https://doi.org/10.1109/ACCESS.2021.3128178

[15] OpenAI. [n. d.]. GPT-4 Technical Report.

[16] Tim Rocktäschel, Michael Weidlich, and Ulf Leser. 2012. Chemspot: A hybrid system for chemical named entity recognition. *Bioinformatics* 28 (6 2012), 1633–1640. Issue 12. https://doi.org/10.1093/bioinformatics/bts183

[17] Burr Settles. [n. d.]. Biomedical Named Entity Recognition Using Conditional Random Fields and Rich Feature Sets. http://us.expasy.org/sprot/

[18] Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, and Guoyin Wang. 2023. GPT-NER: Named Entity Recognition via Large Language Models. (4 2023). http://arxiv.org/abs/2304.10428

[19] Wenxuan Zhou, Sheng Zhang, Yu Gu, Muhao Chen, and Hoifung Poon. 2023. UniversalNER: Targeted Distillation from Large Language Models for Open Named Entity Recognition. (8 2023). http://arxiv.org/abs/2308.03279