

INTERPRETABLE AND GENERALIZABLE HTS CODE CLASSIFICATION FRAMEWORK

Zaruhi Navasardyan

Armenian State University of Economics,

Ph.D. student

zara.navasardyan@gmail.com

Abstract: The automation of Harmonized Tariff Schedule (HTS) code classification - a crucial task in international trade that entails assigning standardized codes to goods for tariff and regulatory purposes - is examined in this paper using GPT-3, a cutting-edge Large Language Model (LLM). The focus on creative prompt-engineering techniques and the usage a frozen model approach eliminates the need for further model fine-tuning. The suggested approach makes use of GPT-3's built-in capabilities to decipher intricate product descriptions and reliably assign them to the appropriate HTS categories. We show through an array of experiments that strategic prompt adjustments including few-shot prompting allowed the model's initial top-1 accuracy of 23 percent to be significantly increased to an average of 73 percent. The main strengths of the approach are the explainability for the chosen category as well as the fully generalizable nature of the methodology.

Keywords: LLM, NLP, HTS classification, GPT-3, generalizability, interpretability, international trade

JEL code: C55

Research objective: With an emphasis on interpretability as well as accuracy and automation, this research aims to develop a highly accurate and autonomous system for classifying Harmonized Tariff Schedule (HTS) codes.

Research Novelty: The development of a fully autonomous HTS code classification system making use of cutting-edge machine learning algorithms including large language models for accuracy, adaptability, and interpretability.

Introduction

Harmonized Tariff Schedule (HTS) codes are part of an internationally standardized system of names and numbers to classify traded products (HTS Revision 1, 2024). The system created and maintained by the World Customs Organization (WCO) assists in identifying goods by assigning a unique code that establishes the applicable tariffs duties and taxes in addition to offering statistical information for global trade. Chapters, headings, and subheadings that specify a product's type, material, and function are part of each unique HTS code which has ten digits maximum.

Around the world, classification is consistent because the first six digits are the same in all nations that have embraced the Harmonized System. Beyond the sixth number, there are country-specific digits that let different nations further categorize goods for statistical and tariff purposes.

The alterations going down to HS codes (as compared to HTS codes) are major, global changes that take place less frequently (e.g. once in 5 years). On the contrary, the closing 4 digits represent country-specific suggestions and are much more dynamic.

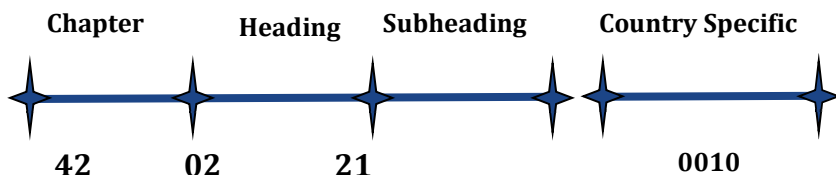


Figure 1. HTS code structure

For companies and governments to efficiently and accurately process and monitor goods as they cross international borders the HTS code system is essential. A smoother trade flow and increased efficiency of the world economy are made possible by properly classifying goods using HTS codes which also lowers the risk of customs delays and guarantees compliance with trade regulations.

HTS code classification automation is a common research area in the last decade given the resources and time spent on manual classification as well. Recent studies utilize extensively different machine learning algorithms for HS code classification using textual descriptions of the products under consideration.

One approach that can be used for the task is the pretraining of encoder language model-based (Devlin et al., 2019). The drawback of this approach is that it lacks any explainability and has limited generalization ability.

Another study uses the Naive Bayes algorithm for HS code classification aiming to optimize customs revenue (Muslim et al., 2022).

Some researchers in collaboration with Korea Customs Service use KoELECTRA-based sentence retrieval algorithms for HTS code

assignment to products (Lee et al., 2021). The same authors later built upon their previous research and added an explainability component to their framework which is partly automated yet requires some human intervention (Lee et al., 2024).

Some studies use not only textual descriptions provided in the declarations but also the image input which can contain information not anyhow present in the textual documents (Li et al., 2019).

This project aims to make a significant contribution to the field by providing a solid and creative solution that tackles the complex challenges of HTS code classification in the contemporary trading landscape utilizing state-of-the-art natural language processing (NLP) models. It is based on the insights gathered from our thorough literature review and the identification of existing gaps.

Research results

Three key obstacles stand in the way of the endeavor to create an interpretable and broadly applicable end-to-end HTS code classification system without the need for human intervention:

- ✓ **Accuracy:** Precisely allocating the appropriate HTS code to a wide range of products while identifying the small variations that set one product category apart from another is crucial to the system's operation. For companies involved in international trade achieving high accuracy is essential because any misclassification could cause logistical problems and financial ramifications.
- ✓ **Generalizability Across Countries:** The system needs to be able to adapt to the distinct HTS code extensions and modifications

that each nation has made. To ensure accurate and consistent classification across different jurisdictions this challenge necessitates a sophisticated model that can adapt to these country-specific nuances.

- ✓ **Adaptability to HTS Code Changes:** The systems recurring updates and modifications are one of its dynamic features. To keep the classification system current and maintain its ability to classify products accurately over time it must be built with the flexibility to adapt to these changes.
- ✓ **Explainability:** Among the challenges facing the development of an autonomous HTS code classification system, explainability emerges as a critical factor.

In the complex realm of international trade, where decisions on product classification have significant financial and regulatory implications, the ability of customs officers and trade professionals to understand the rationale behind algorithmic decisions cannot be understated. Explainability ensures that the system's suggestions are not only accurate but also transparent, fostering trust and facilitating easier review or appeal processes.

It is critical to address these issues to develop a reliable self-governing HTS code classification system. A framework like this would make international trade transactions easier lessen the workload associated with manual classification and guarantee adherence to the complex network of international trade laws.

Considering the intricate challenges identified in the literature and the pressing need for innovative solutions in the classification of Harmonized Tariff Schedule (HTS) codes, this research

endeavors to address these critical issues head-on. The goal is to develop an autonomous, precise, and explainable HTS code classification system that satisfies international trade's strict efficiency requirements and pushes the boundaries of algorithmic decision-making transparency and reliability. Our research aims to revolutionize international trade classifications by addressing the issues of explainability, generalizability across countries accuracy and adaptability to changes in HTS codes. This will lessen the reliance on manual processes and improve the capacity of traders and customs officials to navigate the complexities of global commerce confidently. The suggested methodological approach includes the usage of large language models with decoder transformer architecture for the HTS code classification as well as explanation generation with different prompting strategies like zero-shot or few-shot prompting.

Recent advancements in Large Language Models (LLMs) like GPT-3 (Brown et al., 2020), have significantly elevated their performance across a broad spectrum of downstream tasks. These models, operating under the principle of scaling laws, have demonstrated remarkable improvements in performance, sometimes extending across seven orders of magnitude. Seminal works by Kaplan et al. (2020) and Hoffmann et al. (2022) have solidified the understanding that LLMs adhere to scaling laws, positing that for any specific downstream task, an LLM is expected to outperform a conventional Pretrained Language Model (PLM).

In this context, our research aims to delve into the efficacy of LLMs in the specific task of matching product descriptions with Harmonized Tariff Schedule (HTS) descriptions, a crucial aspect of

global trade that demands high accuracy and understanding of complex trade language. By focusing on this comparison, we seek to illuminate the potential of LLMs to revolutionize tasks traditionally reliant on detailed human expertise, marking a significant step forward in automating and streamlining processes critical to international trade and commerce.

Moreover, the product descriptions provided to customs officers do not always contain comprehensive information, which can enough to have a detailed HTS classification. The experiment results show that large language models being trained on huge amount of data from web which should include product information from e-commerce websites has knowledge on the products and may act like a product description enhancing tool in behind. In the sake of simplicity, all the experiments are done and reported with GPT-3 API. Table 1 below presents some short search term enhancement results by a simple prompt "What is {Search terms}?". The first containing only a code would be difficult to classify even for customs officers without the background knowledge, which the LLM does have.

By embedding the HTS descriptions directly into the prompts, our methodology for product description to HTS code mapping not only enhances robustness against the fluctuations in HTS descriptions but also seamlessly integrates updates into the system. This intrinsic adaptability, combined with the proven state-of-the-art performance of LLMs in both academic and practical spheres motivated our exploration into this area.

Table 1. Product description enhancement results

Search term	Generated Description
e3181	E3181 is a code used to designate a particular type of electrical connector.
electronic scale us	An electronic scale is a weight measuring device that uses sensors and a digital display to show the weight of an object.
ladies 100 cashmere knitted pullover n f swtr csmr cr ls 2 nvy b	A ladies 100 cashmere knitted pullover is a sweater made from 100% cashmere wool. It is typically a crew neck sweater with long sleeves. The navy-blue color is a deep, rich blue color.

Our decision to use a frozen model, without additional fine-tuning, aims to ensure that updates to the HS code which are the output of the models do not impact the model's performance.

Overall, the process of getting the results using GPT3 was fully focused on prompt engineering effort using otheur knowledge of the tasks, understanding of GPT3 inner workings and using results of prior experiments. The example below shows one of the prompts submitted to GPT3 for inference together with a one-liner instruction which includes:

Provide the best possible matching option for the product description instead of <RESULT> and explain why you chose that category

The prompt itself includes the product descriptions and suggested categories.

There is a category of Harmonized Traffic Schedule which can be described as Tableware, kitchenware, other household articles and hygienic or toilet articles, of plastics: Tableware and kitchenware.

there are 4 options of subcategories below that can match a product description;

- (a) Salt, pepper, mustard and ketchup dispensers and similar dispensers
- (b) Plates, cups, saucers, soup bowls, cereal bowls, sugar bowls, creamers, gravy boats, serving dishes and platters
- (c) Trays
- (d) Other

description: plastic plate 100 box

correct subcategory: (b) Plates, cups, saucers, soup bowls, cereal bowls, sugar bowls, creamers, gravy boats, serving dishes and platters

description: plastic trays

correct subcategory: (c) Trays

description: squeeze bottle 24oz

correct subcategory: (a) Salt, pepper, mustard and ketchup dispensers and similar dispensers

description: bellows n straws

correct subcategory: <RESULT>

The output looks like this:

(d) Other

Explanation: The described items, "bellows" and "straws," do not neatly fit into the explicitly mentioned categories of salt and pepper dispensers, plates and kitchenware, or trays. Bellows, typically used for directing air, and straws, for drinking, do not align with the functionalities or categories of products described in options (a) through (c). Therefore,

they fall into the residual category of "Other," which accommodates items not explicitly covered by the more specific subcategories.

The structured approach detailed above, incorporating a two-step classification process for Harmonized Traffic Schedule (HTS) codes, demonstrates a significant enhancement in model accuracy. Initially, the model achieved a top-1 accuracy of 23% for the specific task of completing HTS codes with the last 4 digits, based on correctly predicted first-6-digit descriptions. Through comprehensive experimentation and prompt engineering, this accuracy was boosted to an average of 54%, a considerable improvement that underscores the effectiveness of the methodology when each step is correctly executed. Moreover, the introduction of additional in-context examples, or few-shot learning, further propelled the model's performance, achieving up to 73% accuracy on a selected subsample. This improvement highlights the potential benefits of few-shot learning in enhancing model accuracy.

Conclusion

Through timely engineering without further fine-tuning our study demonstrates the potential of GPT-3 a Large Language Model (LLM) in the automated classification of Harmonized Tariff Schedule (HTS) codes. Notable accuracy improvements are achieved. Our approaches improved performance with few-shot learning which in some cases further increased accuracy to 73 percent a 50-percentage point increase from a base accuracy of 23 percent. These findings highlight few-shot learnings importance and LLMs' efficiency in challenging classification tasks. Since a frozen model

was employed, it became completely generalizable to modifications in HTS documentation and could be tailored for use in different countries. Additionally, the method offers a thorough justification for the category selection. The advancement of AI integration in the efficiency of global commerce is expected to be facilitated by further developing and applying these techniques which should greatly improve international trade processes.

References:

1. **Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D.** (2020). Language Models are Few-Shot Learners. *ArXiv, abs/2005.14165*.
2. **Devlin, J., Chang, M., Lee, K., & Toutanova, K.** (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *North American Chapter of the Association for Computational Linguistics*.
3. **Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D.D., Hendricks, L.A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G.V., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., Rae, J.W., Vinyals, O., & Sifre, L.** (2022). Training Compute-Optimal Large Language Models. *ArXiv, abs/2203.15556*.
4. HTS Revision 1 2024. <https://hts.usitc.gov/>
5. **Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D.** (2020). Scaling Laws for Neural Language Models. *ArXiv, abs/2001.08361*.
6. KoElectra. <https://github.com/monologg/KoELECTRA>
7. **Lee, E., Kim, S., Kim, S., Jung, S., Kim, H., & Cha, M.** (2024). Explainable product classification for Customs. *ACM Transactions on*

Intelligent Systems and Technology, 15(2), pp. 1–24., <https://doi.org/10.1145/3635158>

8. **Lee, E., Kim, S., Kim, S., Park, S., Cha, M., Jung, S., Yang, S., Choi, Y., Ji, S, Song., M, and Kim, H.** (2021). Classification of goods using text descriptions with sentence retrieval. *In proc. of the Korea Artificial Intelligence Conference (KAIA)*.
9. **Li, G., & Li, N.** (2019). Customs classification for cross-border e-commerce based on text-image adaptive convolutional neural network. *Electronic Commerce Research*, 19, pp. 779 - 800.
10. **Muslim, H.A.** (2022). Naive Bayes algorithm in HS code classification for optimizing customs revenue and mitigation of potential restitutions. *Journal of Information Technology and Its Utilization*.

ԱՏԳ ԾԱԾԿԱԳՐԻ ԴԱՍԱԿԱՐԳՄԱՆ ԸՆԴՀԱՆՐԱԿԱՆ ՄԵԹՈԴԱԲԱՆՈՒԹՅԱՆ ՇՐՋԱՆԱԿԸ

Զարուհի Նավասարդյան

Հայաստանի պետական տնտեսագիտական համալսարան,
ասպիրանտ

Բանալի բառեր - LLM, NLP, HTS դասակարգում, GPT-3,
ընդհանրացում, մեկնաբանելիություն, միջազգային առևտուր

Ապրանքների՝ ԱՏԳ (HTS) դասակարգիչներին համապատասխանեցման ավտոմատացումը կարևոր խնդիր է միջազգային առևտրում և գտնվում է մաքսային ծառայությունների և հետազոտողների ուշադրության կենտրոնում: Այս աշխատությունում ուսումնասիրվում են ստանդարտացված ծածկագրերին ապրանքների դասակարգման ավտոմատացման

հնարավորությունները՝ օգտագործելով բնական լեզվի մշակման ժամանակակից ալգորիթմները: Առաջարկվող մոտեցումը օգտագործում է բնական լեզվի մշակման մոդելների հնարավորությունները՝ ապրանքների բարդ նկարագրությունները վերծանելու և դրանք համապատասխան ԱՏԳ ծածկագրերին ճշգրտորեն վերագրելու համար: Մի շարք փորձարկումների արդյունքներ ցույց են տալիս, որ ճիշտ ընտրված մոտեցումը կարող է բարձրացնել մոդելի ճշգրտությունը 50 տոկոսային կետով՝ հասցնելով այն մինչև 73 տոկոսի:

Այս արդյունքները ցույց են տալիս, թե ինչպես LLM-ները, հատկապես GPT-3-ը, կարող են ամբողջությամբ վերափոխել HTS ծածկագրի դասակարգումը՝ բարձրացնելով ճշգրտությունը՝ նվազեցնելով մարդու միջամտությունը և նույնիսկ պարզեցնելով միջազգային առևտրի ընթացակարգերը: Մոտեցման հիմնական ուժեղ կողմերն են ծածկագրի համապատասխանեցման մեկնաբանելիությունը, ինչպես նաև մեթոդաբանության լիովին ընդհանրական բնույթը:

Submitted: 04.04.2024; Revised: 22.04.2024; Accepted: 26.04.2024