

Pluto: A Deep Learning based Watchdog for Anti Money Laundering

Hao-Yuan Chen¹, Shang-Xuan Zou¹, Cheng-Lung Sung¹

¹Data Research & Development Center, CTBC Bank, Taipei, Taiwan
{haoyuan.chen, mi.zou, alan.sung}@ctbcbank.com

Abstract

Banks are faced with Anti-Money Laundering (AML) obligations so they have to identify customers by conducting negative news, aka "adverse media", screening which consists of searching information in the public domain for news items, publications, and government advisories and bulletins for information related an individual or entity's involvement in financial crime matters. Although it is an essential way to determine who in the group poses a higher risk for potential financial crime concerns by catching sophisticated activities from negative news across globe, it also requires heavy human capital and processing time on screening daily produced negative news. Therefore, poor efficiency becomes the most unacceptable obstacle based on this approach. To mitigate this issue, Pluto¹ offers a distributed and scalable batch system embedded deep learning-based Natural Language Processing (NLP) techniques for AML practitioners to improve daily task efficiency. It performs *text preprocessing*, *paragraph embeddings*, and *clustering algorithm* on a set of negative news and provide clustering result with keywords and similarities for AML practitioners. The overall feedback from AML practitioners are very positive on such an impressive enhancement in which Pluto reduces 67% efforts of negative news screening.

1 Introduction

Compliance with Anti-Money Laundering (AML) and Know Your Customer (KYC) duties require identity checks against potential money launderers, terrorists, or people considered high risk in the financial field. Although the concept is reasonable and feasible, there are tremendous challenges on insufficient approaches capable of offering an effective and efficient solution. It triggers lots of techniques been proposed in real world industry [Han *et al.*, 2018].

We leverages GlobalVision's Patriot Officer² to discover whether customers are high-risk individuals or not via check-

ing on a variety of watchlists (e.g., Dow Jones Watchlist) across the globe. One of the required steps is to screen plenty of related documents associated with several watchlists which are related to a single customer. It is an effective method to identify hidden criminal activities but not an efficient one. The volume of daily-crawled content are sometimes greater than the ones consumed by practitioners. It makes practitioners engage in investigation with massive human capital cost and processing time. Therefore, improving operational efficiency is the top priority for such an approach.

We propose an idea of on-demand document clustering through client-server model batch processing system in order to alleviate suffering in practitioners' daily task. It aims to group a set of documents in a way we proposed so that the documents in the same group are similar to each other than to those in other groups. It provides a web-based user interface (UI) not only revealing clustering result in tree-view structure accompanied by related keywords but showing similarity between the latest one and the others within each group as well. The purpose is to support practitioners conducting an effective decision-making in such an efficient way.

2 Design and Implementation

In this section, we introduce the details on Pluto architecture and method based on NLP and fundamental algorithms. It only supports the articles published in Traditional Chinese (zh-TW) at this stage.

2.1 Architecture

The client-server model batch processing system follows a micro-service oriented distributed architecture [Fehling *et al.*, 2014] presented in Figure 1.

The *API* service provides REST endpoints for the interaction between the *UI* and the *Patriot Officer*. It triggers processing pipelines by sending clustering job into *AMQP*³ when receiving requests from *Patriot Officer* and provides the clustering result information for *UI* from *DB*. The *Workers* service fetches jobs from *AMQP* to proceed pipeline for document clustering based on proposed solutions. The *UI* allows

¹A cartoon dog created in 1930 at Walt Disney Productions.

²<http://www.gv-systems.com/products-solutions/patriot-officer/>

³AMQP is an open standard application layer protocol used for queuing and routing messages between the services in a secure and reliable way

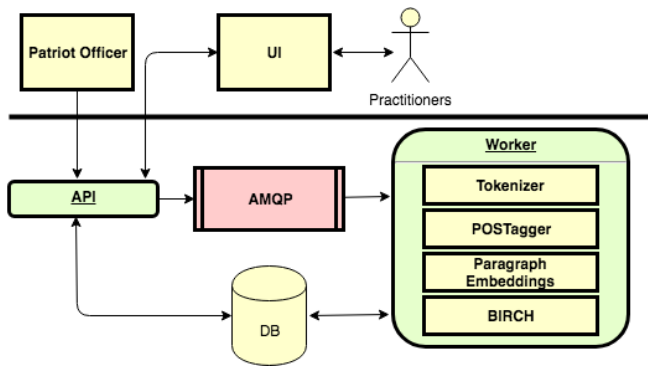


Figure 1: Pluto Architecture

practitioners to retrieve clustering information related to the specific watchlist.

The batch processing system is scalable. The quantity of *Workers* can be scale up and down according to quantity of jobs residing in AMQP, which makes the system achieve throughput guarantee.

2.2 Method

The pipeline comprises three sequential components: text preprocessing, paragraph embeddings construction, and clustering. Input is a set of documents associated to a single watchlist, and the output is a set of clusters containing all documents and the similarities within a single cluster.

Initially, we pick up tokens that match keywords offered by AML practitioners or be tagged as human names and proprietary nouns such as political groups, administrative unit, facilities, locations, organizations, and etc. Sinica CKIP tokenizer and POS tagger [Ma and Chen, 2003] are the major NLP techniques on this topic. The extracted tokens are regarded as representative for each document.

Next, we build the vector representation by *Distributed Bag of Words version of Paragraph Vector (PV-DBOW)* model [Le and Mikolov, 2014] which not only being conceptually simple but also requires to store less data. Moreover, it ignores the context words, which aligns the empirical insights from AML practitioner. The adopted deep learning-based techniques revealed from [Mikolov *et al.*, 2013a] and [Mikolov *et al.*, 2013b] is implemented by Gensim⁴ for this stage.

Eventually, we infer vectors for documents based on paragraph embeddings and apply *Balanced Iterative Reducing and Clustering Using Hierarchies (BIRCH)* [Zhang *et al.*, 1996] to proceed clustering. The reason why we adopt it is that BIRCH, one of the developments in hierarchical clustering, does not require us to pre-specify the number of clusters and is a memory-efficient and time-efficient approach. [Xu and Tian, 2015]

3 Evaluation

To show the effect of Pluto, we cooperate with professionals of AML to quantitate the performance via user testing. We

⁴<https://radimrehurek.com/gensim/index.html>

divide users into 2 groups: (1) reading news without clustered news and (2) reading news with Pluto.

For Pluto users, we implement a web interface to show the performance difference. As shown in Figure 2, the original system provides only an URL list, while Pluto visualizes negative news into clusters with keywords and similarities. With the interface, users can comprehend the news efficiently, and determine the suspect customer fast.

To evaluate the performance, both groups read the same news set contains 3000 news. As a result, Pluto reduces 67% time in average to judge if the customer suspected of mone laundering.

編號	來源
1	Taipei Times_12-Nov-2008_on Factiva.com (https://global.factiva.com/redirect/default.aspx?P=sa&AN=TAIP000020081112e4bc0000&cat=a&op=ASE)
2	The Wall Street Journal Online_05-Jan-2015_on Factiva.com (https://global.factiva.com/redirect/default.aspx?P=sa&AN=WSJ0000020150105eb15002e5&cat=a&op=ASE)
3	Ming Pao Instant News (Chinese Language - Simplified)_11-Sep-2009_on Factiva.com (https://global.factiva.com/redirect/default.aspx?P=sa&AN=MNPACN0020090911e59b018&cat=a&op=ASE)
4	http://irs.judicial.gov.tw/GNN/5/NWSS002.aspx?id=57279&flag=1&key=8&MchInfo=&countid=

(a) Original User Interface

(b) Pluto User Interface

Figure 2: Pluto offers grouped negative news annotated with keywords and similarities. In Figure 2b, the left side is tree-view structure for navigation purpose, and the right side is the pane for annotated negative news.

4 Conclusion and Future Work

In this work, we propose a distributed architecture batch system based on NLP techniques to organize daily negative news and offer an UI for information visualization. At present, the entire system is at piloting stage in our private cloud and facilitates efficient work flow among AML investigations pipeline.

In the future, we will (1) adopt our system to multilingual use cases, especially including Simplified Chinese (zh-CN) and English (en). (2) utilize NLP techniques in further investigation in which we may embrace named entity recognition (NER) and relation extraction (RE) to build the relation network identifying target suspicious entities, events, and time, and location.

References

- [Fehling *et al.*, 2014] Christoph Fehling, Frank Leymann, Ralph Retter, Walter Schupeck, and Peter Arbitter. *Cloud Computing Patterns: Fundamentals to Design, Build, and Manage Cloud Applications*. Springer, 2014.
- [Han *et al.*, 2018] Jingguang Han, Utsab Barman, Jeremiah Hayes, Jinhua Du, Edward Burgin, and Dadong Wan. Nextgen aml: Distributed deep learning based language technologies to augment anti money laundering investigation. In *Proceedings of ACL 2018, System Demonstrations*, pages 37–42, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [Le and Mikolov, 2014] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32, ICML'14*, pages II–1188–II–1196. JMLR.org, 2014.
- [Ma and Chen, 2003] Wei-Yun Ma and Keh-Jiann Chen. Introduction to ckip chinese word segmentation system for the first international chinese word segmentation bakeoff. In *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing - Volume 17, SIGHAN '03*, pages 168–171, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.
- [Mikolov *et al.*, 2013a] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.
- [Mikolov *et al.*, 2013b] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13*, pages 3111–3119, USA, 2013. Curran Associates Inc.
- [Xu and Tian, 2015] Dongkuan Xu and Yingjie Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, Jun 2015.
- [Zhang *et al.*, 1996] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. Birch: An efficient data clustering method for very large databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, SIGMOD '96*, pages 103–114, New York, NY, USA, 1996. ACM.