



Dr. Aravindhan Arunagiri

Manager, Mphasis NextLabs

Aravindhan, an expert in Logics and Algorithms, is part of Mphasis NextLabs. He has hands-on experience in various fields of BPM, Big Data, Cognitive Computing, AI as well as Robotics. He holds a PhD from Department of Management Studies, Indian Institute of Science, Bangalore.



Mr. Aavek Choudhury

Software Engineer, Mphasis NextLabs

Aavek holds a Bachelor's degree from Department of Computer Science and Engineering, SRM University, Chennai. His interests lie in the fields of Big Data and Artificial Intelligence. He is a technology enthusiast, and likes to explore the latest developments in the world of technology.

About Mphasis

Mphasis enables customers to reimagine their digital future by applying a unique formula of integrated cloud and cognitive technology. Mphasis X2C²™ formula for success, (shift anything to cloud and power everything with cognitive), drives five dimensions of business value with an integrated consumer-centric Front to Back Digital Transformation, enabling Business Operations and Technology Transformation. Mphasis applies advancements in cognitive and cloud to traditional application and infrastructure services to bring much needed efficiency and cost effectiveness. Mphasis' core reference architectures and tools, combined with domain expertise and hyper specialization are the foundation for building strong relationships with marquee customers.



An Intelligent System for Email Segregation and Routing: A Cognitive Approach

Dr. Aravindhan Arunagiri

Manager, Mphasis NextLabs

Mr. Aveek Choudhury

Software Engineer, Mphasis NextLabs

Abstract

In today's internet era, emails are used for exchanging information between people. Emails contain structured, semi-structured and unstructured (free text) data that gives information regarding sender/receiver of mail, subject and main topic (content) respectively. Businesses receive a large volume of email (both inbound and outbound), which can be very difficult to handle. Moreover, managing them manually is costly, critical and time consuming. For managing these emails, existing applications can only provide basic capabilities such as tagging/labeling mails, filtering spams, segmenting them based on sender/receiver address, etc. But high-level mail management functions like identification of semantics of mail (context, topics), summarization and actions (routing, archiving, purging) require human cognitive capabilities.

In this paper, an intelligent system is proposed for segregation and routing of mails based on its content topic and context. The system comprises of a cognitive framework that includes a cognitive text analytics (COT) engine and an inbuilt email processor engine that intelligently routes the mails to respective Subject Matter Experts (SME) to take required action. The cognitive framework can interact with any native mail applications and perpetually update its knowledge with ingestion of new mails. This reduces the cost of managing mails and improves the performance of mail applications.

Introduction

Email is a primary channel used to communicate and exchange information among the participating agents. The agents may be a group or individual persons, machines, department, organizations, etc. Emails are characterized based on structured, semi-structured and unstructured data. This data is quite useful for human agents to take appropriate actions that can include summarization, classification based on some features (eg. topics), segmentation, segregation and routing.

The number of emails sent and received per day, even in a midsize organization, is large and requires huge storage size^[3]. With recent development of technologies that can store and manage big data, these high-volume mails can be handled effectively. But managing the mail for taking cognitive decisions (like understanding and acting) on hourly, or even daily, basis requires a lot manual effort and high overhead cost. Modern email applications provide a wide range of capabilities that facilitate basic management of mails like tagging or labelling mails, segmentation of mails based on receiver address, spam and phishing mail filtering, automated mail response, etc. These capabilities are of great assistance to human agents to perform their high-level management functions such as comprehension of mail, summarization, segregation and routing. In addition to huge volume, the variety in its content (subject matter of discussion)

may require human agents to possess knowledge in various domains to be able to manage these emails. Hence, an intelligent system with cognitive capabilities to perform these segregation and routing functions becomes indispensable.

The current research work in text analytics is used to analyze the (free text) data aligned in email header, subject and body structures to create insights^[3] that can include identifying topics, theme, context discussed in mails, summarizations of mails, etc. These research artifacts that act on the standard texts incur performance degradation when applied to emails, due to the sparse topics and lack of contextual information^{[1][2]}.

To address this, an intelligent system that comprehends the email content and automatically takes decision is developed. The system contains cognitive text analytics engine that analyzes and segregates the emails based on the content information (e.g., topics) and maps them to the SMEs. An in-built email processor engine routes the segregated emails to the respective SME.

In the next section, the framework of intelligent email segregation and routing system, and its building components are discussed. The section "Process of Email Segregation and Routing" explains the functional and process aspects of the system, and the last section discusses various limitations of present system and future capabilities of the system.

Intelligent Email Segregation and Routing System Framework

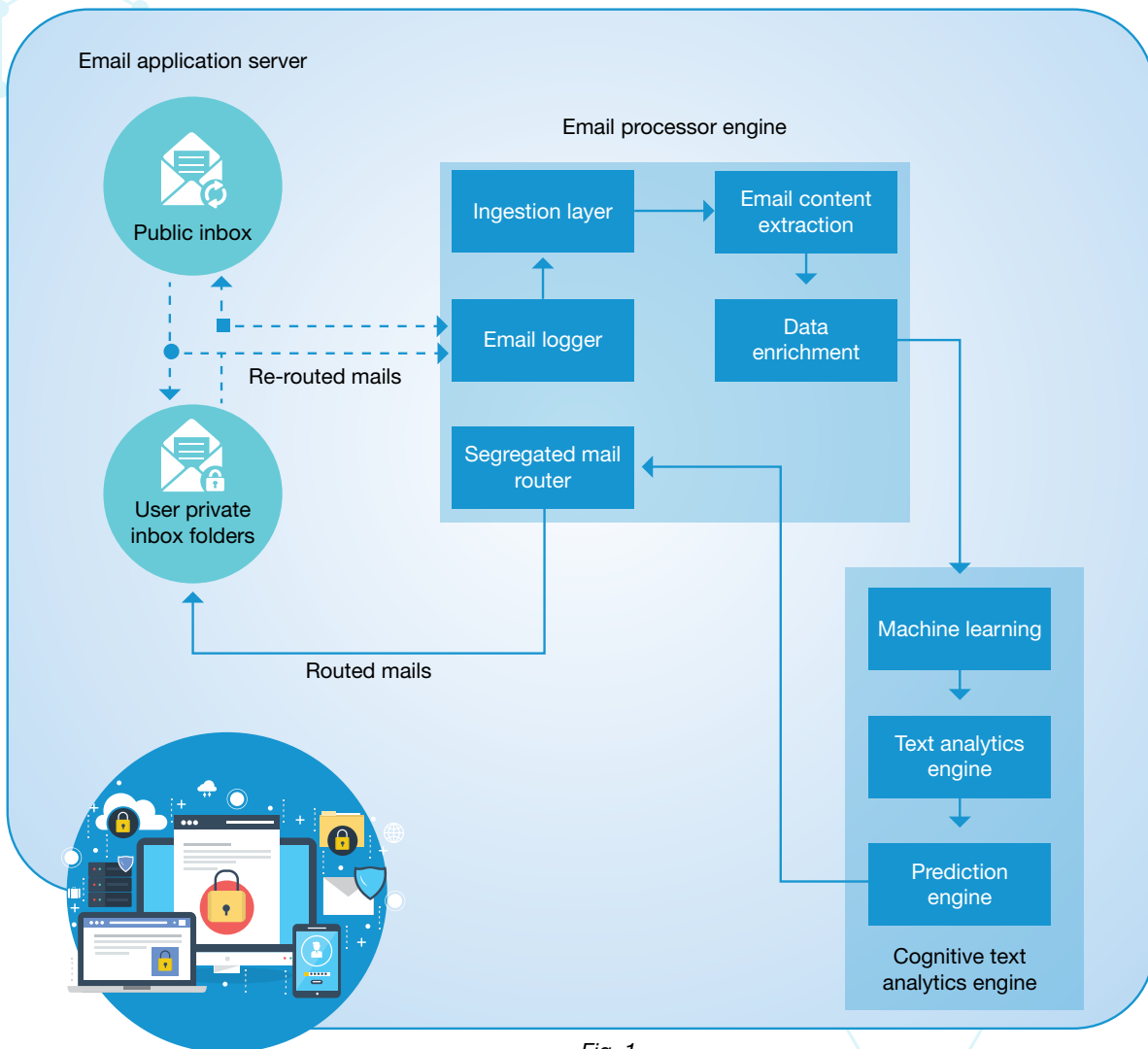


Fig. 1

Cognitive Framework of Intelligent Email Segregation and Routing System

The system intends to segregate the inbound email to a public inbox folder based on its topic of discussion and then routes the segregated messages to respective SME's private inbox for decision action. The framework shown in figure.1 consists of (i) Native email application server, (ii) Email processor engine, (iii) Cognitive text analytics engine.

Native email application server

It consists of a public inbox where the inbound mails from outside the environment are received. This public inbox receives emails that are meant to be sent to a specific SME, but are addressed to public folder (e.g. Help desk inbox). The email server also includes private mail folders of respective users, who can access them using their dedicated credentials.

Email processor engine

The email processor engine consists of (a) Ingestion layer, (b) Email content extraction module, (c) Data enrichment module, (d) Email logger, (e) Segregated mail router.

- 1. Ingestion layer:** It pulls the emails (re-routed between public and private folders) that are stored in the email logger. The pulled email comprises of a header, subject and the body (content). The layer assigns a unique ID to each email it extracts and stores.
- 2. Email content extractor:** This functionality extracts the email body content. The extracted content, still preserving the unique ID, is appended to the other information contained in the email and stored in a data format (e.g. CSV, Data) that is amenable for analysis in Cognitive text analytic engine.
- 3. Data enrichment module:** In this module, the irrelevant and/or noisy attributes are removed from the email data prepared in email content extractor module, to optimize the size of email data in the system.
- 4. Email logger:** It logs each mail along with its batch information (sender, receiver, time stamp) using respective batch ID. It logs any email that are re-routed between public and private folders in the application server.
- 5. Segregated mail router:** This module consists of email services protocol functions (like POP) that use the email ID class label assigned for each mail (by cognitive text analytics engine) to route them to respective Subject Matter Expert (SME) for an action.

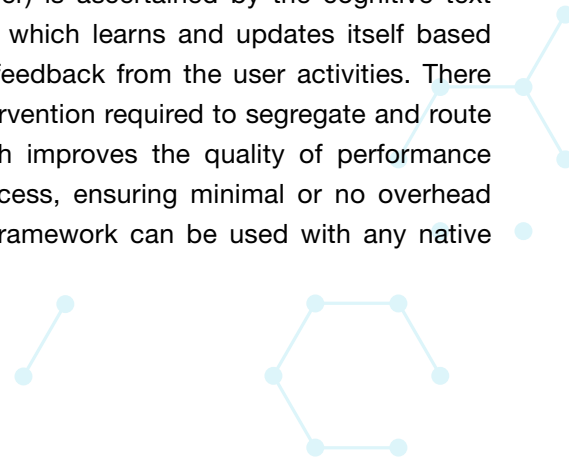
Cognitive text analytics engine

This engine includes text analytics module that performs feature selection and feature representation of the email content. The machine learning modules train the learning models using the new batch of sample emails (with class labels) received from the email processor engine. The cognitive engine then uses these trained models to predict another batch of emails and assigns a class label to each of these emails. The engine can update its learned models at regular intervals to maintain its recentness.

Process of Email Segregation and Routing

The current business practice requires the human agents for email segregation and routing. For example, a public inbox of help desk receives number of emails with variety of queries and request for actions that are intended to subject matter experts, but sent to help desk address. This help desk employs dedicated human agents to read, comprehend the mail content and re-route the mail to the respective SMEs. Another case is where the email is addressed to private folders of unintended SMEs. These mails are re-routed by the SMEs, back to the public inbox where the emails are re-routed to respective SMEs. These cases require manual intervention from the human agents to segregate and re-route emails. The volume of emails and varied types of their requests increase phenomenally, and the decisions over these mails are required to be made on daily or sometimes even hourly basis. The present manual practice is time-consuming, error-prone, and incur high overhead cost (that increases proportionally with volume of mails processed).

To address these challenges, an intelligent email segregation and routing framework is used to re-route the emails that reach a public inbox or private folders that are not intended for them but are addressed to them. In this framework, the relevance of the emails routed to specific user (private folder) is ascertained by the cognitive text analytics engine, which learns and updates itself based on the valuable feedback from the user activities. There is no manual intervention required to segregate and route the emails, which improves the quality of performance of the entire process, ensuring minimal or no overhead cost. Also, this framework can be used with any native mail application.



Limitations and Future Work

The present system assumes that the emails are received and processed in the fashion shown in the framework. Further the cognitive text analytics engine semantic capabilities can be used to leverage other email management functions like summarization, traffic management, business process discovery, etc. The future versions of the system can include the characterization of SMEs (w.r.t availability and performance) to optimize the routing and allocation of emails.

References:

1. AmeniBouaziz et al, "Short text classification using semantic random forest", Int. conference of data warehousing and knowledge discovery, DAWAK 2014, 2014, pp-288-299.
2. AthanasiaKolovou et al, "Tweesterar SemEval-2017 task 4: Fusion of Semantic-Affective and pairwise classification models for sentiment analysis of Twitter", Proc. Of International workshop on Semantic Evaluations (SemEval-2017), Association for Computational Linguistics, Vancouver, Canada, pp-675-682.
3. Ghulam Mujtaba et al, "Email classification research trends: Review and open issues", IEEE. transactions, Vol-5, 2017.

