A SURVEY ON SUMMARIZATION OF EMAILS WITH CLUE WORDS

Er. Kuldeep Kaur and Asst. Prof. Anantdeep Kaur
Department of Computer Engineering, Punjabi University Patiala

Abstract: Popularity of emails is increasing day by day in each field of the industry, email overload [1] becomes a major problem [1] for many email users. Users consume a lot of time for reading, replying and fixing or organizing their emails. A new paradigm is comes in summarization i.e email summarization that consider only summary of whole text email. To help users organize their email folders, many forms of support have been suggested, including email classification, email visualization and Spam filtering.

Keywords; industry, summarization, email folders, email visualization, classification, email

I. INTRODUCTION
EMAIL SUMMARIZATION

Summarization Technology is employ in the wide range sector in the industries now days. A summary [2] can be explain as a text that is generate from one or more texts or text documents, that contain a useful portion of the information in the original text, and that is no longer than half of the original text. Summarization [2] is the process of reducing a text document in order to create a summary that contains the most significant points of the text document. Text summarization [3] is useful for make the short text story of all the documents, newspaper articles, and emails correspondence or to extract significant facts for the search engine. To reduce the size, the sentences which are not near to the centroid are not to be considered in the output of the summary. Automated summarization is the way of decreasing a textual content file with pc software with the purpose to create a summary that keeps the most vital points or key elements of the authentic original document. Technologies that may make a coherent summary that include some variables like inclusive of period, writing style and syntax. Automated statistics summarization is part of machine mastering and records mining.

II. TYPES OF SUMMARIZATION

1. Abstractive summarization
2. Extractive summarization

➤ **Abstractive Summarization:**
An Abstractive summarization [4] concentrates to create an understanding of the main concepts in a document and then represent those concepts in clear natural language. Automatic abstractive summarization is a very complicate task and is still in its infancy. Abstractive summarization refers to create a document with new sentences that contain the core messages of the original one. This second form produces more fluent summaries as it respects natural language semantic rules.
In the abstractive summarization [5] the aim is to represent main concepts and ideas of a text document by interpretation of the source document in clear natural language. It uses linguistic methods to examine and translate the text and then to find the new facts and expressions to best identify it by generating a new smaller text that express the most important information from the initial text document. Problems that come with the abstractive summary are:
- The largest challenge in abstractive summary is the representation problem.
- Systems cannot summarize what their representations cannot capture. In limited domains, it may be possible to devise adapted structures, but a typical-purpose solution relay upon open-domain semantic analysis.
- Systems that can really "understand" natural language are beyond the effectiveness of today's technology.

➤ **Extractive Summarization:**

An extractive summary contains a subset of the text related to the original email. Since with this method the text is anticipated from their original context, reordering may be needed to make the summary understandable. In extractive summarization [5], the goal is to predict the most meaningful parts of documents like sentences, paragraphs, etc. to express main concepts of the document. Automatic extractive summarization is a much more developed area with many other approaches and tools. Extractive summaries present a number of advantages, such as:
- An extractive process is more lightweight than an intelligent procedure of summary composition. This translates into a reduced computation time.
- By using entire parts of the text included in the original email, it is impossible to compose new phrases with incorrect synonyms. Even if the flow between parts might result shaky, the internal meaning of every single part remains the same.
- When users read some text in the summary, they can easily link it back to the original email if needed. On the contrary, tracing back a topic from an abstractive summary to the original email requires more time.

III. HOW EMAIL WORK

Email [6] formulates an important means of communication in our daily exchanges, as a repository of corporate information used not only for personal conversation. Email does not work so differently than it used to when it first look

as. It depends on two basic communications protocols: SMTP (Simple Mail Transfer Protocol), which is used to send messages and POP3 (Post Office Protocol), which is used to receive messages.
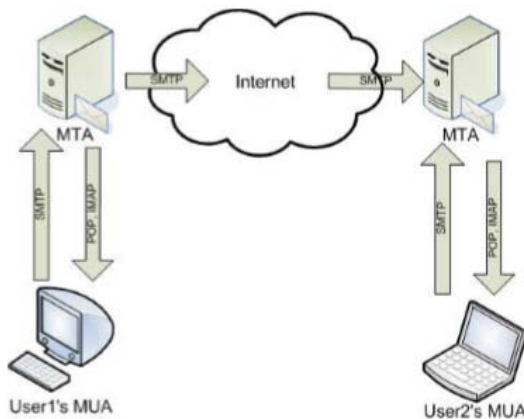
Some logical elements of the Internet Mail System are[7]:

1. Mail User Agent (MUA) – It helps the user to read and write email messages. The MUA is usually referred to as "email client" usually implemented in software. Two popular email clients are Mozilla Thunderbird and Microsoft Outlook. These programs convert a text message into the appropriate internet format in order in which these message to reach its destination.

2. Mail Transfer Agent (MTA) –It takes a message passed to either an MUA or another MTA and then decides for the appropriate delivery method and the route that the mail should follow. It uses the SMTP protocol to send the message to another MTA or an MDA.

3. Mail Delivery Agent (MDA): It receives messages from MTAs and delivers them to the user's mailbox in the user's mail server

4. Mail Retrieval Agents (MRA): It delivers mail messages from the user's mail server to the user's local inbox. MRAs are often embedded in email clients.

Email Mining

Email Mining [7] can be determined as an application of the latest upcoming research area of Text Mining on email data. Text Mining is an approaching field that has affectionate the interest of researchers from areas like Data Mining, Machine Learning, Natural Language Processing.

However, there are some specific characteristics of email data that set a distinctive separating line between Email and Text Mining:

1. Email includes additional information in the headers of email that can be employed for various email mining tasks.

2. Text in email is significantly shorter and, therefore, some Text Mining techniques might be inefficient in email data.

3. In Email Spelling and grammar mistakes also appear frequently because linguistic well-format is not guaranteed.

4. In an email message, different topics are appears; a fact that makes e.g. mail classification more difficult.

5. Email is generic techniques are difficult to be effective to individuals.

6. Email is a data stream and facts or distributions of target classes may change over. Algorithms should work in two ways: instance-wise and feature-wise, as new features (e.g. words) may appear.

7. Email contains probably have noisy data. Additional text like HTML tags and attachments must be removed in order to apply a text mining technique.

IV. CONCLUSION AND FUTURE SCOPE

In this work we studied the overview of the Email Summarization. Email Summarization is based on text Summarization that create summary of the original text document like half of the initial paragraph.With the ever increasing popularity of emails, it is very common nowadays that people discuss specific issues, events or tasks among a group of people by emails [8]. Those discussions can be viewed as conversations via emails and are valuable for the user as a personal information repository. In the base paper they adopted three cohesion metrics, clue words, semantic similarity and cosine similarity, to measure the weight of the edges.

V.REFERENCES

[1] Alam, M., & Kakkar, M. (2015, Oct 10). Email Summarization-Extracting Main Content from the Mail. International Journal of Innovative Research in Computer and Communication Engineering, 3, 7.

[2] Carenini, G., Ng, R. T., & Zhou, X. (2007, May 8). Summarizing Email Conversations with Clue Words. ACM, 10.

[3] Gupta, H., Kottwani, A., & Gogia, S. (2016). Text Analysis and Information Retrieval of Text Data. IEEE WiSPNET 2016 conference, (p. 6). Mumbai.

[4] Gupta, V., & Lahal, G. S. (2010, aug 3). A Survey of Text Summarization Extractive Techniques. JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, 2, 12.

[5] Lenjewer, B. (2015, May 5). Automatic text summarization with context based keyword extraction. International Journal of Advance Research in Computer Science and Management Studies, 3, 9.

[6] Loza, V., Lahiri, S., & Mihalcea, R. (2011). Building a Dataset for Summarization and Keyword Extraction from Emails. university of North Texas, Computer Science and Engineering. America: Samsung Research America.

[7] Katakis, I., Troumaka, G., & Vlahavas, I. (2006). Email Mining: Emerging Techniques for Email Management. ACM, 32.

[8] Fisher, D. and Moody, P., 2002. Studies of automated collection of email records. Institute for Software Research, University of California, Irvine.