Predictive process approach for email response recommendations

Ralph BOU NADER

Marwa ELLEUCH

Meriana KOBEISSI

Walid GAALOUL

Boualem BENATALLAH

Nour ASSY

Overview

Process mining involves **analyzing business process traces to identify inefficiencies and optimize their elementary steps**.

These **traces** can be **found in the logs of information systems** used by business actors during process execution, including emailing systems.

Process prediction uses data mining and machine learning techniques to identify future activities in a business process.

Email traces can be used for process prediction, but their **non-structured textual nature poses challenges**.

Existing works on email prediction: **focus on enhancing email management, with limited consideration for the context of business processes**.

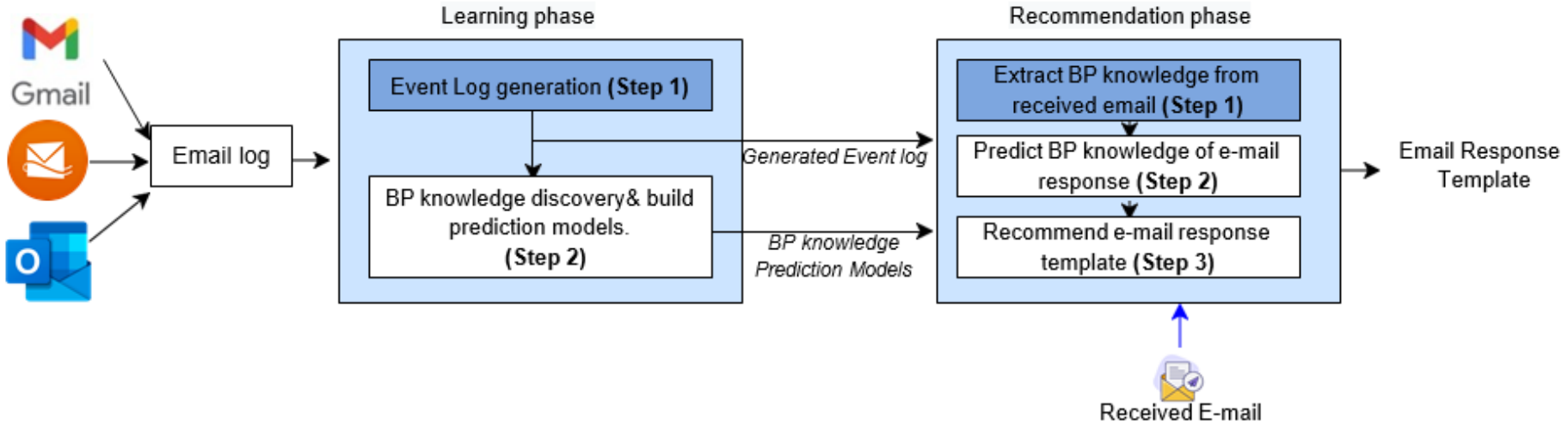
For those that combined email management with the notion of BP: **mostly limited to the stage of BP discovery from email logs or at most classifying incoming emails into BP activities.**

Overview

Making predictions in the context of process-oriented emails is **not limited to identify future BP activities to be performed through emails**. It **requires recommending the emails enabling BP actors performing these activities**.

Solution:

- A process-activity aware email response recommendation system.
- It generates an email response template of its main body based on the process and activity context of a received email. Our approach involves two main phases.



Learning phase (Phase 1)

The goal of this phase:

Learn BP knowledge-oriented from previously exchanged emails to generate BP-oriented predictions for an email response.

Step 1 - Event Log generation

- To generate a structured event log from email logs .
- Each event is primarily associated with one activity's occurrence (i.e., appearance) in an email. The following attributes characterize it:

  (1) the occurred activity-related attributes, which include the activity name, related business data values, and business context information

  (2) the speech act of the occurred activity

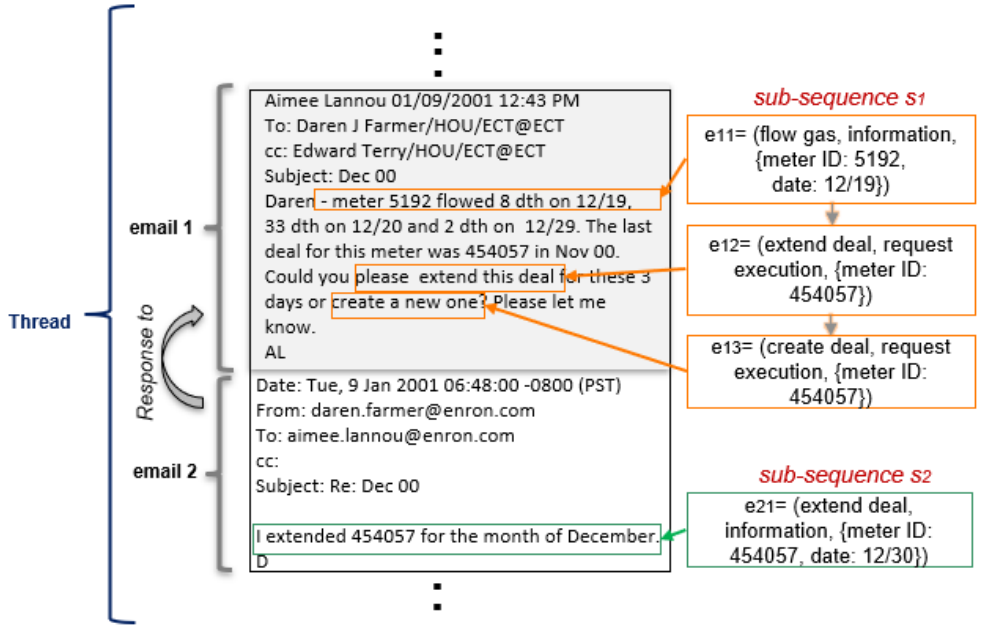
  (3) the email-related attributes where the activity appear (e.g., email ID, body, timestamp, sender & recipients)

  (4) the thread ID to which the email belongs.

we have approximated the concept of trace in process mining with the notion of thread, which refers to a sequence of related activities or events. A thread comprises a set of conversations that share relevant information values, such as business data values and email addresses, that are considered as a unit of analysis in our research.

Step 2A - BP knowledge discovery

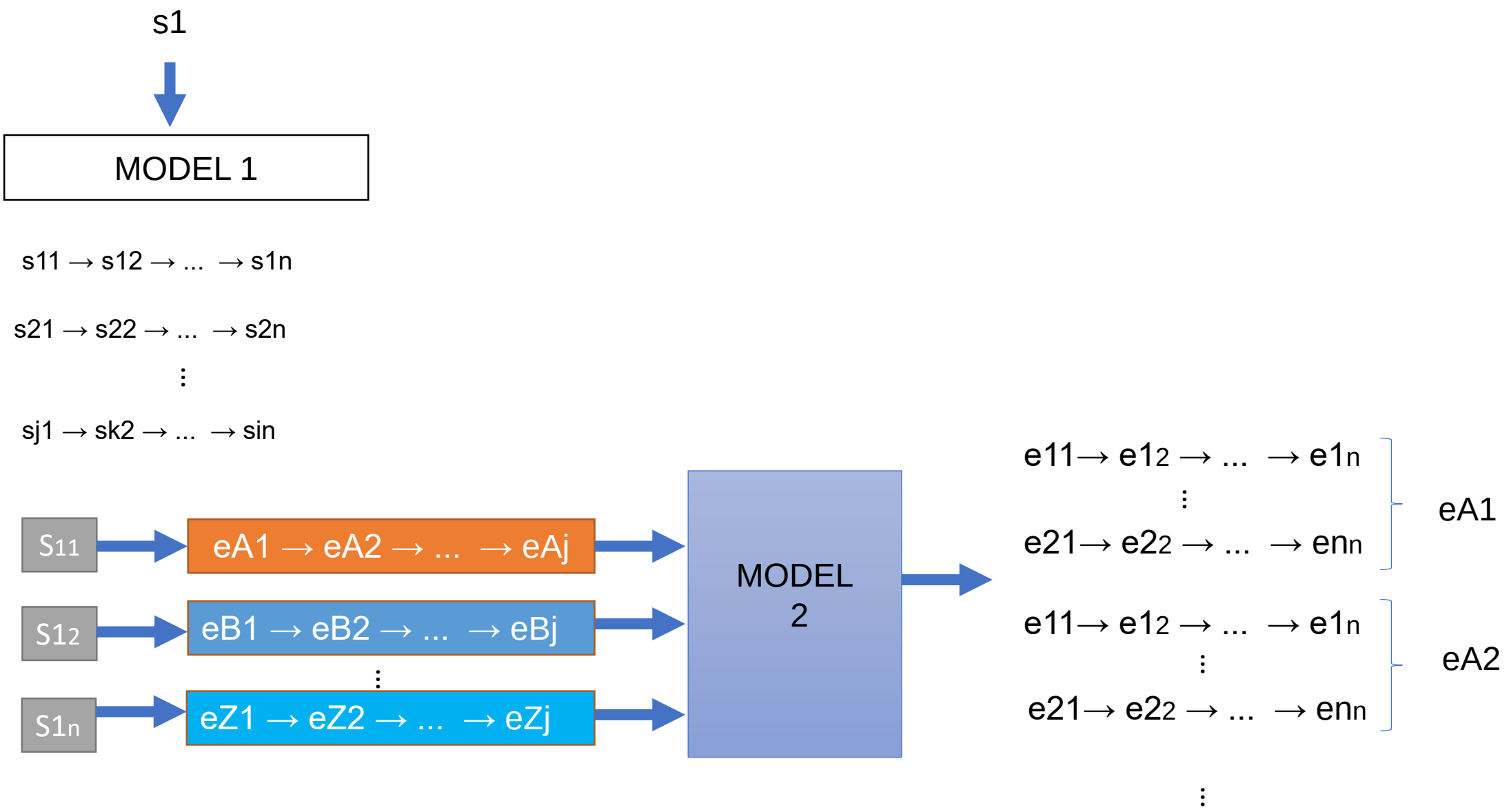
- Learn from the generated event log reduced into a training data-set of event sequences, the sequential relation between events.
- The goal is to predict the set of ordered events that follows the one in relation to an incoming e-mail considering its antecedents in the same email thread.
- we convert each thread in our event log into a sequence of events.



Step 2B - build prediction models

Building two prediction models:

- A first model takes as input the sub-sequence of events appearing in a received email and predicts the next combination of events that would appear in the email response.
- The second prediction model takes as input each event (belonging to the predicted events) and generates the sub-sequence of events following it in the same received email.


s1
MODEL 1
s11 → s12 → ... → s1n
s21 → s22 → ... → s2n
sj1 → sk2 → ... → sin
S11
S12
S1n
eA1 → eA2 → ... → eAj
eB1 → eB2 → ... → eBj
eZ1 → eZ2 → ... → eZj
MODEL 2
e11→ e12 → ... → e1n
e21→ e22 → ... → enn
eA1
e11→ e12 → ... → e1n
e21→ e22 → ... → enn
eA2

Recommendation phase (Phase 2)

Step 1 – Extract BP knowledge from received email

The first step is to discover the sub-sequence of events in the received email

We apply, the procedure supplied by Elleuch et al which proceeds as follows for discovering:

- **Activity and BD occurrences:** It considers an activity, or a BD as discovered in an email if one of its patterns reflecting its expression occurs.
- **Speech acts:** It applies a rules-based method. The rules are derived from natural language and are summarized into a decision tree that maps activity verb tenses to activity speech acts (e.g., the future tense refers to the 'intention' speech act and the past refers to the 'information' speech act)

Step 2 – Predict BP knowledge of e-mail response

- predict the BP knowledge to be included in the email response using the prediction models built in the first phase.
- feeds the sub-sequence of events extracted in the first step into the first prediction model. This later will return a list of combinations of events that would appear in the email response.
- each event in the retained combination is fed into the second prediction model and predicts the set of ordered events following it in the same received email.

Step 3 – Recommend e-mail response template

- generate an email response template, by discovering the textual content related to the BP knowledge of the email response.
- We assume that the suggested email responses: (i) must share a similar business context with the received email, (ii) contain the predicted BP knowledge, and (iii) have the same author writing style.
- for each event in the selected predicted list of events, we proceed as follows:

Step 3 – Recommend e-mail response template

- Retrieve all e-mail sentences from the previously sent emails that contain the same predicted BP knowledge.
- Predictive process approach for email response recommendations. Identify the business context of the received e-mail and the retrieved email sentences from the first step, using Yake, an unsupervised automatic keyword extraction technique.
- Filter the retrieved sentences by keeping those that have a similar business context to the received email based on cosine similarity matrix. A similarity greater than a chosen threshold (e.g. 0.5) indicates that the retrieved email sentence has the same business context as the received e-mail

Step 3 – Recommend e-mail response template

- Keep sentences with the same writing style as the recipient of the email
- Use stylometric analysis to classify different writing styles in the email
- Focus on three major features: lexical, vocabulary richness, and readability scores
- Perform classification using K-Means Clustering

Step 3 – Recommend e-mail response template

- Concatenate selected sentences in the order of appearance for each event in the predicted sequence of events.
- Use the Python natural language analysis library Stanza to locate and classify named entities mentioned in the sentences, such as date, person name, and location.
- Replace the extracted entities in the email template with their tags, which represent modifiable words that the employee should replace with real business data values.
- The generated confirmation email template contains modifiable words enclosed in curly brackets, such as the candidate's name, and shares similar business contexts with the application email, including job title and organization name.

Dear {name},

This letter is to inform you that we received your application for {job_title}, in our research group {organization}. We are currently in the process of taking applications, if you are selected for an interview, our human resources department will be in contact with you by {datetime}. Thank you again for the time you invested in your application.

Kind regards,
{signature}