Searching Attentive Tasks with Document Analysis Evidences and Dempster-Shafer Theory

Kristin Stamm, Andreas Dengel
German Research Center for Artificial Intelligence, Kaiserslautern
{firstname.lastname}@dfki.de

Abstract

Many enterprises strive toward the integration of input communication channels into their internal business processes. To help them, we propose to drive input channel document analysis (DA) by formalizing information expectations from current process instances in Attentive Task (AT) templates. This requires, however, to map incoming request documents to the related AT from a set of ATs. For this purpose, we present a search approach that prioritizes a set of ATs based on DA evidences. Our algorithm relies on the theory of Dempster-Shafer to iteratively handle DA results and further uses the string edit distance of Levenshtein to provide robustness to errors in DA results. We evaluate the search performance in terms of influence of evidences, error robustness, and ease of calibration for our approach.

1. Introduction

The establishment of new communication channels, such as email, represents a real challenge for enterprises in terms of information overload since, according to Bellotti et al. [1], employees must deal with increasing incoming request quantity but also complexity. Enterprises, therefore, aim at better supporting input channel management by automating document understanding and integrating incoming requests from multiple channels into the underlying business processes, e.g., mail, email, fax, and eDocs. However, most existing channel management systems do not support multiple channels and require manual operations leading to higher costs and lower processing quality.

We already proposed an approach of process-driven document analysis (DA) that formalizes information expectations from existing process instances in form of *Attentive Tasks (ATs)* [7] for guiding the overall DA with methods described in a *Specialist Board* [3]. An overview of this approach is depicted in Fig. 1. We showed that our approach is relevant for enterprises and can lead to performance improvements. However, achieving good results for a given document requires finding the corresponding AT based on the DA results only. The relevancy of the information resulting from the extraction methods is, therefore, crucial for searching the AT associated to a process instance.

In this paper, we propose and evaluate a search approach that prioritizes the current ATs relying on evidences generated during DA. The algorithm computes a degree of belief (DoB) for each AT based on these evidences and aggregates the results with a combination rule based on the Dempster-Shafer (DS) theory [6]. Robustness to errors in the DA is achieved by using the edit distance of Levenshtein. As first evaluations, we apply the algorithm to an email corpus representing the processes of a financial institution and analyze search performance. We first give an overview of the related work, then explain our approach, and finally present the results of some first evaluations.

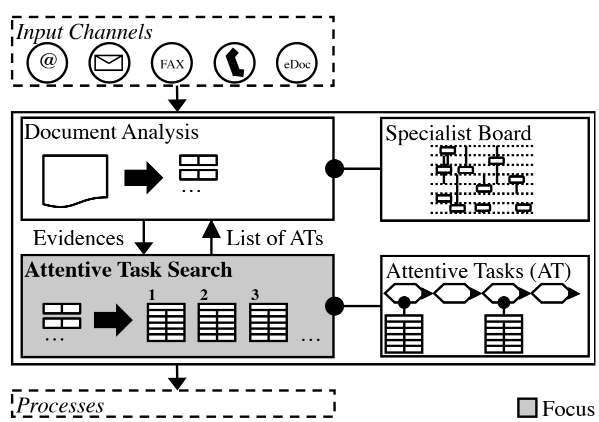

Figure 1. Overview of the process-driven document analysis and of the Attentive Task search.

Table 1. Example of an Attentive Task.

Descriptor	Value	Type	Constraints	Slot
SenderName	Ina Mueller	Person	isCustomer	Ident.
SenderEmail	ina@mueller.org	EmailAddress	Related(senderName)	Ident.
LoanType	Dispo	LoanType	{dispo, longTerm}	Other
LoanAmount	?	Money	LargerThan(0)	New
StartDate	?	Date	After(today)	New
EndDate	?	Date	After(startDate)	New

?: New value expected; Ident.: Identifying

2. Related Work

The problem of mapping input documents to task instances has already been addressed by many different researchers, especially in the field of email management. Unfortunately, existing approaches are not sufficient to solve this problem since they either consider uninstantiated processes only, or are limited to a specific domain. Some approaches use simple heuristics, e.g., Bellotti et al. [1] with their *thrasks* that are conversation threads with manual adaptation. Other approaches apply traditional classification approaches, e.g., Naïve Bayes or Support Vector Machines. Cohen et al. classify emails into sender intentions based verb-noun pairs called *speech acts* [2]. Dredze et al. combine a set of specific classification methods that rely on involved people or topics [4]. Unfortunately, all these methods are too domain specific to be applied to ATs. Kushmerick and Lau have developed an approach based on unsupervised learning to identify task structures from emails that is very efficient for personal email management with highly unstructured and implicit processes [5]. However, these approaches have a limited applicability for well-defined processes appearing in organizations, especially for customer interaction.

3. Search Approach

In the following, we give a brief introduction to the concept of Attentive Tasks (ATs) that has been used in our existing process-driven document analysis (DA) approach for efficiently scheduling DA [7]. This approach iteratively switches between prioritization of active ATs and execution of DA methods. We first focus on the design of ATs and then present the key elements of the new evidence based search algorithm.

3.1. Attentive Tasks

Attentive Tasks (ATs) are the formalization of information expectations toward incoming documents. An example of an AT is given in Table 1. An AT includes a set of information slots with a *descriptor*, a *value*, an information *type*, some value *constraints*, and the *slot* role. Some slots already have known values, like the email address of the customer, while other slots have unknown values remaining to be identified under some given constraints such as, for example, the type of loan being part of the current enterprise's portfolio.

ATs are generated for process instances that cannot proceed without external input, e.g., when a customer needs to send additional information. Active *ATs* are collected on a central server to be prioritized by the search module according to existing DA evidences.

3.2. Evidence-Based Search Approach

The search Algorithm 1 performs on a large set of ATs stored on a central task server and matches incoming requests to the corresponding AT using all text annotations so far extracted by DA specialists as evidences, i.e., a list of descriptor d and associated value v.

Algorithm 1 Search Attentive Tasks $atList$ given evidences $evidList$ and string distance $maxD$

for all e in $evidList$ **do**
  for all a in $atList$ **do**
    $d \leftarrow a.containsDescr(e.descr)$
    $v \leftarrow a.valueMatch(e.descr, e.value, maxD)$
    $vs \leftarrow a.valueSetMatch(e.descr, e.value)$
    $p_e(a) \leftarrow mass(d, v, vs)$
  end for
  $m_e \leftarrow normalize(p_e)$
  $m_{all} \leftarrow combine(m_{all}, m_e)$
end for
return $heapsort(atList, m_{all})$

The search algorithm computes a matching function p_e for each evidence with a matching value $p_e(a)$ for each AT a depending on a matching descriptor, a matching value or a set of constraints if no value is available in the AT. With normalization to 1, mass functions m_e are generated and combined with the Dempster-Shafer (DS) rule into a single mass function representing the degree of belief (DoB) for an AT based on the available evidences. Finally, the AT list is (heap-)sorted according to the DoB and returned to the DA module. The main ingredients of the algorithms are the following:

Matching parameters. When comparing an evidence to an AT, we distinguish matching parameters for the following 5 cases - evidence match (1.0), value mismatch (p_{mm}), evidence not found (p_{nf}), value expected (p_{ve}), and only value match (p_{ov}):

$$p_e(a) = \begin{cases} 1.0 & \text{if } \exists s \in a | e.d = s.d \wedge e.v = s.d \\ p_{mm} & \text{if } \forall s \in a | e.d \neq s.d \wedge e.v \neq s.d \\ p_{ve} & \text{if } \exists s \in a | e.d = s.d \wedge e.v = \emptyset \\ & \wedge\ e.v \in s.C \\ p_{ov} & \text{if } \exists s \in a | e.d \neq s.d \wedge e.v = s.v \\ p_{nf} & \text{else} \end{cases}$$

Where evidence e with descriptor d and value v is compared to each AT a that is a set of slots s with descriptor d, value v, and constraints C. *ValueSet-Match* is used when a value is expected (p_{ve}) from the incoming request and expressed by constraints (= set, regular expression). A match is given when the evidence value meets the constraints.

Combination rule. The matching values are normalized to one mass function m_e so that $\sum_a m_e(a) = 1$. The mass functions for all evidences are combined with Dempster-Shafer (DS) rule [6]. Two mass functions m_1 and m_2 are combined to m_{12} for AT a as follows:

$$m_{12}(a) = \frac{\sum_{B \cap C = a} m_1(B) m_2(C)}{1 - K} \text{when } a \neq \emptyset$$

$$m_{12}(\emptyset) = 0, K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$

Edit distance. For matching evidence values with AT slots, we allow a maximum Levenshtein distance (LD) to straighten out typos in the DA results. LD counts the minimum number of add, $delete$, and $substitute$ operations to equalize two strings. The threshold $maxD$ has to be optimized toward reducing errors and not deteriorating search results.

4. Evaluation

The evaluations aim at understanding the main influencing factors on search performance. We evaluated the influence of evidence types, the optimization of error robustness, and the calibration of search parameters.

4.1. Experimental Setup

We conducted our first evaluations on an email corpus based on two customer request processes of our case study partner, an international financial institution [7]. The corpus includes 49 emails from 10 customer probands requesting toward 2 processes. We generated Attentive Tasks (ATs) for each email and used

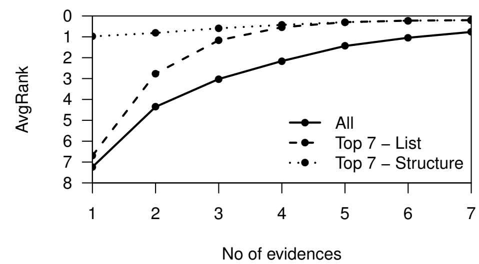

Figure 2. Influence of evidence sets.

our existing DA prototype to generate DA evidences driven by the AT. Even though our search algorithm is applicable to all input channels, we focused on email since this channel covers all document structures, e.g., image (as attachments), text and metadata. We examined the following aspects:

Influence of evidences. We evaluated if the selection of evidence types influences search performance by comparing three scenarios: (1) random use of available evidences (All), (2) simple selection of evidence types (Top 7 - List), (3) structured use of evidence types (Top 7 - Structure).

Robustness. We introduced robustness by using Levenshtein distance (LD). Since we believe that allowing deviations may lower search results, we examined LDs from 0 to 9 to optimize robustness.

Calibration. To fine tune the algorithm, we examined how the search is influenced by the different mass calculation values and its sensitivity to change.

4.2. Influence of Evidences

For the evaluation of evidence influence, we first conducted the search for different types and number of evidences by randomly selecting the available evidences from the DA results and generating AT sets (20,000 repetitions). The used evidences comprise document, input channel, class, sender, as well as process specific information (person, organization, address, and date). The average rank of evidence types shows tremendous differences in search performance. Therefore, we selected the best performing and most probable appearing evidences. When using a single evidence, only three evidence types perform with average rank 2 or better. With more evidences, seven types perform well (rank < 2). We repeated the experiments for the *Top 7 - List*, and a two-step structure (*Top 7 - Structure*) where the three best performing evidence types are included in the first, and all remaining well performing types in the second step. Fig. 2 depicts the average

Table 2. Levenshtein distance.

LD	0	1	2	3	4	5	6	7	8	9
Avg Rank	1.18	1.17	1.17	1.37	1.56	2.19	3.78	4.91	5.07	5.41

ranks for each evidence group and evidence numbers from one to five. We conclude that search performance improves with the number of evidences and that choosing the best performing evidences stepwise is crucial.

4.3. Error Robustness

Experiments were conducted with LD from 0 to 9 with 3 evidences per search and only *Top 7* evidences. Table 2 depicts the average rank for each LD value. We observe that for LD 0 to 4 the performance is stable, but with LD higher than 4, search results lower. Stability can be explained by the good DA results of this corpus. We recommend to allow a LD of 3 or 4 to keep search results optimal and allow for error smoothening.

4.4. Calibration

To better understand how sensitive the choice of matching parameters is, we iterate the parameters for value mismatch p_{mm}, evidence not found p_{nf}, value expected p_{ve}, only value match p_{ov} in steps of 0.1 from 0.0 to 1.0. Experiments were conducted for 3 evidences from *Top 7* List and from *All* evidence types. Fig. **??** depicts the average ranks for all configurations:

1. *Value mismatch* is stable in a range between 0.1 and 0.4 for both evidence sets. Giving ATs with mismatching value a higher score leads to a deterioration of search results, especially with 1.0.

2. *Evidence not found* has no influence on search results for the *Top 7* set, but for *All* evidences, because the *Top 7* set contains only evidence types that appear in all ATs. Still, we recommend calibrating this parameter.

3. *Matching evidence types waiting for value* produces stable results between 0.1 and 0.9 and must be considered for search (> 0.0), but must be lower than matching values ($= 1.0$).

4. *Value match and evidence type mismatch* parameter does not influence search results, as it does not appear in our corpus.

All four parameters are relatively insensitive to small changes, which can be explained by the normalization step when computing the mass function. An initial calibration of parameters with all evidence types helps handling errors in the evidence top lists.

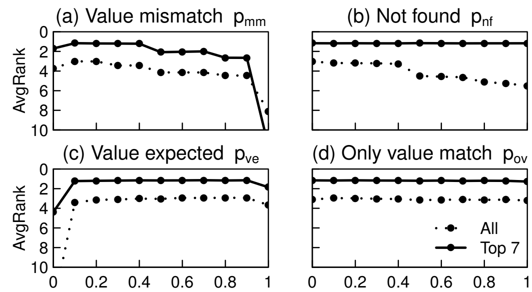

Figure 3. Parameters.

5. Conclusion and Future Work

We propose an approach to search the Attentive Task (AT) to an incoming document based on the available document analysis results with a degree of belief relying on the Dempster-Shafer theory allowing addition of new evidences. First evaluations show that types and number of evidences are crucial for search performance and that error robustness can be allowed up to Levenshtein distance 4 without deteriorating search performance. The calibration of the main parameters is insensitive to small changes and relatively simple to conduct. If parameters are calibrated for all evidences instead of prioritized ones, search results remain better.

Our next steps are evaluating the approach on a larger AT corpus, handling documents that are not related to any AT, and introducing a learning mechanism for the best performing evidence types.

References

[1] V. Bellotti et al. Quality vs. quantity: Email-centric task-management and its relationship with overload. *Human-Computer Interaction*, 20:1–2, 2005.

[2] W. W. Cohen, V. R. Carvalho, and T. M. Mitchell. Learning to classify email into ”speech acts”, 2004.

[3] A. Dengel and K. Hinkelmann. The specialist board - a technology workbench for document analysis and understanding. In *IDPT, 2nd World Conference on Design & Process Technology, Austin, TX, USA*, volume 2, pages 36–47, December 1996.

[4] M. Dredze, T. A. Lau, and N. Kushmerick. Automatically classifying emails into activities. In *IUI*, pages 70–77, 2006.

[5] N. Kushmerick and T. A. Lau. Automated email activity management: an unsupervised learning approach. In R. S. Amant, J. Riedl, and A. Jameson, editors, *IUI*, pages 67–74. ACM, 2005.

[6] G. Shafer. *A Mathematical Theory of Evidence*, volume 1. Princeton university press Princeton, 1976.

[7] K. Stamm and A. Dengel. Attentive tasks: Process-driven document analysis for multichannel documents. *Proceedings of DAS 2012*, 2012.