AI Trust in Business Processes: The Need for Process-Aware Explanations

Steve T. K. Jan,[1] **Vatche Ishakian,**[2] **Vinod Muthusamy**[2]
[1]Virginia Tech, [2]IBM Research AI

Abstract

Business processes underpin a large number of enterprise operations including processing loan applications, managing invoices, and insurance claims. There is a large opportunity for infusing AI to reduce cost or provide better customer experience and the BPM literature is rich in machine learning solutions. More recently, deep learning models have been applied to process predictions. Unfortunately, companies have applied or adopted very few of these innovations. We assert that a reason for this lack of adoption is that business users are risk-averse and do not implicitly trust AI models. We challenge the BPM community to build on the AI interpretability literature, and the AI Trust community to understand what it means to take advantage of business process artifacts in order to provide business level explanations.

Introduction

Business processes underpin a large number of enterprise operations including loan origination, invoice management, and insurance claims processing (Van Der Aalst and others 2011). The business process management (BPM) industry is expected to approach $16 billion by 2023 (Marketwatch 2019). There is a great opportunity for infusing AI to reduce cost or provide better customer experience (Rao and Verweij 2017), and the BPM literature is rich in machine learning solutions to gain insights on clusters of process traces (Nguyen et al. 2016; 2019), predict outcomes (Breuker et al. 2016), and recommend decisions (Mannhardt et al. 2016). Deep learning models including from the NLP domain have also been applied (Tax et al. 2017; Evermann, Rehse, and Fettke 2017).

Unfortunately, very little of these innovations have been applied and adopted by enterprise companies (Daugherty and Wilson 2018), and those adopted are limited to narrow domains such as customer services, enterprise risk and compliance (Wilson, Alter, and Shukla 2016).

We assert that a large reason for the lack of adoption of AI models in BPM is that business users are risk-averse and do not implicitly trust AI models. There has been little attention paid to *explaining model predictions to business users with process context*. These business users are typically experts in their fields but not data scientists, and explanations must be presented in their business domain vocabulary. We challenge the BPM community to build on the AI interpretability literature, and the AI Trust community to take advantage of business process artifacts.

Example of process-aware explanations

Consider the example loan application process in Fig. 1. Suppose we build a sequence model that takes as input the activities and features observed in the process and predicts the outcome, in this case whether a loan will be approved. Such models have been shown to achieve accuracies of up to 85% (Evermann, Rehse, and Fettke 2017). We can then use tools such as LIME (Ribeiro, Singh, and Guestrin 2016) to explain the prediction, and as we see in Fig. 2, LIME suggests that the presence of the skilled agent activity causes the application to be rejected.

A subject matter expert, however, would understand from the process description that large loan requests from borrowers with low credit scores are the ones most likely to end up routed to a skilled agent. This is an example of a causal relationship that can be inferred from the process description: the LOAN AMOUNT and/or CREDIT SCORE features are the cause of the feature associated with the skilled agent activity.

The explanations from LIME are based on sampling perturbations around the input features and measuring how the predictions change with the perturbations. It turns out in our example, that many of the perturbations do not conform to the process description and hence can never occur. Once we apply the causal relationship above to constrain the perturbation sampling, LIME offers the explanation in Fig. 3 where the CREDIT SCORE is now prominent.

This simple example illustrates how directly applying interpretability techniques to process models results in incomplete or potentially misleading explanations. We also see how by being process-aware we can augment existing algorithms to improve the quality of explanations. This is a nascent and fertile research area that we have only scratched the surface of.

As stated earlier, virtually all AI models in the BPM literature train models with features from process traces, depicted as "state of the art" in Fig. 4.There is an opportunity to apply known interpretability approaches in a BPM context, denoted as LEVEL 1 in Fig. 4. For example, a regression model to predict process completion time (Polato et al. 2014) can be augmented by techniques to make regression models more interpretable (Schielzeth 2010). The same can apply to deep learning models for process prediction tasks. Most of the current techniques, however, are based on sequence models including LSTMs and RNNs (Tax et al. 2017; Evermann, Rehse, and Fettke 2017). While there has been a lot of research on explaining deep learning models based on CNNs for image classification tasks (Gan et

Figure 1: Example loan application business process

Figure 2: Vanilla LIME: The explanation is incomplete showing only that the loan is rejected because it is sent to a skilled agent.

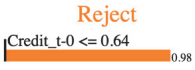

Figure 3: Process-aware LIME: The explanation now highlights the influence of the causal CREDIT SCORE feature.

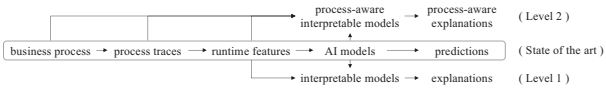

Figure 4: Approaches for interpretability in BPM AI models

al. 2015), there is a growing interest in similar problems for sequence models (Li et al. 2015). LEMNA attempts to offer high fidelity explanations for deep learning models (Guo et al. 2018) but assumes a security context where adjacent features are dependent on each other, an assumption that does not always hold true in business processes. Note that LEVEL 1 interpretability only uses the trained model (either as a black box or white box) and possibly features from the training data. Of course, applying known interpretability techniques to the new business process domain may not be straightforward and require additional innovations, but LEVEL 1 at least offers researchers a pathway to begin experimenting.

Challenges to trustworthy AI for BPM

A more ambitious, and promising, approach is to bring *process-awareness* to the problem, marked as LEVEL 2 in Fig. 4. Here, interpretability models would take advantage of the knowledge of the business process definitions and full runtime process traces. Some of the information in these artifacts is typically lost when preparing the data for the predictive models. There is, unfortunately, a dearth of solutions that apply this approach. An example of LEVEL 2 was presented in the previous section where the black box interpretability model in LIME is augmented with knowledge of the causal relationships derived from the business process definition. An understanding of the feature causality graph avoids misleading or incomplete explanations.

Business process datasets[1] are publicly available, but as part of this research area the community will need to develop metrics to measure the quality of the explanations. Standard techniques that measure how model accuracy degrades as features are removed from the dataset may make sense in other machine learning domains, such as visual explanations of image classification models (Adebayo et al. 2018), but may not be appropriate for business process.

We think process-aware explanations is an interesting research area with potential for high business impact.

[1]https://data.4tu.nl/repository/collection:event_logs_real

References

Adebayo, J.; Gilmer, J.; Muelly, M.; Goodfellow, I.; Hardt, M.; and Kim, B. 2018. Sanity checks for saliency maps. In *The 32Nd NeurIPS*.

Breuker, D.; Matzner, M.; Delfmann, P.; and Becker, J. 2016. Comprehensible predictive models for business processes. *MIS Quarterly*.

Daugherty, P. R., and Wilson, H. J. 2018. *Human+ machine: reimagining work in the age of AI*. Harvard Business Press.

Evermann, J.; Rehse, J.-R.; and Fettke, P. 2017. Predicting process behaviour using deep learning. *Decision Support Systems*.

Gan, C.; Wang, N.; Yang, Y.; Yeung, D.-Y.; and Hauptmann., A. G. 2015. Devnet: A deep event network for multimedia event detection and evidence recounting. In *The 28th CVPR*.

Guo, W.; Mu, D.; Xu, J.; Su, P.; Wang, G.; and Xing, X. 2018. Lemna: Explaining deep learning based security applications. In *The 25th ACM CCS*.

Li, J.; Chen, X.; Hovy, E.; and Jurafsky, D. 2015. Visualizing and understanding neural models in nlp. *arXiv preprint arXiv:1506.01066*.

Mannhardt, F.; De Leoni, M.; Reijers, H. A.; and Van Der Aalst, W. M. 2016. Decision mining revisited-discovering overlapping rules. In *CAiSE*.

Marketwatch. 2019. Business Process Management (BPM) Market 2019: Key Findings, Regional Study, Size, Growth and Global Trends by Forecast to 2023. Online; accessed 2019.

Nguyen, P.; Slominski, A.; Muthusamy, V.; Ishakian, V.; and Nahrstedt, K. 2016. Process trace clustering: A heterogeneous information network approach. In *In SIAM SDM*.

Nguyen, P.; Ishakian, V.; Muthusamy, V.; and Slominski, A. 2019. Summarized: Efficient framework for analyzing multidimensional process traces under edit-distance constraint. *arXiv preprint arXiv:1905.00983*.

Polato, M.; Sperduti, A.; Burattin, A.; and de Leoni, M. 2014. Data-aware remaining time prediction of business process instances. In *IJCNN*.

Rao, D. A. S., and Verweij, G. 2017. Sizing the prize: Whats the real value of AI for your business and how can you capitalise? *PwC Publication, PwC*.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. Why should I trust you?: Explaining the predictions of any classifier. In *the 22nd ACM SIGKDD*.

Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* 1(2):103–113.

Tax, N.; Verenich, I.; La Rosa, M.; and Dumas, M. 2017. Predictive business process monitoring with lstm neural networks. In *CAiSE*, 477–492. Springer.

Van Der Aalst, W., et al. 2011. Process mining manifesto. In *BPM*. Springer.

Wilson, H.; Alter, A.; and Shukla, P. 2016. Companies are reimagining business processes with algorithms. *Harvard Business Review*.