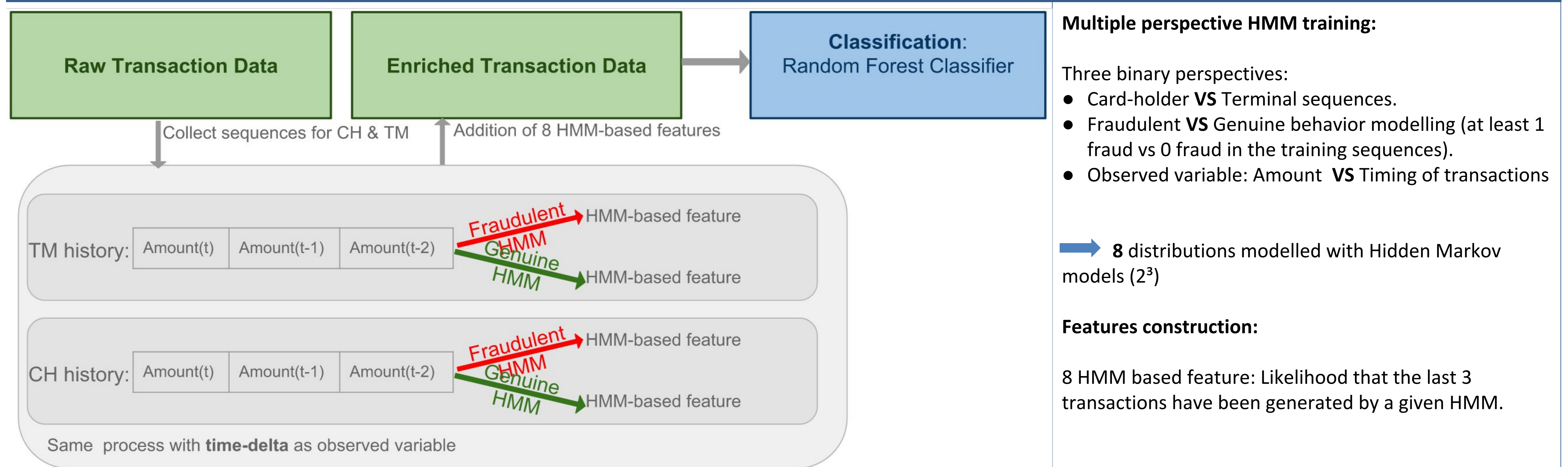


State of the art State of the art feature engineering for credit card fraud detection consists in creating descriptive features that contain historical knowledge about the card-holder. These single statistics are called aggregated features and can be for example: the amount spent or the number of transactions done in a certain amount of time with the same merchant category or country [2],[3].

Motivation

State of the art weaknesses	Proposed approach
Descriptive features are single statistics	Explicit generative models of the sequences of transactions
Single statistics based on expert knowledge	Supervised modelling of amount and timing behaviors
Only card-holder based features	Modelling card-holder and terminal sequences

Multiple perspective HMM-based feature engineering



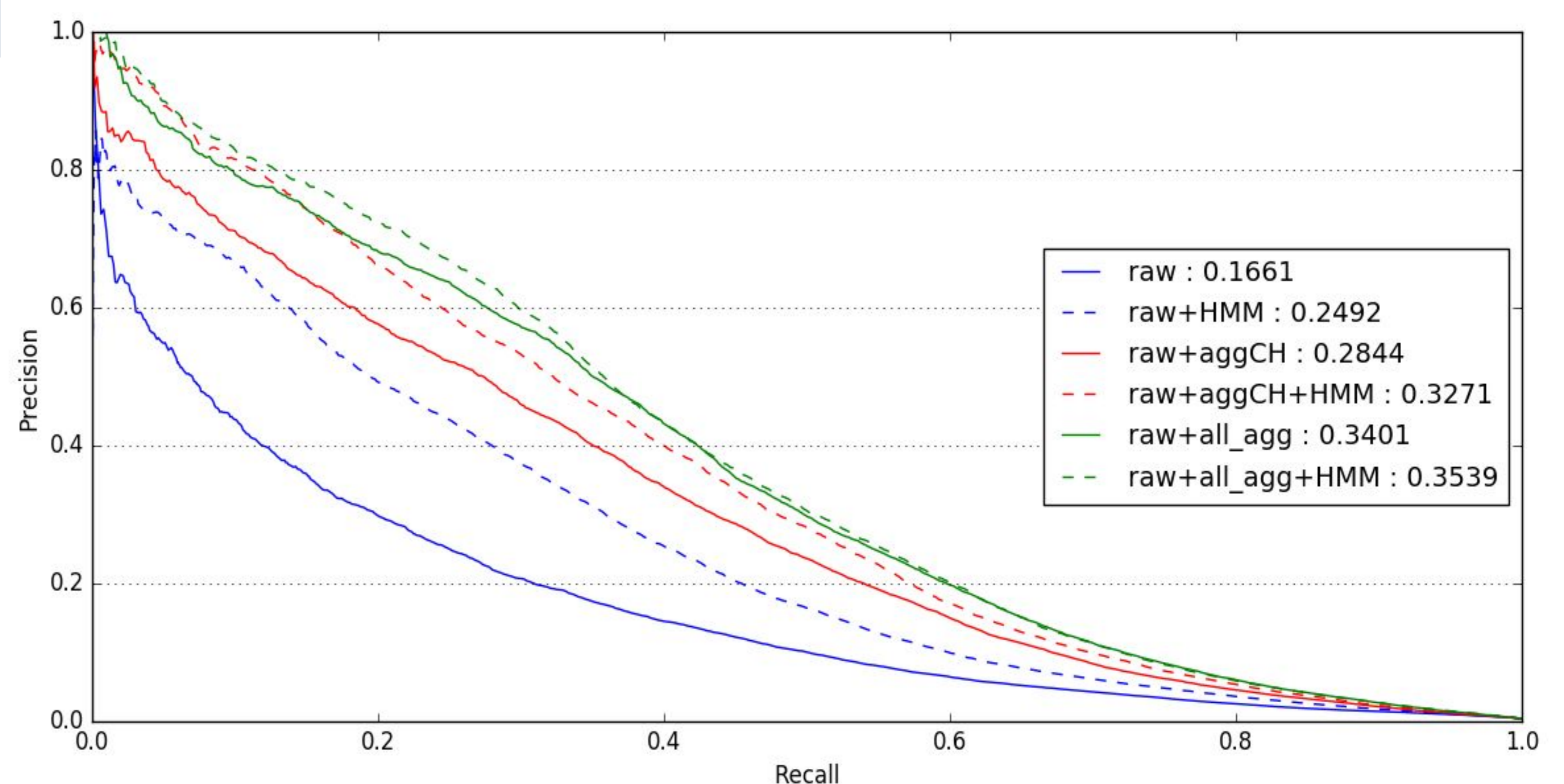
Experimental setup

Real world dataset: $4.7 \cdot 10^7$ credit card transactions from Belgium between 01.03.2015 and 31.05.2015. **raw** feature set contains 15 features describing the card-holder, the merchant and the transaction.
Classifier: Random forest classifier.
Metric: Precision Recall AUC [4].

For comparison purpose:

- 4 card-holder centered descriptive features [2], [3] (**aggCH**).
- 4 terminal centered descriptive features in order to quantify the increase due to HMM modeling and the increase due to terminal sequences information. (contained in **all_agg**)

Dotted curves: prediction with the addition of 8 HMM-based features



Improvement in fraud detection when using HMM-based features

Multiperspectives HMM-based features are a strong tool to improve fraud detection :

- 15% increase in fraud detection compared to state of the art [3].
- Terminal perspective brings 20% increase by itself.

Interesting properties :

- HMM models the dependencies between the events of the sequences.
- Modelling different perspectives provides a fine-grained description of the past of the transactions.
- This feature engineering can be extended to any dataset with interactions between users.

The framework can be found at: https://gitlab.com/Yvan_Lucas/hmm-ccfd

References

1. Bolton R., Hand D. J. (2001) Unsupervised profiling methods for fraud detection. *Credit scoring and credit control VII*.
2. Bahnsen A. C., Aouada D., Stojanovic A., Ottersten B. (2016) Feature engineering strategies for credit card fraud detection. *Expert systems with Applications*.
3. Whitrow C., Hand D. J., Juszczak P., Weston D. J. and Adams N. M. (2008) Transaction aggregation strategy for credit card fraud detection. *Data Mining and Knowledge Discovery 18 (1)*.
4. Davis J., Goadrich M. (2006) The relationship between precision-recall and roc curves. *ICML 06 Proceedings of the 23rd international conference on Machine learning*.