Dataset	Damerau–Levenshtein Distance	
	Mean	SD
BPI2013.Incidents	.563	.199
BPI2013.Problems	.616	.177
BPI2012 (completion events)	.659	.203
BPI2012.W (completion events)	.703	.205
BPI2012.W (all events)	.697	.211
BPI2012.A	.545	.241
BPI2012.O	.532	.191

Table 8: Mean and standard deviation of Damerau-Levenshtein distance between actual trace remainders and those predicted from a prefix of length 5. Hallucinations produced using probability sampling ($k = 1$) and element-wise feedback ($m = 1$); smaller is better.

hallucination is initialized with a trace prefix and then continued until it produces an end-of-case indicator. The hallucination can then be compared with the actual trace continuation using a string-edit distance. We train nets with 32 embedding dimensions and 5 unroll steps for 100 epochs. Using trace prefixes of length 5, we produce hallucinations and compare them to the actual continuation using the normed Damerau–Levenshtein distance, which ranges from 0 to 1 (Table 8).

The application of hallucinations might be interesting in other areas of process mining as well, such as improving event log completeness in cases where particular mining algorithms benefit from better or larger logs[5].

6.6.2. Hidden State Dynamics

Recent work on understanding RNNs also focuses on visualization. In particular, visualizations of the embedding matrix, the state activation and the state dynamics are useful in understanding how an RNN encodes knowledge (Karpathy et al., 2015; Li et al., 2015; Yosinski et al., 2015; Strobelt et al., 2016).

We export embedding matrices after completion of training to create 2D t-SNE plots (Maaten and Hinton, 2008). They are not included here as no significant or

[5] We thank one of the anonymous reviewers for this suggestion.