Use of Process Crowding in Conditional WGAN for Remaining Process Events Prediction

Yoann Valero - LIST3N | Livejourney - yoann.valero@utt.fr
Pr Frédéric Bertrand - LIST3N - frederic.bertrand@utt.fr
Dr Myriam Maumy - LIST3N - myriam.maumy@utt.fr

Background

Process data → event logs:

Case ID	Activity	Date / time
YV	Plane	2022-09-05 10:25
YV	Airport	2022-09-05 19:50
YV	SDSS	2022-09-05 20:25

Event (rows) / Case (columns)

From an unfinished case (**prefix**), we need to predict the remaining events (**suffix**). We use a Conditional WGAN with **"crowding"** information to do so.

Definitions

In a log:
- U: space of units (= cases)
- T: space of recorded times
- A: space of possible activities

Event: $e = (u, t, t', a),\ u \in U,\ t < t' \in T,\ a \in A$.
Let said log contain n events with $i = 1, \ldots, n$:

$$act(u,t) = \left\{ e_4^{(i)}; (u_i = u) \wedge (t \in [t_i, t_i'[) \right\}$$

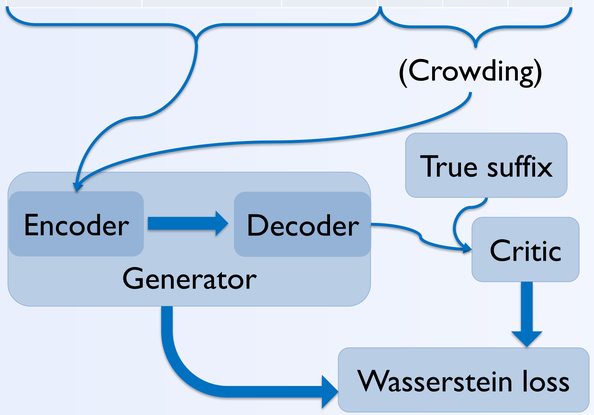
act(u, t) gives the activity u is in at time t.
Then :

$$C(t) = \left(\sum_{i=1}^{n} \mathbb{I}\left\{ u_i; act(u_i,t) = a \right\}, a = a_1, \ldots, a_{|A|} \right)$$

C(t) gives the crowding of all activities at time t.

Methods

Case ID	Activity	Date	A1	A2	A3
U1	A1	t1	10	5	35
U1	A2	t2	9	6	28



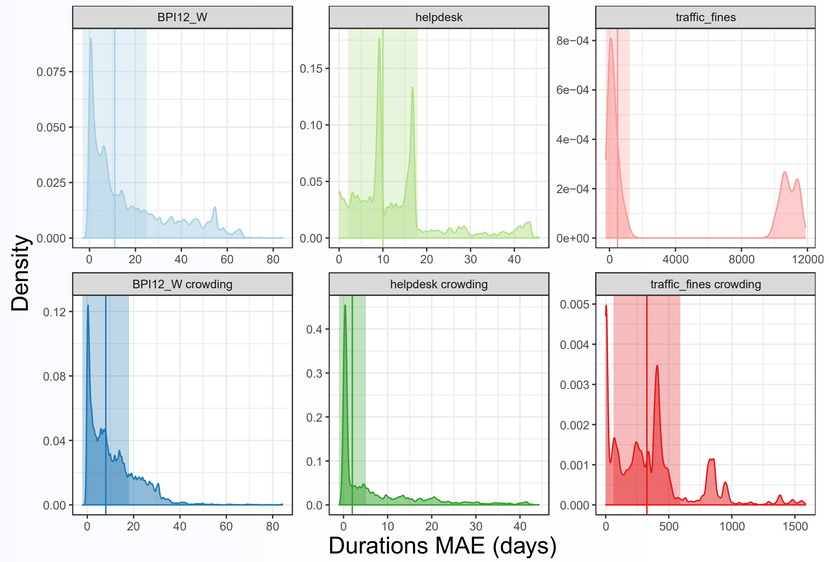
Results

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{t}_i^{\,\text{crowd}} - t_i\right|}{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{t}_i - t_i\right|} :$$

- BPI 2012 (W) : **0,538**
- Helpdesk : **0,611**
- Traffic fines : **0,086**

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\mathrm{DL}\left(\hat{y}_i^{\,\text{crowd}}, y_i\right)}{\frac{1}{n}\sum_{i=1}^{n}\mathrm{DL}\left(\hat{y}_i, y_i\right)} :$$

- BPI 2012 (W) : **1,04**
- Helpdesk : **0,98**
- Traffic fines : **1,01**



Conclusions

- **Crowding** greatly **improves times** predictions
  - Like a GPS: including traffic is better
- **Precision** of **activity** predictions is maintained
- **Wasserstein** loss makes training way more **stable**

References

1. F. Taymouri et M. L. Rosa, Encoder-Decoder Generative Adversarial Nets for Suffix Generation and Remaining Time Prediction of Business Process Models, 2020. arXiv : 2007.16030[cs.LG]
2. M. Arjovsky et al., Wasserstein GAN, 2017. arXiv : 1701.07875 [stat.ML]
3. M. Mirza et S. Osindero, Conditional Generative Adversarial Nets, 2014. arXiv : 1411.1784 [cs.LG]