

Inference Networks in SPENs

Mechanical Turks

(Kalpesh Krishna, Arka Sadhu, Suman Swaroop and Anish Senathi)

Main Goals

- Study the whole set of papers on SPENs
- Implement an ICLR '18 paper on Inference Networks for SPENs
- Investigate performance gains on a KB completion task
- Investigate advantages of W-GAN style training

Related Work

- Belanger & McCallum 2016 - “Structured Prediction Energy Networks”
- Belanger et al. 2017 - “E2E Training of SPENs”
- Tu & Gimpel 2018 - “Learning Approximate Inference Networks for Structured Prediction”

Non SPEN work -

- Yaghoobzadeh et al. 2015 - “Corpus-level Fine-grained entity typing using contextual information.”
- Hinton et al. 2015 - “Distilling the Knowledge in a Neural Network”
- Goodfellow et al. 2014 - “Generative Adversarial Networks”
- Gulrajani et al. 2017 - “Improving the Training of W-GANs”

Datasets

- **BibTeX** - Standard MLC dataset, 1836 input features, 159 labels. 4880 / 2515 train / test split.
- **Bookmarks** - Standard MLC dataset, 2150 input features, 208 labels. 48000 / 12000 / 27856 train / dev / test split
- **FIGMENT** - KB Completion, 200 input features, 102 labels. 101266 / 40220 / 60447 train / dev / test split

Implementation

- **Stage 0** :- Pretrain $F(x)$ + Local Energy Term
- **Stage 1** :- Adversarial Training of Inference Network vs Energy Network
- **Stage 2** :- Re-tuning of Inference Network using Energy Network (similar to distillation)

(lots of details in report)

Extension to the Paper

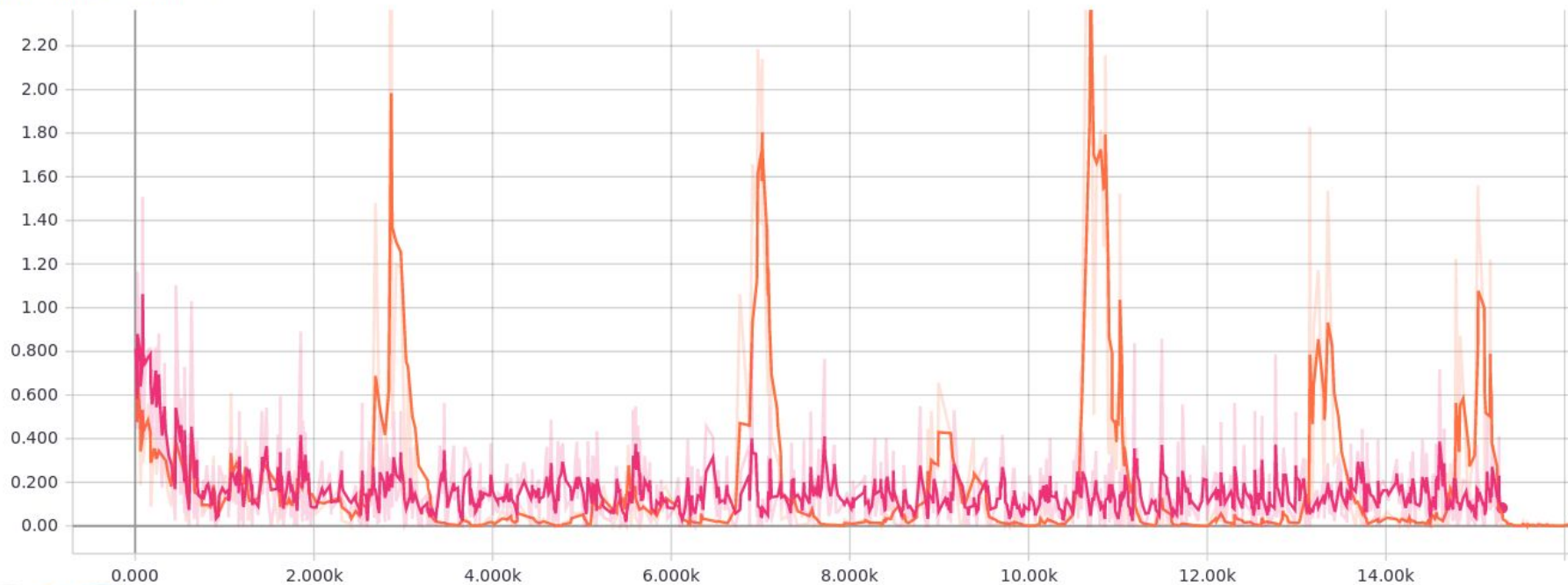
- Energy Networks act like discriminators
- If we can enforce Lipschitz Continuity, we can model W-GANs
- Following Gulrajani et al. 2017, we penalize the gradient of the Energy Network with respect to the input labels (y)

Code & Experimental Platform

- <http://github.com/theshadow29/infnet-spen>
- TensorFlow 1.4+
- 1881 lines of code in the repository
- Dell XPS 13, running an 8th generation Intel i5 processor with 8 cores (after hyperthreading) and 8GB RAM

Convergence of Base Objective

logger/base_objective_1



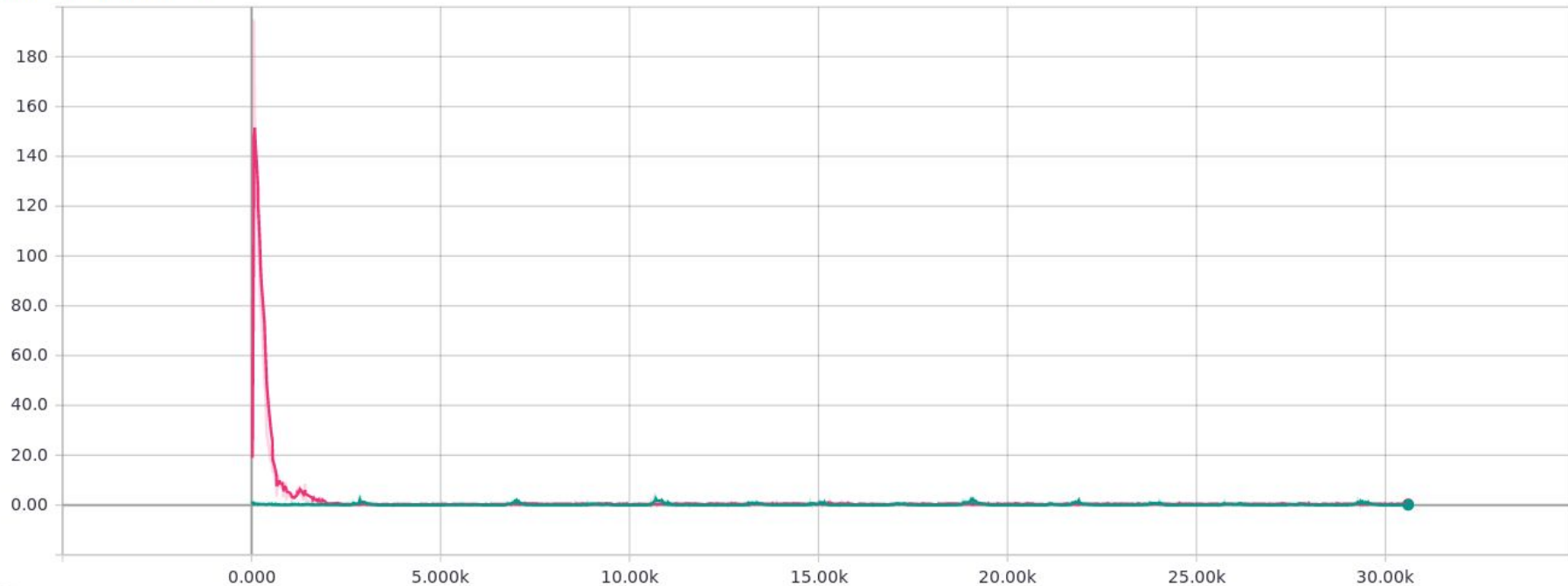
Convergence with W-GANs

logger/base_objective_1



Convergence with W-GANs - 2

logger/base_objective_1



F1-Score Results

Table 1: F1-scores macro-averaged. Reported figures are taken from [Tu and Gimpel \[2018\]](#)

		BibTex	Bookmarks
MLP	OurModel-MLP	0.3787	0.3203
	Reported-MLP	0.389	0.338
Infnet	OurModel	0.4221	0.3574
	Reported-SPEN-BM16	0.422	0.344
	Reported-SPEN-E2E	0.381	0.339
	Reported-InfNet	0.422	0.376

FIGMENT Accuracy Results

Table 2: Accuracy measures for the Figment Dataset. Reported measures is taken from Yaghoobzadeh and Schütze [2016]

	Test Set
OurModel - MLP	0.4463
OurModel	0.4565
Reported-Figment(GM)	0.426

Timing Analysis

Table 3: Timing Analysis. Batch size of 32 is used everywhere. All values are examples processed per second

		Training Time (On train set)			Inference Time (On dev set)		
		BibTex	Bookmarks	Figment	BibTex	Bookmarks	Figment
MLP	OurModel	33904.14	35723.41	50643.78	47985.57	46793.62	88182.58
	Reported-MLP	21670	19591	-	90706	92307	-
Infnet	OurModel	13275.04	12344.86	15013.35	49817.87	46429.22	85749.82
	Reported-SPEN-E2E	551	559	-	1420	1401	-
	Reported-InfNet	5533	5467	-	94194	88888	-

Timeline

- March (week 1 to week 3) - Abstract Submission + Reading
- March (week 4) to April (week 1) - FIGMENT pre-processing, initial implementation
- April (week 2 to week 3) - Technical correspondence with authors, important misunderstandings fixed. Partly implemented Belanger et al. 2016. BibTeX added
- April (week 4) to May (week 1) - Fixed mistakes, Bookmarks added, investigated Gulrajani et al. 2017 style training.

Challenges

After talking to authors, we noticed that these four mistakes (which were not clear in paper) hampered our progress -

- Not having a classification threshold
- Not pre-training local energy jointly with $F(x)$
- Using $\text{sum}(\text{loss})$ rather than $\text{mean}(\text{loss})$ with mentioned hyperparameters
- Using label-averaged F1 instead of example-averaged F1

Acknowledgements

- **Lifu Tu** :- Exchanging 10-15 emails with us, helping us through the very frustrating debug phase
- **Prof. Kevin Gimpel** :- Suggesting FIGMENT + some novel energy function ideas + ideas on synthetic sequence labelling tasks
- **Prof. Soumen Chakrabarti** :- Sharing his copy of the ClueWeb09 dataset

Thank You!

Questions?