

## **DL4MT Winter School Final Report**

### **Overview**

The first DL4MT Winter School was organised to provide an intensive week of training in the application of emerging Deep Learning technologies to Machine Translation. Professors Kevin Duh, Hermann Ney, and Kyunghyun Cho each presented two days of lectures, delving into the details of Artificial Neural Network implementations for translation tasks. Practical sessions during the afternoons gave participants the opportunity to work with real systems, and to develop their expertise by implementing the components of deep models for several Natural Language Processing tasks. Participants who attended all of the sessions received an excellent introduction to Deep Learning for Machine Translation, and left the school with the tools to use Deep Learning in their own research.

After an application and review period, 75 participants registered for the school – four of whom were granted a 500 euro scholarship along with a waiver of the registration fee. The participants came from five continents and 22 different countries. The geographic distribution of the participants is given below:

- 43 Europe
- 13 Asia
- 5 North-America
- 1 South-America
- 1 Africa
- 13 Unknown / not in the application form

Other participants include:

- 5 organisers
- 9 mentors (7 external + 2 internal)
- 10 volunteers
- 1 visitor

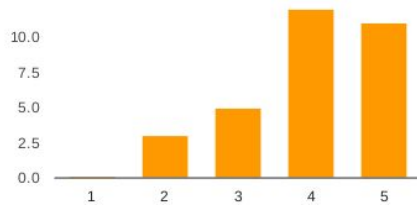
### **Event Summary**

Each day of the Winter School was divided into a three-hour block of lectures in the morning, and three hours of practical sessions in the afternoon. A poster session was organized for Tuesday evening, in which six workshop participants presented original research on topics related to Deep Learning for Machine Translation. A guest lecture from Dr. Michael Lysaght of the Irish Centre for High-End Computing (ICHEC) about GPUs and related technologies in HPC took place on Wednesday evening. On Thursday, a full-day excursion visited several Neolithic sites in Ireland. See the appendix for a full schedule of the school.

## Feedback

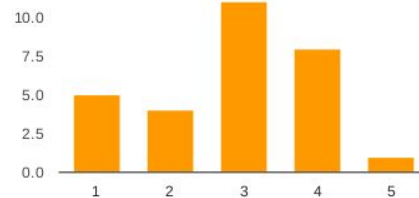
We conducted an anonymous survey of all participants in order to collect feedback about the DL4MT Winter school, and to improve the organization of future events. The overall response was very positive; however, the lab sessions generally received much lower ratings than the rest of the school, primarily due to complaints about the space not being ideal for programming exercises, and the different labs not having a clear relationship between one another.

**Overall Experience**



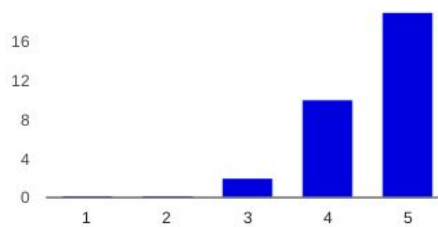
Very Bad: 1	<b>0</b>	0%
2	<b>3</b>	9.7%
3	<b>5</b>	16.1%
4	<b>12</b>	38.7%
Very Good: 5	<b>11</b>	35.5%

**Lab Sessions**



Very Bad: 1	<b>5</b>	17.2%
2	<b>4</b>	13.8%
3	<b>11</b>	37.9%
4	<b>8</b>	27.6%
Very Good: 5	<b>1</b>	3.4%

**Invited Mentors and Morning Talks**



Very Bad: 1	<b>0</b>	0%
2	<b>0</b>	0%
3	<b>2</b>	6.5%
4	<b>10</b>	32.3%
Very Good: 5	<b>19</b>	61.3%

## **Materials**

The school also produced valuable resources which have already proven useful to many students who were not able to attend the event. Most of these resources are open-source, and provide an excellent starting point for those interested in the application of Deep Learning to Machine Translation and Natural Language Processing. Two repositories of lab exercises are freely available on github, and Kyunghyun Cho's repository implementing the encoder-decoder model for translation has already reached more than 100 stars, indicating a wide-reaching impact of this work beyond the participants of the DL4MT school.

## **Conclusions**

Overall, the school was a great success, thanks to the hard work of the local team from DCU, and the excellent presentations given by the mentors. The documentation, shared folders, correspondence, and all minutes from the planning meetings have been archived in order to facilitate the organization of future events by this team.

## **Appendix:**

### **Detailed Timetable for DL4MT Winter School 18-24.10.2015**

18.10.15 (Sunday) Kevin Duh	Morning 9:00-12:30	Open Session (10 min) <a href="#">Fundamentals for DL4MT I</a> Quick review of MLP & Backprop Deep Belief Nets, Stacked Auto-Encoders Tricks: drop-out, distributed training, Hessian-free optimization, ReLu, etc.
	Lab 14:00-17:30	Goal: Prep and setup. Compare logistic regression, MLP, and stacked auto-encoders on same data Challenges: Implement tricks
	Evening 19:00-20:00	Free
19.10.15 (Monday) Kevin Duh	Morning 9:00-12:30	<a href="#">Fundamentals for DL4MT II</a> Survey of recent papers from machine learning (NIPS/ICML/ICLR) and related application fields besides NLP (ASR/CV) Some example uses of Deep Learning in NLP: word embedding,

		phrase/sentence embedding, (this part by Tsuyoshi Okita)
	Lab 14:00-17:30	Goal: word embedding (SGNS), visualization. Challenges: implementation tricks
	Evening 19:00-20:00	Free
20.10.15 (Tuesday) Hermann Ney	Morning 9:00-12:30	Neural LMs and TMs for SMT I fundamentals of neural networks (NN, including deep neural networks); NN structures feedforward multilayer perceptron recurrent neural network (RNN) for sequence modelling Long-Short Term Memory RNNs for sequence modelling.
	Lab 14:00-17:30	Goal: Rescore n-best lists using RNN LM Challenges: implement activation functions, compare full vs. class-factored output layers, implement stopping criteria for training data preprocessing FF/RNN/LSTM LM training (toy example) n-best rescoring
	Evening 19:00-20:00	Poster Session (with spotlight presentation)
21.10.15 (Wednesday) Hermann Ney	Morning 9:00-12:30	Neural LM and TM for SMT II use of NNs for language modelling (in speech recognition and translation) use of NNs for translation modelling.
	Lab 14:00-17:30	Goal: n-best rescoring using uni/bidirectional translation and joint models Challenges: implement the concatenation of source and target embeddings and compare that to the sum, introduce parallel sections to speed up the RNN evaluation data preprocessing (Bidirectional) joint model training (toy example)

		(Bidirectional) translation model training (toy example) n-best rescoring
	Evening 19:00-20:00	<p>Extreme-scale computing - challenges and opportunities on the road to Exascale. Dr. Michael Lysaght, ICHEC.</p> <p>Abstract: With Exascale systems on the horizon at the same time that conventional compute architectures are suffering from rising power densities, we are facing an era with power, energy-efficiency, and cooling as major constraints for scalable High Performance Computing (HPC). In this talk, I will provide an overview of how GPGPU technology may offer a means to address some of these challenges and also compare GPGPUs against other emerging many-core technologies on the road to exascale computing, as well as highlighting the relevance to machine learning.</p> <p>Bio: Dr. Michael Lysaght leads the Novel Technologies Activity and the Intel Parallel Computing Centre at the Irish Centre for High End Computing (ICHEC), where he has a particular focus on supporting the Irish scientific user community and Irish industry in the exploitation of emerging multi-/many-core technologies. In conjunction with his role at ICHEC, In conjunction with his role at ICHEC, he is a Science Foundation Ireland (SFI) Funded Investigator in the Irish Centre for Software Engineering (LERO)&amp;. Michael joined ICHEC in 2011 after working in the UK as a HPC application expert as part of HECToR's distributed Computational Science and Engineering program, where he worked on re-factoring and optimising community codes for the UK research community. Prior to this he worked for three years as a UK EPSRC Postdoctoral Research Fellow in theoretical atomic physics at Queen's University Belfast, where he pioneered the development of Time-Dependent R-Matrix Theory and associated parallel applications including the TDRM and RMT codes. Michael obtained his PhD in physics in 2006 from University College Dublin and has written 32 journal publications, including a book chapter on theoretical and numerical methods for quantum dynamic imaging in the Springer CRM Series in Mathematical Physics.</p>
22.10.15 (Thursday)	Excursion	Excursion
23.10.15 (Friday) Kyunghyun Cho	Morning 9:00-12:30	Neural MT I n-gram => feedforward LM => RNN LM

		Encoder-decoder architecture for machine translation Kalchbrenner and Blunsom (2013) Sutskever et al. (2014) Cho et al. (2014)
	Lab 14:00-17:30	- Data Preparation: Basic preprocessing - [Optional: RNN-LM with Theano] - Encoder-Decoder with Theano (without attention) - Line-by-line explanation of the code - Trained models will be provided
	Evening 19:00-20:00	Free
24.10.15 (Saturday) Kyunghyun Cho	Morning 9:00-12:30	Neural MT II  Attention-based neural machine translation with large vocabulary Bahdanau et al. (2015) Jean et al. (2015)  Incorporating monolingual corpora for neural machine translation Gulcehre&Firat et al. (2015)  Beyond language translation: image/video caption generation and Q&A  Xu et al. (2015) Yao et al. (2015) Hermann et al. (2015)
	Lab 14:00-17:30	- Attention-based Encoder-Decoder with Theano - Line-by-line explanation of the code - Trained models will be provided
	17:30-17:45	Closing Session