An Introduction to Neural Networks and Uses in EDM

Long Short-Term Memory (LSTM), Attention mechanism and Transformers

Ange Tato
École de Technologie Supérieure
Montreal, Canada

Agenda

- Deep Knowledge Tracing (DKT) on ASSISTments dataset
- DKT + Attention Mechanism on Logic-Muse dataset

Deep Knowledge Tracing (DKT) on Assistments dataset

- **ASSISTments dataset**:
  - *Assistments* is an online tutor that simultaneously teaches and assesses students in grade school mathematics.
  - Official link of the dataset : https://sites.google.com/site/assistmentsdata/
  - We will use the skill builder 2009-2010 version. You have access to it through the first notebook.
  - Dataset size = 525534
  - Number of students = 4217 and the Number of problems = 124

Deep Knowledge Tracing (DKT) on Assistments dataset

- **Data Format**:
  - Two files: a **dataset file** and a **split file**.
  - The dataset file contains **student**, **skill** and **performance** information and the split file specifies which students belong to the training set.
  - The dataset file is a 3-column space-delimited file. Each row in the file indicates whether a particular student answered a specific problem correctly or not. The first column is the student id, the second column is the skill id associated with the problem and the last column is whether the student got the problem correctly (1) or not (0).

Deep Knowledge Tracing (DKT) on Assistments dataset

- **Data Format**: dataset file
  - The dataset file is a 3-column space-delimited file. Each row in the file indicates whether a particular student answered a specific problem correctly or not.
  - The first column is the student id, the second column is the skill id associated with the problem and the last column is whether the student got the problem correctly (1) or not (0).

```
0 7 1
1 82 0
1 82 0
1 82 0
1 82 0
1 82 0
1 82 0
1 82 0
2 6 1
2 6 1
2 6 0
2 6 1
```

Deep Knowledge Tracing (DKT) on Assistments dataset

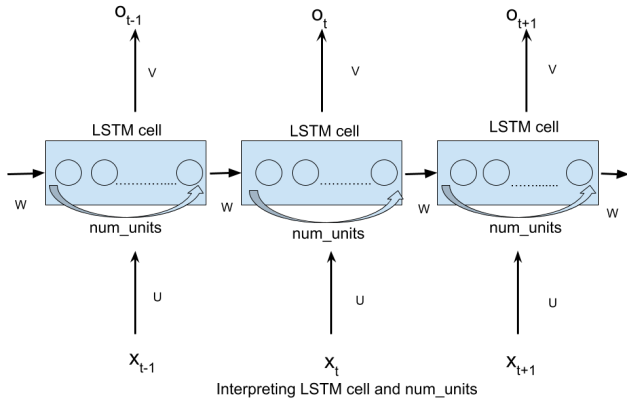
- **Data Format**: split file
  - The split file is a space-delimited file where each column indicates whether the corresponding student id should be in the training set (1) or not (0).
  - For example, the split file:

```
1 1 0 1 1 0
```

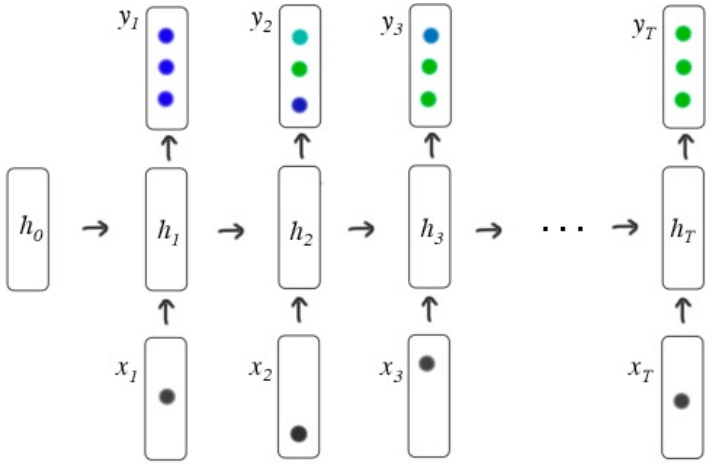
  - indicates that students 0, 1, 3 and 4 should be in the training set and the rest will be in the test set.

Deep Knowledge Tracing (DKT) on Assistments dataset

- **Model** : DKT (similar to the original paper – Piech et al. 2015)
  - Input dim = number of skills * 2
  - Output dim = number of skills
  - Time step = 100
  - Hidden units = 200
  - Batch size = 5



Interpreting LSTM cell and num_units




```
Layer (type)                 Output Shape              Param #
=================================================================
masking_1 (Masking)          (5, 100, 248)             0
_________________________________________________________________
lstm_1 (LSTM)                (5, 100, 200)             359200
_________________________________________________________________
time_distributed_1 (TimeDist (5, 100, 124)             24924
=================================================================
Total params: 384,124
Trainable params: 384,124
Non-trainable params: 0
```

Deep Knowledge Tracing (DKT) on Assistments dataset

- **Data preprocessing** :
  - **Input X** is size **2 * num_skills = 2 * 124**, one hot encoding
  - **Example** : "**2 3 1"** in dataset file => student id 2, skill = 3, answer = 1 (correct)
    - 3 (skill) * 2 + answer = 7, X[7] = 1 => X = [0,0,0,0,0,0,0,1,0,0,0,0,…]
  - Output **Y** (pred) is size equals to num_skills **= 124**
  - **Y** (true) is size equals to num_skills **= 124 + 1** (first 124 to specify the skill and the last cell is to specify the answer).

Deep Knowledge Tracing (DKT) on Assistments dataset

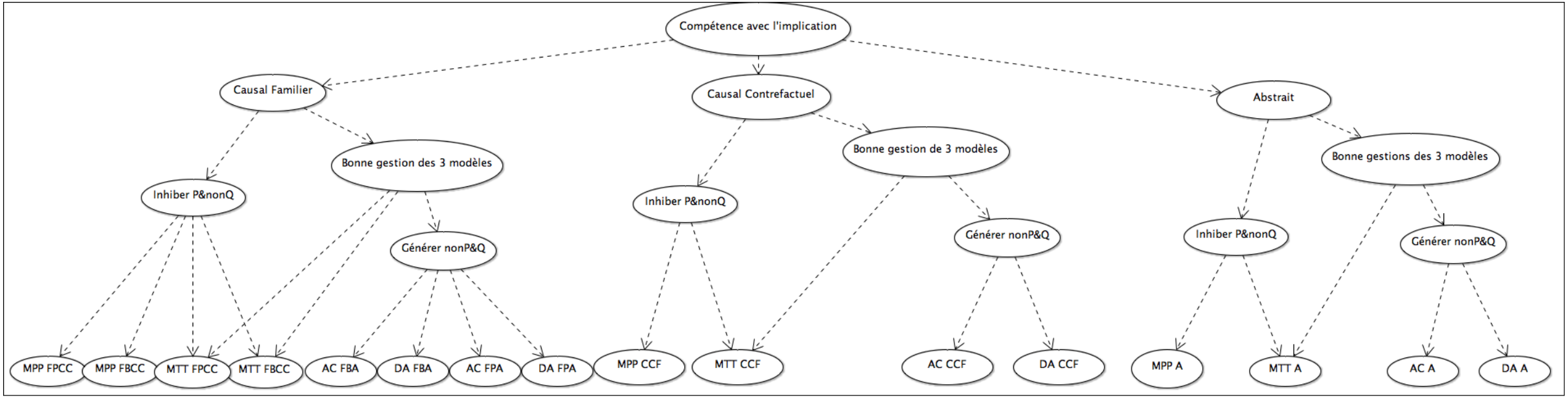
- **Results (first jupyter notebook)**:
  - AUC Deep Knowledge Tracing = **0.85** after 50 epochs**.**
  - AUC Bayesian Knowledge Tracing = 0.69 (see the original paper)

DKT + Attention Mechanism on Logic-Muse dataset

- **Logic-Muse dataset**:
  - Logic-Muse is a web-based Intelligent Tutoring System that simultaneously teaches and assesses students in logical reasoning.
  - 16 skills , 3 exercises for each skills
  - Number of students = 294 and the Number of exercises = 48
  - All students completed the 48 logical reasoning exercises => sequence length = 48

DKT + Attention Mechanism on Logic-Muse dataset

- **Logic-Muse Bayesian network (BN)**:
  - Built by experts
  - 16 observable skills , 12 latent skills



DKT + Attention Mechanism on Logic-Muse dataset

- **Data Format**:
  - Two files: a **dataset file** and a **file containing predictions made by the BN**.
  - The dataset file contains **student**, **skill** and **performance** information.
  - The dataset file is a 49-column comma-delimited file. Each row in the file indicates whether a particular student answered a specific exercise correctly or not. The first column is the student id, the rest of the columns are vectors of length 17 where each one encodes an exercise and whether the student got that exercise correctly (1) or not (0).

```
,0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47
0,"[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]","[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]","[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
1,"[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]","[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]","[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
2,"[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]","[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]","[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
3,"[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]","[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0]","[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
4,"[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]","[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1]","[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

DKT + Attention Mechanism on Logic-Muse dataset

- **Data Format**:
  - The file containing predictions made by the BN is also a 49-column comma-delimited file. Each row in the file indicates the probabilities of mastering each skills for a particular student.

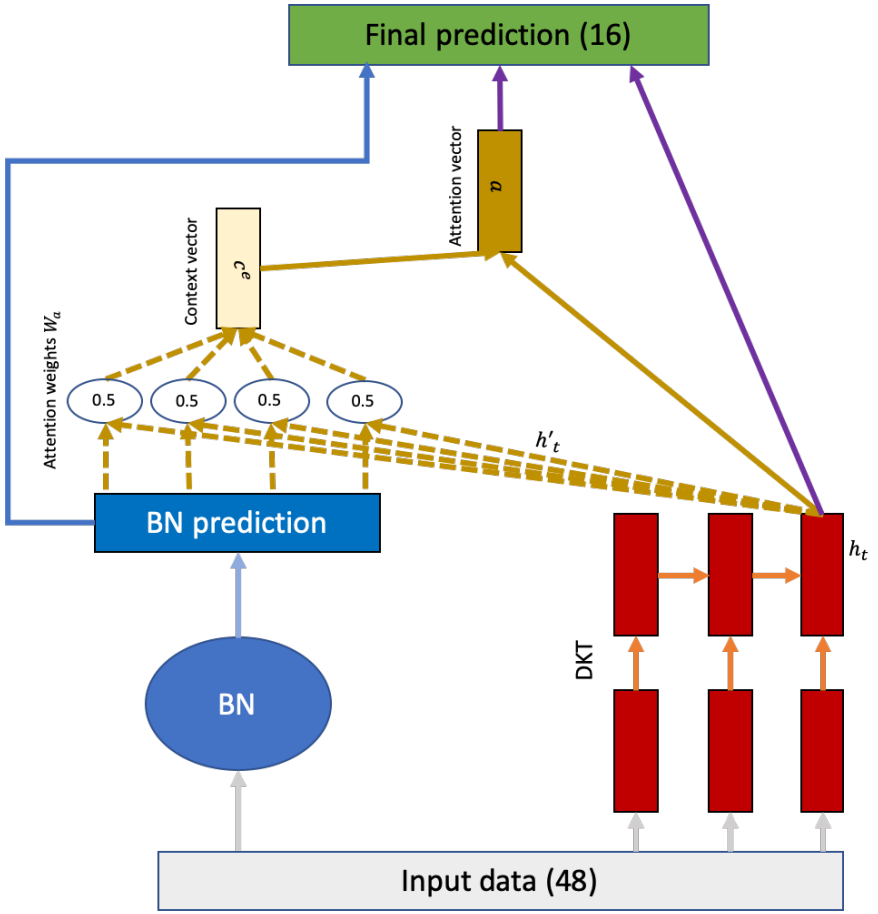
```
,0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47
0,"[0.84375, 0.7200000000000001, 0.7625, 0.7775000000000001, 0.745, 0.745, 0.745, 0.745, 0.7200000000000001, 0.7625000000000001, 0.745, 0.745, 0.69, 0.6832500000000001,
0.7274999999999999, 0.7274999999999999]","[0.6024340770791076, 0.7200000000000001, 0.7625, 0.7775000000000001, 0.745, 0.745, 0.745, 0.745, 0.7200000000000001,
```

Deep Knowledge Tracing (DKT) on Assistments dataset

- **Data preprocessing** :
  - **Input X** is size **2 * num_skills = 2 * 16**, one hot encoding
  - **Example** : "**[0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1]**" in dataset file => skill = 3, answer = 1 (correct)
    - 3 * 2 + answer = 7, X[7] = 1 => X = [0,0,0,0,0,0,0,1,0,0,0,0,…]
  - Output **Y** is size equals to num_skills **= 16.**

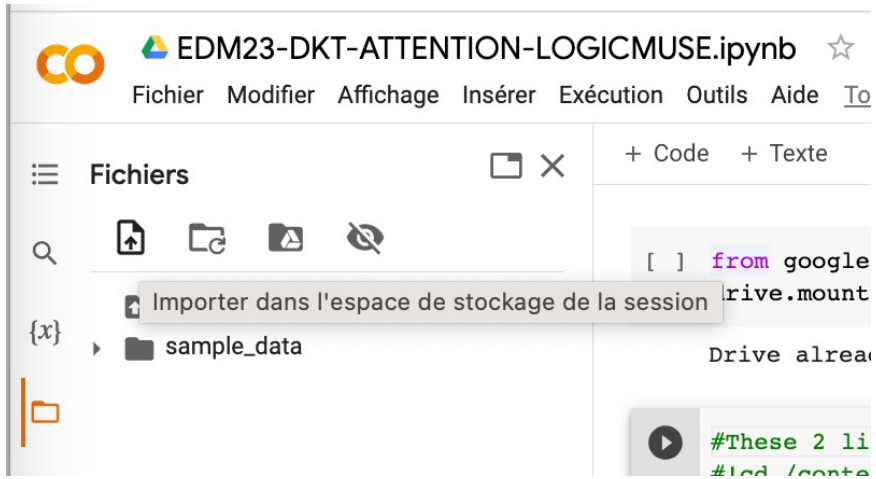
DKT + Attention Mechanism on Logic-Muse dataset

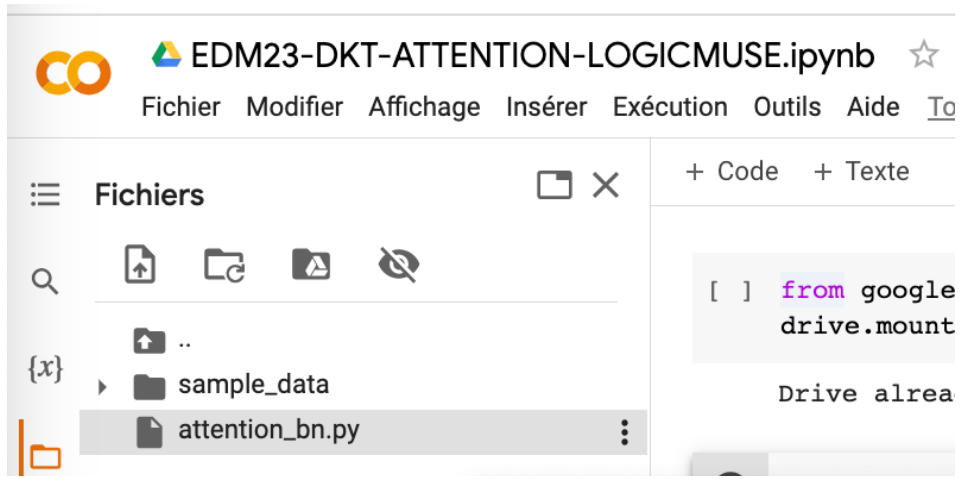
- **Model**:
  - Attention on BN
  - 16 observable skills , 12 latent skills



DKT + Attention Mechanism on Logic-Muse dataset

- **To do before running the second notebook:** import the file '**attention_bn.py**'





- **Results (– your turn)**:
  - AUC DKT = ?
  - AUC DKT + attention = 0.8784
  - AUC BN = ?