

Goals

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

- 1 Give a historical overview of the development of PDP
- 2 Start a conversation about neural networks and machine learning
- 3 Identify resources to learn about cutting-edge “machine intelligence” and “deep learning”

Outline

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

Parallel Distributed Processing (PDP)

- 1 History of PDP: Its situation in time**
- 2 Application of PDP: On Learning the Past Tenses of English Verbs**
- 3 TensorFlow Example: Learning to classify hand-drawn digits**
- 4 “Deep learning”**

Why did I choose Chapter 18?

- Worked with electrical engineers in signal processing applied to physics-based imaging systems; they were machine learning experts. Been curious ever since and this is my chance to learn.
- At first this chapter seemed like a good way for me to learn a "real-world" computational language problem. Found out it is an example that both lays foundation for and is quickly superseded by more recent work.
- As such, look at more recent work in machine learning, including deep learning, which I know even less about. Please teach me!

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

Origins of PDP

- “earliest roots of the PDP approach can be found in the work of...neurologists”
- “Feldman and Ballard (1982) laid out many of the computational principles of the PDP approach (under the name of *connectionism*), and stressed the biological implausibility of most... computational models in (AI)” ⁽¹⁾
- “Hopfield’s (1982) contribution of the idea that played a prominent role in the development of the Boltzmann machine” which re-appears in neural networks and machine learning later

¹ David E. Rumelhart and James L. McClelland. *Parallel Distributed Processing*. Massachusetts Institute of Technology, 1986. ISBN: 0262521873. DOI: 10.1016/0010-0277(93)90006-H. arXiv: 9809069v1 [arXiv:gr-qc]. URL: <http://stanford.edu/%7Bjlmcc/papers/PDP/Chapter18.pdf>, 1, p. 41.

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

The PDP renaissance

Promised to imitate human learning and physiology in solving a number of problems including:

- 1 Processing sentences
- 2 Place recognition
- 3 **Learning the past tense of English verbs (Ch. 18)**

See [²].

²David E. Rumelhart and James L. McClelland. *Parallel Distributed Processing*. Massachusetts Institute of Technology, 1986. ISBN: 0262521873. DOI: 10.1016/0010-0277(93)90006-H. arXiv: 9809069v1 [arXiv:gr-qc]. URL: <http://stanford.edu/%7Bjlmcc/papers/PDP/Chapter18.pdf>, Vol 2.

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

PDP application: Past tenses of English verbs

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

Unlike the formal grammar approach, the rules for learning past tense are *learned* as correct examples are shown to the system.

The connectionist/PDP “perspective eschews the concepts of symbols and rules in favor of a model of the mind that closely reflects the functioning of the brain.” [³]

³ Marc F. Joanisse and James L. McClelland. “Connectionist perspectives on language learning, representation and processing”. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 6.3 (2015), pp. 235–247. ISSN: 19395086. DOI: 10.1002/wcs.1340.

PDP application: Past tenses of English verbs

- Model behaves like a human in that it goes through three stages of past tense learning
 - 1 Phase I: ten most common verbs; 8 irregular, 2 regular
 - 2 Phase II: a large set of regular verbs; at this point the model gets confused and makes mistakes on the original irregular verbs it had previously learned
 - 3 Phase III: expansion to more examples including the irregular verbs again; mistakes stop for first-learned irregulars
- “Learning” means presenting correct conjugations to the model, e.g. GO → WENT, and changing network based on error
- Mathematical vectors represent phoneme structure in input/output data. More specifically, phonemes are translated to a more complex structure, “Wickelphones” which can build to form “Wickelfeatures”

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

PDP application: Past tenses of English verbs

Introduction & Outline
History of PDP
PDP application: Past tenses of English verbs
TensorFlow
Deep learning

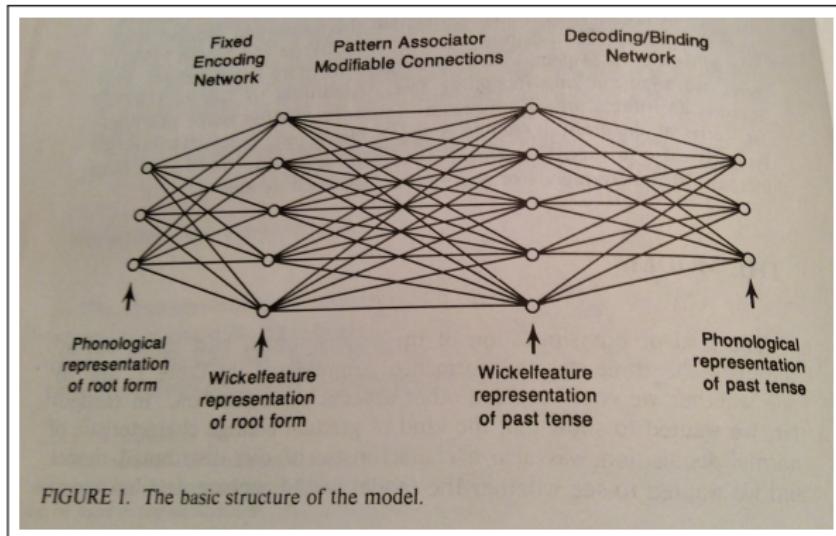


FIGURE 1. The basic structure of the model.

From⁴

⁴ David E. Rumelhart and James L. McClelland. *Parallel Distributed Processing*. Massachusetts Institute of Technology, 1986. ISBN: 0262521873. DOI: 10.1016/0010-0277(93)90006-H. arXiv: 9809069v1 [arXiv:gr-qc]. URL: <http://stanford.edu/%7B~%7Djlmcc/papers/PDP/Chapter18.pdf>.

Major weakness of this approach: Wickels

“Wickelphones” and “Wickelfeatures” that never caught on. They were improved upon just a few years later as explained in [5] and acknowledged in Joanisse and McClelland (2015) [6].

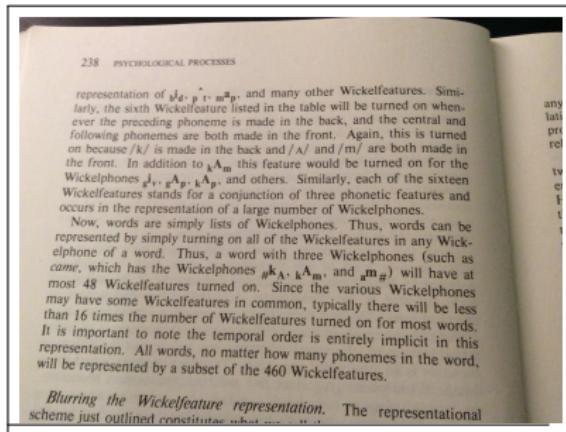
Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning



⁵ Charles X. Ling. “Learning the Past Tense of English Verbs: The Symbolic Pattern Associator vs. Connectionist Models”. In: *Journal of Artificial Intelligence* (1994), pp. 209–229. arXiv: cs/9402101.

⁶ Marc F. Joanisse and James L. McClelland. “Connectionist perspectives on language learning, representation and processing”. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 6.3 (2015), 225–247. doi: 10.1002/wics.1396

Weaknesses of the original PDP program

Criticisms centered on the

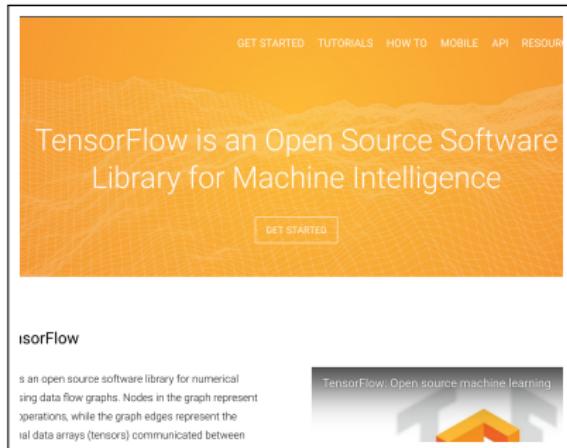
- 1 “issues of high error rates and low reliability of the experimental results”
- 2 “the inappropriateness of the training and testing procedures”
- 3 “hidden features of the representation and the network architecture that facilitate learning”
- 4 “opaque knowledge representation of the networks”

(list is made of selected quotes from [⁷])

⁷ Charles X. Ling. “Learning the Past Tense of English Verbs: The Symbolic Pattern Associator vs. Connectionist Models”. In: *Journal of Artificial Intelligence* (1994), pp. 209–229. arXiv: cs/9402101.

TensorFlow: “Library for Machine Intelligence”

Introduction & Outline
History of PDP
PDP application:
Past tenses of English verbs
TensorFlow
Deep learning



- 1** Python or C++
- 2** Optimized for deep learning model building, training, validating, and testing
- 3** Can configure automatic GPU utilization on Linux and Mac

MNIST For ML Beginners: handwritten digit classification

Hello, World! for machine learning.

Each digit is assigned to a class which is represented as an output vector with value 1 at the index corresponding to the digit, and 0 elsewhere.

So for example,

$$0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, 1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, 9 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \quad (1)$$

Introduction & Outline

History of PDP

PDP application:
Past tenses of English verbs

TensorFlow

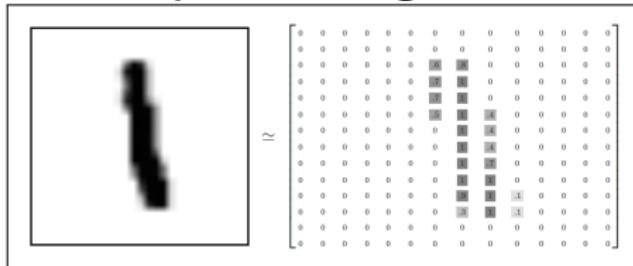
Deep learning

MNIST For ML Beginners: handwritten digit classification

TensorFlow Python library includes MNIST data:

```
from tensorflow.examples.tutorials.mnist import input_data  
mnist = input_data.read_data_sets("MNIST_data/", one_hot=True)
```

The data: pairs of images and actual digit



Corresponds to
digit 1

- 1 55,000 for training
- 2 10,000 for testing
- 3 5,000 for validation

MNIST For ML Beginners: handwritten digit classification

“The training set is used to fit the models; the validation set is used to estimate prediction error for model selection; the test set is used for assessment of the generalization error of the final chosen model. Ideally, the test set should be kept in a “vault,” and be brought out only at the end of the data analysis”⁸

- 1** 55,000 for training
- 2** 10,000 for testing
- 3** 5,000 for validation

⁸Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Vol. 1. 2009. ISBN: 9780387848570. DOI: 10.1007/b94608. URL: <http://statweb.stanford.edu/~tibs/ElemStatLearn/>, p. 222.

Training & back-propagation

The back propagation technique was one of the major contributions of the PDP research group, by Rumelhart, et al., 1986. The conceptual outline of training the neural network is below.

9

- Present the inputs, calculate outputs using the current weights (initialize weights to random values for first presentation)
- Check difference between input and output
- Adjust weights according to difference and the back propagation algorithm

⁹ David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323.6088 (1986), pp. 533–536. ISSN: 0028-0836. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0). arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).

Training & back-propagation

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

“the backpropagation equations are so rich that understanding them well requires considerable time and patience as you gradually delve deeper into the equations. The good news is that such patience is repaid many times over.”¹⁰

¹⁰ Michael A. Nielsen. *Neural Networks and Deep Learning*. 2015. URL: <http://neuralnetworksanddeeplearning.com/>.

Some confusing terminology

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

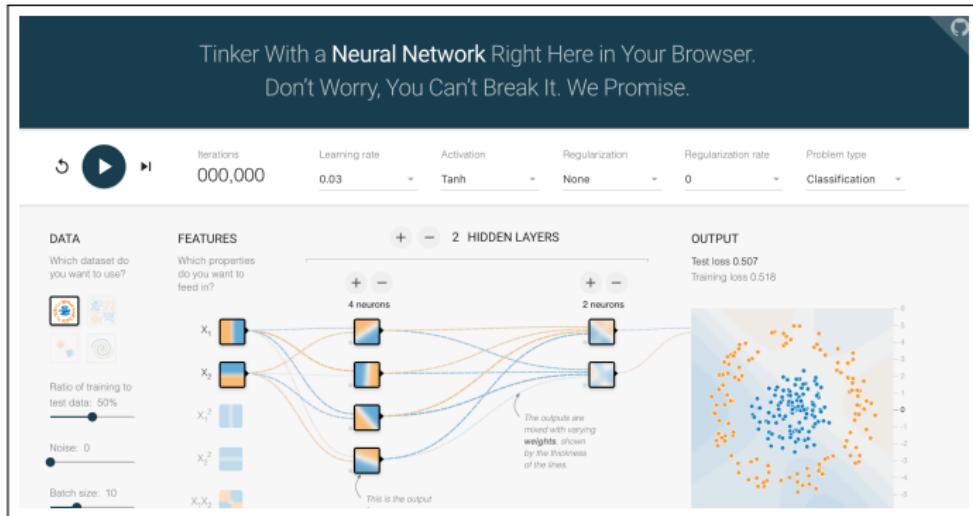
“Hidden layer”: not an input or an output layer

Perceptrons: neurons where the output activation is either 0 or 1

Sigmoid neurons: digit learning used sigmoid neurons where activation is a decimal between 0 and 1

ML beginners: “Neural network in the browser”

Introduction & Outline
History of PDP
PDP application:
Past tenses of English verbs
TensorFlow
Deep learning



playground.tensorflow.org

Deep learning

The depth of deep learning comes from the number of layers used in the network. As the number of layers is increased, there is also a hierarchical dependency added. This allows the network to represent the types of hierarchical patterning that occurs in many natural data sources¹¹.

See this comparison of popular deep learning frameworks on GitHub and read for more info on deep learning frameworks including TensorFlow.

¹¹ Michael A. Nielsen. *Neural Networks and Deep Learning*. 2015. URL: <http://neuralnetworksanddeeplearning.com/>, Tensorflow.org. *TensorFlow — an Open Source Software Library for Machine Intelligence*. URL: <https://www.tensorflow.org/> (visited on 09/18/2016).

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

Introduction
& Outline

History of
PDP

PDP
application:
Past tenses of
English verbs

TensorFlow

Deep learning

According to Wikipedia Google has been running Tensor Processing Units for over a year in their data centers. These TPUs are like GPUs, but specialized not just for matrix multiplications, but for the "tensors", the multidimensional arrays, of TensorFlow.