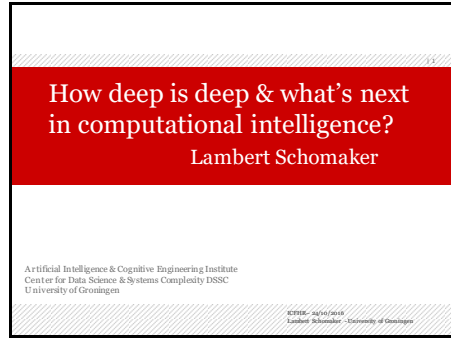


Slide 1

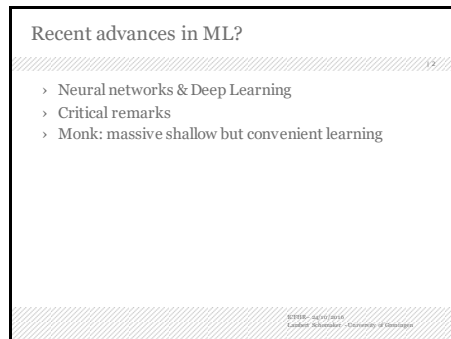


Schomaker (2016) How deep is deep and what is next in computational intelligence? [keynote lecture] International Conference on Frontiers in Handwriting, 23-26 Sept. Shenzhen, China.

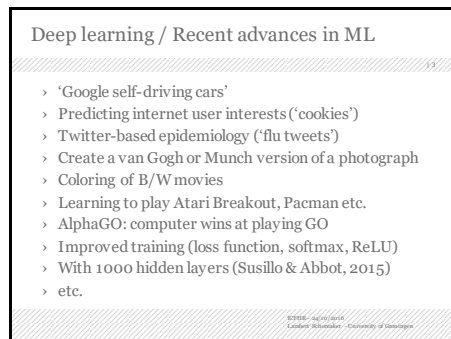
Copyright 2016 L. Schomaker

Note: not all image material may have been cleared as regards copyright. l.r.b.schomaker@rug.nl

Slide 2



Slide 3



List of current successes in deep learning

Slide 4

History of NN's

- > 1957 - 1<sup>st</sup> generation (Rosenblatt's Perceptron)
- > 1983 - 2<sup>nd</sup> generation (Werbos/Rumelhart)
- > 1996 - NN - winter
- > 2000 - 3<sup>rd</sup> generation: Deep Learning (Hinton/Lecun)
  - Computer vision
  - Speech/handwriting: sequence classification LSTM/BLSTM (Schmidhuber/Liwicki/Graves)
  - Remark: handwriting recognition played an important role. Early 2D convolutional nets by LeCun: IWFHR 1990, Cenparmi, Montreal

LECTURE 24/26/2015  
Ludovic Schreiber - Université de Sherbrooke

Brief history of NNs

Slide 5

Rosenblatt's Perceptron 1957

- Input: row of numbers (pixels)
- Output: row of numbers (code)
- wires: 'learning' is tuning their electrical resistance

No XOR mapping (Minsky & Papert, 1969)

It is remarkable that after Minsky and Papert the rejection of the perceptron was so massive. After all, linear systems with only an input and an output layer still can do a lot and also were in use. Consider for instance Widrow & Hoff, telephone line echo cancellation using a linear system. But indeed, non-linear mappings are impossible and the fact that the output units are thresholded does not introduce a non-linearity in the forward mapping itself.

Slide 6

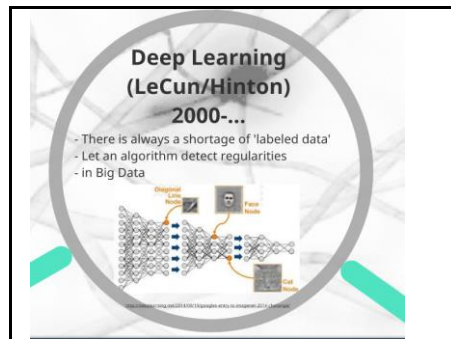
Multilayer Perceptron (1983)

- Rumelhart & McClelland generalized error backpropagation: Werbos (1974)
- Universal function approximator by summation of scaled and translated sigmoids
- Handles more complex mappings
- Training requires a lot of computing
- ... and data
- Limitations become clear (~1996)

Finally: non-linear mappings are possible. Rumelhart & McClelland came from psychology. They published two books, one blue, one brown, as a set, with a yellow third book for students. It had a diskette with C code. By 1996 there were many frustrations with MLP. There was not enough labeled data, computers were slow and generalization was not good. The SVM was developed at AT&T by Isabelle Guyon, her husband Bernard Boser and Vapnik, on the basis of the problems in training

handwriting recognizers. The bosses at AT&T were not happy with the fact that NNs yielded different solutions from different randomisations.

Slide 7



Could have been a better slide, but we all know the drill: CNNs finally have their breakthrough. It must be admitted that for a long time Yan LeCun was the only one with very good results on CNNs. The community was also surprised with the guts of Hinton to publish in Nature about what many were already doing quite extensively: using autoencoders or diabolo MLP for dimensionality reduction because there is no need for labels. But Hinton added some very useful tricks that would ultimately allow for the deep learning revolution. Both researchers were essential, in any case. Hinton, G. E. and Salakhutdinov, R. R

**Reducing the dimensionality of data with neural networks.**


Science, Vol. 313. no. 5786, pp. 504 - 507, 28 July 2006.

LeCun, Y., Bengio, Y. and Hinton, G. E. **Deep Learning.**

Nature, Vol. 521, pp 436-444.

Slide 8

The 'GO' example



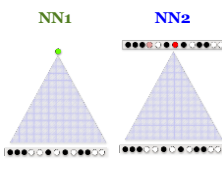
- > Board game, black/white
- > Enclose the opponent
- > GO:  $10^{170}$  possible states (chess:  $10^{47}$  states)
- > **Google/DeepMind: Very limited game knowledge,**
- > bootstrap with a limited data set of expert games.
- > NN1: Learn the **value** of any given board configuration
- > Then NN2 learns the **policy** to choose the best move per board
- > by playing 30 million times against 'itself'

- > Computer won 15 March 2016 from human world champion Lee Sedol

AlphaGO: a real system entails much more than just one single deep net

Slide 9

The 'GO' example



- > NN1: Learn the **value** of any given board pattern from human expert games
- > Then train NN2 to detect the **policy** to choose the best move given a board state

by playing 30 million times against 'itself'

But don't forget that these networks are embedded in a smart minimax search architecture with ply exploration etc. !!

The point is that Deep Learning in itself is hardly interesting. Only by integrating multiple networks into a functional architecture for the operational stage, they will be useful, as in AlphaGO. My favorite metaphor for an isolated NN in this respect is the Ferrari engine bolted to a workbench in your shed. Very powerful but utterly useless.

Slide 10

Example: finding DATE blocks in handwritten manuscripts

Zhenwei Shi (2016)




FIGURE 3.12: A schematic of the CNN used for date region detection from handwritten documents.

©2016-2017 DeepLearning.AI. Licensed under the Creative Commons Attribution 4.0 International License.

We are currently (evidently) also working on CNN and LSTM. The example here is challenging because of the low prior probability of finding a data in a sea of regions of interest that are not representing a DATE block.

Slide 11

Results for DATE detection

TABLE 4.13: Final evaluation of the proposed DateFinder system.

Methods	Recall	Precision	F <sub>0.5</sub> score
SVM-based Classifier	25.9% ± 7.0%	13.5% ± 6.6%	14.9% ± 7.0%
ConvNets Model	24.3% ± 5.2%	14.8% ± 7.1%	15.4% ± 7.5%
Date Positional Expectancy Model and SVM-based Classifier	40.6% ± 10.8%	26.7% ± 5.9%	28.5% ± 6.2%
Date Positional Expectancy Model and ConvNets Model	43.1% ± 14.0%	31.4% ± 9.7%	33.0% ± 9.8%

Please note: the 'class' concept is challenging

©2016-2017, 2018  
Ludovic Schrodiner - University of Strasbourg

Results on generic DATE block detector. Please note that the prior is very small, 30% is not so bad. Here positional density was used to weigh likelihoods (positional expectancy model)

Slide 12

**Deep?**  
**basic principle:**  
**Dimensionality reduction**

"Reconstruct large patterns via a relatively small number of values"

Look Ma, no labels!

Here: images

But as in GO, any pattern can be used,

e.g., Affymetrix RNA expr. arrays

50k dim.

With a dimensionality of 50k, it really becomes useful to do dimensionality reduction. NNs are in many ways convenient, but it is difficult to convince biomedical researchers.

Slide 13

Traditional 'deep learning': autoencoder

133

- > Study with Nitsuma: recog copyist from notation
- > How to remove staff lines?
  - By image processing? Destroys individual notes!
  - In deep feature space: by subtracting large *eigenvectors*

Residual  $r=f-\hat{f}$

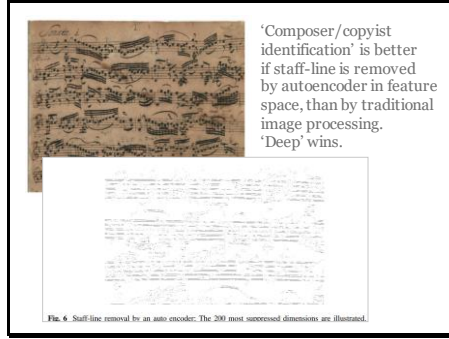
↓

Classify in r

©2016-2017, 2018  
Ludovic Schrodiner - University of Strasbourg

The goal of this study was writer identification on musical script. The staff lines do not represent relevant information. How to get rid of them? In the image or in feature space?

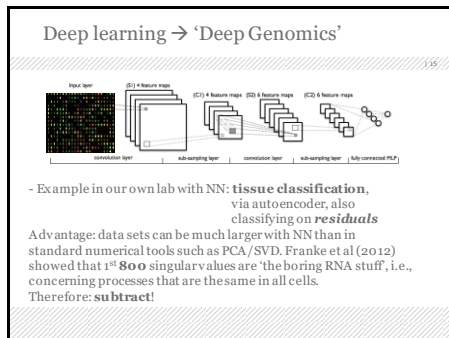
Slide 14



Niitsuma, Masahiro and Schomaker, Lambert, van Oosten, Jean-Paul, Tomita, Yo and Bell, David (2016). Musicologist-driven writer identification in early music manuscripts, *Multimedia Tools and Applications*, 75(11), pp. 6463—6479

abstract="Recent renewed interest in computational writer identification has resulted in an increased number of publications. In relation to historical musicology its application has so far been limited. One of the obstacles seems to be that the clarity of the images from the scans available for computational analysis is often not sufficient. In this paper, the use of the Hinge feature is proposed to avoid segmentation and staff-line removal for effective feature extraction from low quality scans. The use of an auto encoder in Hinge feature space is suggested as an alternative to staff-line removal by image processing, and their performance is compared. The result of the experiment shows an accuracy of 87 % for the dataset containing 84 writers' samples, and superiority of our segmentation and staff-line removal free approach. Practical analysis on Bach's autograph manuscript of the Well-Tempered Clavier II (Additional MS. 35021 in the British Library, London) is also presented and the extensive applicability of our approach is demonstrated  
<http://dx.doi.org/10.1007/s11042-015-2583-8>

Slide 15



Rudolf Fehrmann, Lude Franke et al. found this. They have several publications, among which one on SSVD (sparse svd) in *Nature*.

Slide 16

Define 'deep'!

---

- > Is it the convolutional aspect?
- > Is it the number of layers?
- > Is it the dimensionality reduction?

©2016 - 2017/2018  
Ludwig Schröder - University of Duisburg-Essen

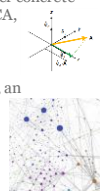
Define 'deep'

Slide 17

Two notions of 'deep'

---

- > Deep: is hidden, geometric, averaged over concrete instances, a subspace, etc., like PCA, correlation patterns, non verbal  
*Yes!*
- > Deep: is an exact fact, hidden in a graph, an unexpected explanation, precise, explainable to humans?  
*Not yet!*  
As, e.g., in causality inference ('deep cause')



©2016 - 2017/2018  
Ludwig Schröder - University of Duisburg-Essen

Two notions of 'deep'

Slide 18

Time to identify what **cannot** be done!

---

- > Deep Learning is no computational intelligence, yet
- > Intelligence by proxy: over the supervised labels
- > A smart human PhD is always necessary
- > No general intelligence: each experiment is a *one-trick pony*
- > Extensive, laboratory-based training
- > Generalisation to real, new data from new sensors, from new contexts is still difficult: k-fold evaluation is still a scam:
- > 'i.i.d.' and sampled from one cleaned pool of data yields overly optimistic performance estimates

©2016 - 2017/2018  
Ludwig Schröder - University of Duisburg-Essen

At 10 minutes

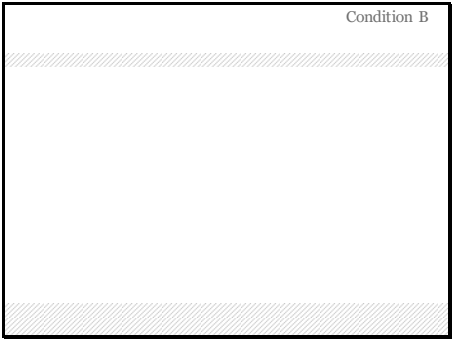




Slide 22



Slide 23



Slide 24



Slide 25

Results featural context experiment

- > Test words were handwritten (natural, shopping note)
- > same-writer flanking words (72% correct):
  - better human word recognition than different-writer flanking words (54%)
- > Conclusion: human readers exploit similarities at the letter and feature level, 'live' pertrial isogeny principle (Baird/Nagy)

ETROS - 2010/2011  
Lecture 20: Reading - University of Dundee

Sriharsha Veeramachaneni and George Nagy (2005). Style context with second-order statistics, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 14--22

**Scaling Up Whole-Book Recognition** Pingping Xiu & Henry S. Baird Computer Science & Engineering Dept Lehigh University 19 Memorial Drive West, Bethlehem, PA 18017 ...  
Published in: international conference on document analysis and recognition · 2009  
Authors: Pingping Xiu · Henry S Baird

Slide 26

Functions that are needed

- > One-shot learning → attribute classification
- > Transfer learning → reusability of skills
- > Improved unsupervised learning
- > Systems that are adaptive 24/7 (always on)
- > Active learning: - knowing what you don't know
  - identifying information that would help to disambiguate
- > → cognitive architecture as opposed to rigid pipeline processing

ETROS - 2010/2011  
Lecture 20: Reading - University of Dundee

Timing: You really should be halfway now

Slide 27

Large-scale processing of  
handwritten historical documents:  
The Monk system

Lambert Schomaker

ERC project Quirna/Popovic (2-4-0035, Jerusalem)

At 20 minutes

Slide 28

Monk e-Science web service addressing these questions:

- **What?** Word retrieval by 24/7 machine learning
  - A m bition: A European Google for handwriting
- **When?** Medieval manuscript dating
- **Where?** Geographical localization
  - Goal: uploading of charters from 1300-1550 on a server
- **Who?** Writer identification
  - On Monk server over internet
  - Using GIWIS Windows tool

©2006, 2007, 2008  
Lambert Schomaker - University of Groningen

Slide 29

Amsterdam  
Groningen  
Strafgedingnis

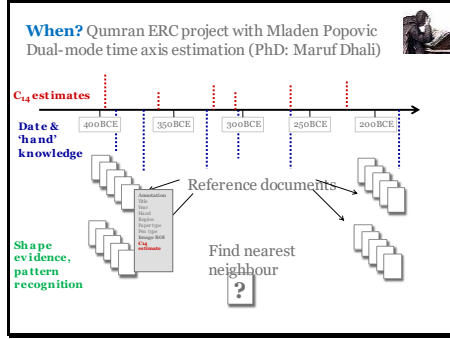
What?

芳

Colours  
Coop Jumbo

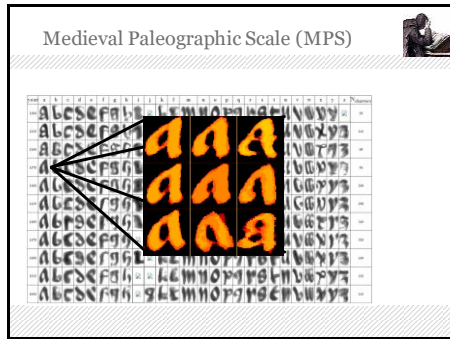
What? (is written): recognition and retrieval of text

Slide 30



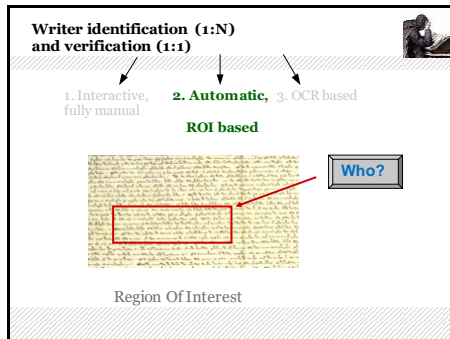
When? (has it been written)

Slide 31



In the MPS project (PhD student: Sheng He), we developed textural methods for dating of acts. However, also individual segmented characters show paleographical developments in a traditional manner.

Slide 32



Who? (wrote it)

It is much less work to just define a region of interest for writer identification (or dating)

Slide 33


Researchers & Monk 

- > Marius Bulacu, Axel Brink, Katrin Franke
- > Ralph Niels, Louis Vuurpijl, Jean-Paul van Oosten, Sheng He, Jan Burgers, Petros Samara, Olarik Surinta
- > "The Nijmegen Handwriting Group 1984-1993"
- > Netherlands Forensic Institute: Ton Broeders, Wil Fagel, Elisa van den Heuvel
- > *Isabelle Guyon, Rejean Plamondon*
- > 2016: Maruf Dhali, Mahya Ameryan, Sukalpa Chanda
- > **Users (humanities researchers):**
  - Jinna Smit, Mark Aussems, Masahiro Niitsuma, Mladen Popovic, Daniel Stoek, Jetze Touber, Grace Fong, Elaine Treharne, Dominique Stutzmann, Andreas Weber, Maxim Romanov, many others

People who has worked on Monk or had an influence on its development.

Also humanities scholars: They are high

Slide 34

Case: Optical Character Recognition 


- > "The process of segmenting a text image into individual `c``h``a``r``a``c``t``e``r` images and classifying each as being a letter in an alphabet"
- > Impressive results on contemporary printed text in machine fonts: with some linguistic postprocessing results are close to 100%
- > Solved?

©2016 - 4402/2016  
Lecture 3/Handwritten - University of Groningen

People who has worked on Monk or had an influence on its development.

Also humanities scholars: They are high

Slide 35

OCR ? 


- > 'OCR' on historic documents does not work well
- > On *handwritten* manuscripts it doesn't work, at all!
- > Problems:
  - image quality
  - unknown character shapes
  - unknown statistical language models
- > However: pattern recognition and machine learning make enormous progress these days!
- > Which methods? How to apply them?

©2016 - 4402/2016  
Lecture 3/Handwritten - University of Groningen

People who has worked on Monk or had an influence on its development.

Also humanities scholars: They are high

Slide 36

Current technology: neat text/known language 


- > Why is 'OCR', i.e., letter by letter transcription on handwriting so difficult?
- > Machine print: per character, per font, 8000 training examples are needed, typically
- > E. Barney-Smith: 200k instances of printed e vs e
- > Address reading: reduced lexicon, zip codes etc., **help**
- > In linguistic modeling: 20<sup>th</sup> century newspaper corpora do very little for 15<sup>th</sup> century acts
- > Literary text, acts and charters each need their own knowledge models in order for OCR to work


©2006-2007/2008  
Ludwig Schröder - University of Stuttgart

For example: What would a TREC corpus do for medieval administrative text?

In any case, shape recognition needs to be strong if such additional sources cannot be used.

Slide 37

Handwriting recognition: eat this! 




- Many languages, scripts
- Over historical periods
- Contractions of letters
- 'Suggested' sloppy letter shapes
- Individual writer styles
- Image problems
- Sliding window for character search usually problematic:

exit 'OCR'

Exit OCR as: Exit the methods that assume identifiable individual characters in the input stream for all letters of the intended word.

Slide 38


Monk - **Design considerations** 

1. Don't promise perfection
2. Don't promise 'transcription'
3. Don't promise exhaustive coverage (as in databases)
4. Make use of human trainers, volunteers

- > Word retrieval / word spotting:
  - "a Google for handwritten documents"

©2006-2007/2008  
Ludwig Schröder - University of Stuttgart

Slide 39




**Monk - Design considerations**

- > Word based:
  - “a Google for handwritten documents”
- > The word is a reliable chunk of information with many shape features: redundancy
 

minimum
- > Big Data: With sufficient data, there is always a reasonable response on a query

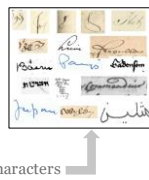
©2016 - 2017/2018  
Lodewijk Schreier - University of Groningen

Slide 40




**Monk's world model:**

- > Institutes
  - Collections
    - Books (i.e., documents)
    - Pages
      - Paragraphs
      - Lines
        - Word zones and characters
        - Pixels



©2016 - 2017/2018  
Lodewijk Schreier - University of Groningen

Slide 41



**Coarse overview of current contents Monk,**  
400+ documents, 75k page scans, 1.8M lines, 777k labels

<ul style="list-style-type: none"> <li>&gt; KdK Dutch administrative 1893-1906</li> <li>&gt; 20 books</li> <li>&gt; Dutch Admiralty 1760-1823 5 books</li> <li>&gt; Printed, Elsevirium, 1646</li> <li>&gt; Quinon 1618 (1413) 1620161</li> <li>&gt; Middelnits (example document)</li> <li>&gt; Accounts, 1425 Gefra</li> <li>&gt; Schepenbank Looisim, 1421-1539, 3 books</li> <li>&gt; Colonial diary 1939</li> <li>&gt; Municipal year report 1855</li> <li>&gt; 20k illuminated initials</li> <li>&gt; Russian handwritten newspaper 1672</li> <li>&gt; Scholarly correspondence 1674-1682</li> <li>&gt; Chansicon Boemorum 1201</li> <li>&gt; Hamiltonian Opusculorum 1150</li> <li>&gt; Resolutions States General 1627</li> <li>&gt; Medieval charters 1300-1350 in 2007 periods</li> <li>&gt; Welch 1741 1605</li> <li>&gt; Chancery Philippe le Bel 130x</li> <li>&gt; Hieratic (Egyptian)</li> <li>&gt; Hieroglyphs</li> </ul>	<ul style="list-style-type: none"> <li>&gt; 'Beowulf' and related, 4 books</li> <li>&gt; Ming Qing Chinese poetry, 300+ documents</li> <li>&gt; Arabic documents, 201 16181</li> <li>&gt; Charlotte Perkins-Stetson diaries 1883 (682p)</li> <li>&gt; Wittenbergeches (fraktur machine print)</li> <li>&gt; Etc. ...</li> </ul> <p><b>Cooperations with:</b></p> <ul style="list-style-type: none"> <li>o <i>Hugens institute, Serboene,</i></li> <li>o <i>Harvard,</i></li> <li>o <i>Stanford,</i></li> <li>o <i>Czech National Library,</i></li> <li>o <i>Dutch National Library, McGill Univ.,</i></li> <li>o <i>Utrecht Univ.,</i></li> <li>o <i>Ume Uppsala,</i></li> <li>o <i>Cy archive Louvain,</i></li> <li>o <i>Leiden Univ.,</i></li> <li>o <i>CNRS</i></li> </ul>
---	---

©2016 - 2017/2018  
Lodewijk Schreier - University of Groningen

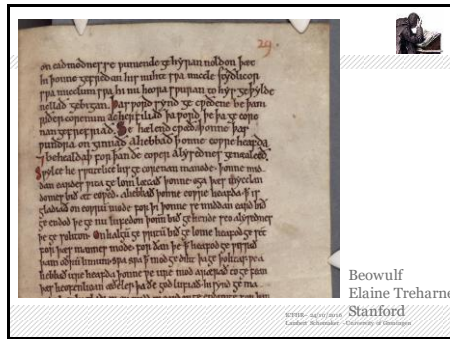
Note that almost all documents are handwritten, but some machine printed text is also in Monk. This entails difficult material such as German fraktur, Arabic and printed hieroglyphs.

Slide 42



While binarisation is usually too destructive, Otsu-based contrast enhancement works well, especially if local Otsu (or other method) is used.

Slide 43

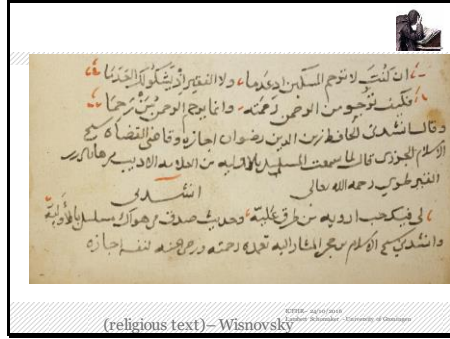


Slide 44





Slide 45



(religious text) – Wisnovsky

Slide 46

**[صاحب إفريقية ينسلخ من الإمارة ويتصوّف]**

وقها قوت أمور أبي عبدالله الشيعي بالمغرب، فصنع صاحب إفريقية صنيع محمد بن يعقوب ملك اليمن، فانسلم من الإمارة، وأظهر توبته، وليس الصوف، ورأه المظالم، وخرج إلى الروم غازياً. فقام بعده ابنه أبو العباس<sup>(١)</sup>.

وكان خروج إبراهيم بن أحمد صاحب إفريقية<sup>(٢)</sup> منها وركوبه البحر سنة تسع وثمانين، فوصل إلى صقلية، ومنها إلى طبرستان، فافتتحها، ثم حاصر كتيسة، فعرض بإسهاال، ومات في ذي القعدة. وكانت ولايته ثمانية وعشرين عاماً ونصف<sup>(٣)</sup>، ودفن بصقلية<sup>(٤)</sup>.

(١) انظر هذا الخبر في:  
تاريخ الطبري، ٩٦/١، والميون والحدائق ج ٤ ق ١٨٢/١، وجزاير الاسم ٣٢/٥، والتكامل ٥٢٢/٧، والبدية والنهاية ٩٥/١١.

(٢) هو الأمير إبراهيم بن محمد: كما في: الميون والحدائق ج ٤ ق ١٦٥/١، وهو: عبدالله بن إبراهيم بن أحمد أبو العباس، كما في: الحلة الشوار لابن الأثير ١٧٤/١ الذي وصفه بأنه كان شجاعاً بطلاً ذا بصيرة بالحروب والتدبير، عاقلاً أميناً عالماً، له نظر في العدل وصناعة بالغة والأدب...<sup>(٤)</sup> وذكره ابن عذاري بكنيته ولم يسمه فقال: أبو العباس بن إبراهيم بن أحمد: أظهر الشقاق، والحولس على الأرض، وإصناف الظلوم، وجناس أهل العلم، وشاورهم، وكان لا

Dhahabi, Tarih Islam, Tahmir/Itizi – Maxim Romanov


Slide 47

Geolocation and Date  
Scientific name of species  
Name of field scientist (=artist)

Fig. 4. Field drawing of a red-throated Barbet (*Begalaima mystacophanos*), Bu itenzorg, Java, May 1827. Each field drawing contains place names, dates, scientific names and person names. The NC collections contains ca. 2000 of such field drawings.

NWO/  
Making Sense  
project

Slide 48



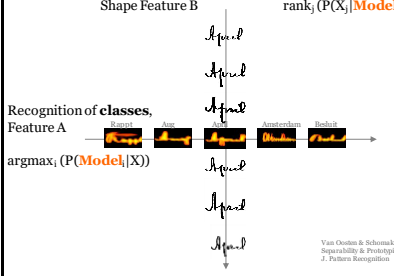
Pattern recognition and machine learning

- Several years of experimentation (online 3300+ days)
  - started in 2005
  - Monk was switched on, to a largely autonomous mode, in 2009
- For developing and optimizing two functions:
- **Retrieval**: return images for a given keyword
- **Recognition**: return the most likely word given an image

Slide 49

Retrieval of instances, Shape Feature B  $\text{rank}_j(P(X_j|\text{Model}))$

Recognition of classes, Feature A  $\text{argmax}_i(P(\text{Model}|X))$




Van Oosten & Schomaker (2013) Separability & Prototypicality ... J. Pattern Recognition

Contrary to expectation, a good classifier for recognition (in terms of recall and precision, e.g.) is not guaranteed to provide an a posteriori likelihood that is useful for intuitive ranking. For retrieval, other methods may be more applicable. The distance to a centroid is more informative in this respect than the distance to a separating boundary.

Slide 50

Boosting performance

- > Old mine-shaft elevator principle: *Fahrkunst*
- > After having trained with method A to its 'max'
- > An orthogonal method B can reach a higher performance
- > Then method A again, etc.
- > Until the real asymptote is reached



Method A - Method B

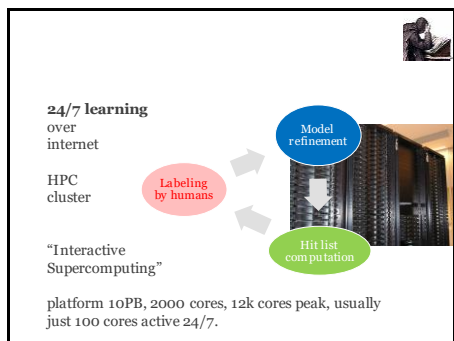
© 2014 - 2015  
Lambert Schomaker - University of Groningen

<http://www.sciencedirect.com/science/article/pii/S0031320313003786>

Jean-Paul van Oosten and Lambert Schomaker(2014). Separability versus prototypicality in handwritten word-image retrieval, Pattern Recognition, 47(3), pp. 1031 - 1038, issn = "0031-3203", abstract = "Hit lists are at the core of retrieval systems. The top ranks are important, especially if user feedback is used to train the system. Analysis of hit lists revealed counter-intuitive instances in the top ranks for good classifiers. In this study, we propose that two functions need to be optimised: (a) in order to reduce a massive set of instances to a likely subset among ten thousand or more classes, separability is required. However, the results need to be intuitive after ranking, reflecting (b) the prototypicality of instances. By optimising these requirements sequentially, the number of

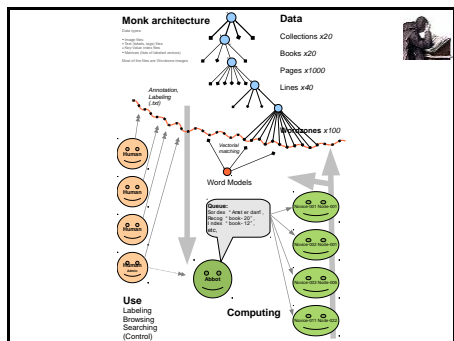
distracting images is strongly reduced, followed by nearest-centroid based instance ranking that retains an intuitive (low -edit distance) ranking. We show that in handwritten word-image retrieval, precision improvements of up to 35 percentage points can be achieved, yielding up to 100% top hit precision and 99% top-7 precision in data sets with 84 000 instances, while maintaining high recall performances. The method is conveniently implemented in a massive scale, continuously trainable retrieval engine, Monk. "

Slide 51



Users look at lines of text or at hit lists, a word model is computed, a new ranking is computed, presented to the user(s), and so on. In the early stage of the Monk development, Blue Gene (12k cores) was used for doing a grid search on optimal MLP configurations. However, centroid search proved to be much more convenient such that the current computational requirements are less demanding.

Slide 52



Labeling induces changes in the knowledge state, and leads to retraining, using a queuing of jobs in HPC.

As another metaphor, the Monk system is like a blossoming tree presenting flowers (word zone candidates) to bees (human labelers). The large surface that is needed was provided by the IBM gpfs file system.

Slide 53

Schomaker, L. (2016). Design considerations for a large-scale image-based text search engine in historical manuscript collections, Information Technology, 59, ISSN: 2196-7032

**Table 2:** File system statistics, May 2015. The maximum storage is > 1 PB. Please note the high number of inodes (read "files"): such a quantity would be considered extraordinarily high in most file-system implementations. The users sees a Posix file system which is realized through a ZFS-based metadata server using regular data infrastructure, i.e. the server

Filesystem	Storage	Used	Avail	Use%
/dev/gpfs9	2.0 G	1.3 G	775 M	63%
	104 TB	93 TB	11 TB	90%

Maximum number of files

Current number of files

Slide 54

Lessons learned during Monk development - 1

- A shape feature which is powerful for Retrieval may not be strong in Recognition!
- Requirement B: hit list should provide nice, intuitive ranking in a satisfying 'hit list'
- Requirement A: target word class should survive competition with the other word classes (needle from the hay stack)

©2016 - 2017  
Ludovic Schomaker - University of Groningen

Slide 55

Problems with products of probabilities - 2

me that all  $p_i \approx P$ , where  $P$  is a given DC

$$P_{true} = P^n$$

$$P_{measured} = (P + \epsilon)^n$$

absolute-valued error can be expressed as

$$\epsilon' = E|P^n - (P + \epsilon)^n|$$

however, is the relative error, as it will be used in a classification problem.

$$\epsilon'_{rel} = \epsilon' / P^n$$

©2016 - 2017  
Ludovic Schomaker - University of Groningen

Draft can be found on arXiv, I am still working on this.

Slide 56

Lessons learned during Monk development -3

- **Ballpark principle**
- no label: kmean, Kohonen, neural gas etc.
- one label: 1NN – first nearest neighbor
- ~5 labels: NC – nearest centroid (mean)
- >20 labels: SVM
- > 100 labels: MLP's
- But nearest centroid is by far the most satisfying in a real big data context

©2010-2011, 2013  
Ludovic Schott, University of Strasbourg

Slide 57

Character 'Shin' from Dead Sea Scrolls

0001

©2010-2011, 2013  
Ludovic Schott, University of Strasbourg

Movie of Shin converging to a stable probability landscape for ink. Normalisation on the basis of center of gravity and standard deviation of the radius, times a factor such as 2.5 to cover the ink sufficiently. At N=1878 instances:

1878

Slide 58

'Amsterdam' from 19xx administrative coll.

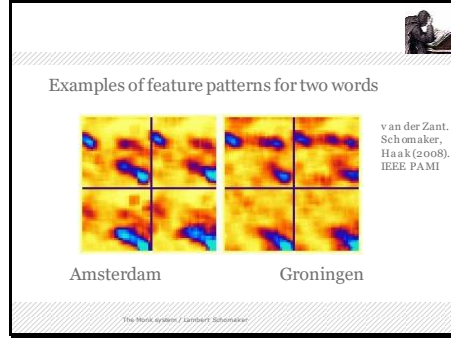
0001

©2010-2011, 2013  
Ludovic Schott, University of Strasbourg

Movie for the word Amsterdam, much less samples but still converging to a stable mean. This principle yields an attractor, also in other feature spaces, of which the inertia prevents drift in an on-line (24/7) learning setup. At N=498:

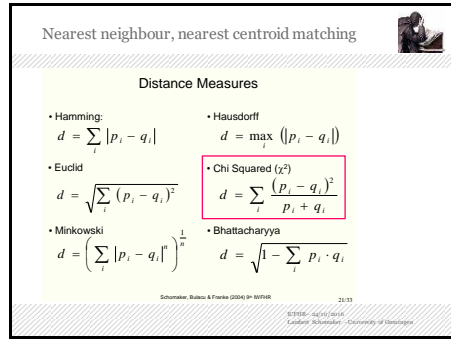
0498

Slide 59



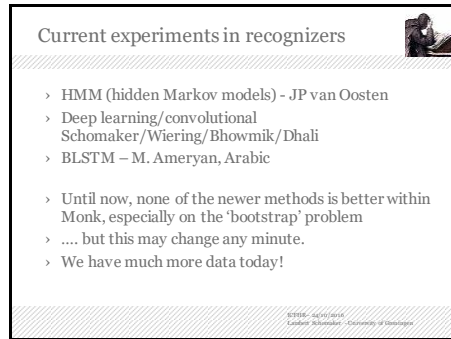
The paper in PAMI 2008 gave us the confidence that whole-word feature approaches can be very powerful. The Serre/Poggio neural network was a bit complicated and Matlab based, so I developed a more technical feature method for words in this period, that was generic enough to handle a wide range of handwritten scripts. Not only in averaged images but also in averaged feature spaces, big data or more than a hundred examples per class yield very stable models.

Slide 60

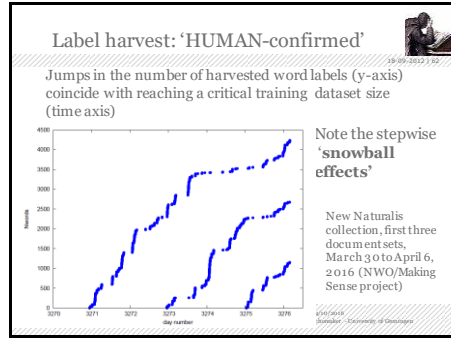


It is advisable to use Bhattacharyya or Chi-square on feature vectors that represent probability. I am sorry if that reduces the shine of your more complicated method. Up to 10% improvements in performance with respect to Euclidean or Manhattan can be observed in a wide range of pattern recognition problems.

Slide 61



Slide 62



Each trace represents the harvest curve for a book

Please note phase transitions: at some point during label harvesting, the visual word models become successful in attracting new unseen instances and provide a clean hit list that is easy to confirm by the human users:

Phase transitions in the training!

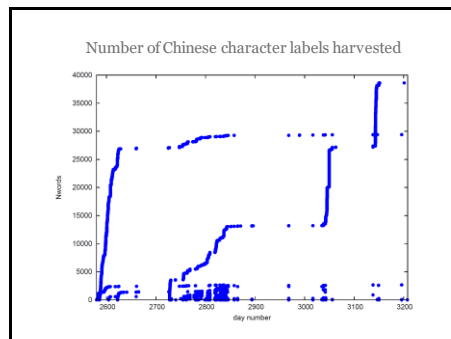
Slide 63

Learning to read Chinese woodblock printed text, from scratch

- > Columnar layout
- > Rotate scans 90 degrees
- > Many columns have table-line separators
- > Adapted the line segmentation step in Monk to remove these lines
- > Otherwise **no code changes**

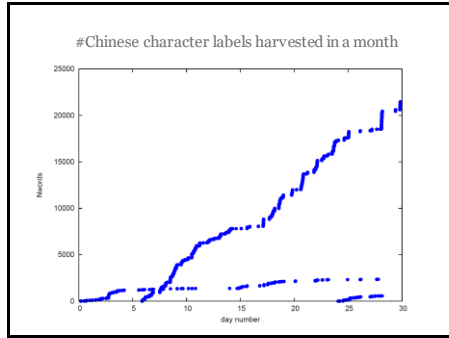
© 2016, Leiden University - University of Groningen

Slide 64

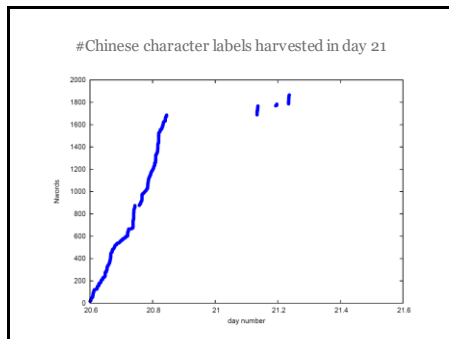


Woodblock-printed documents (higher curves) elicit more labels than handwriting, as can be expected. Still, this category is considered 'difficult' by many researchers in Chinese script recognition and by the users in the humanities.

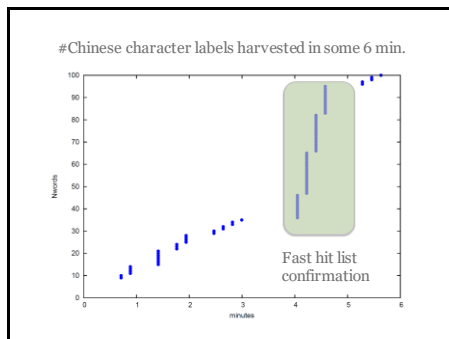
Slide 65



Slide 66



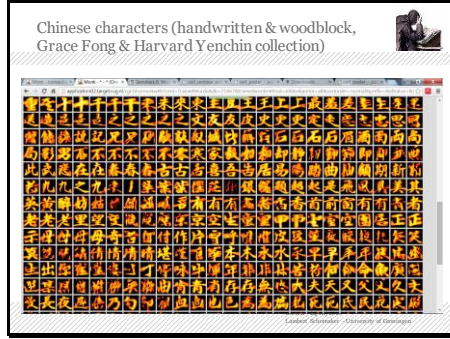
Slide 67



This slides shows a **phase transition** due to easy labeling of a correct hit list with about 60 correctly recognized and top-ranked instances.



Slide 68



Catalogue of mixed printed and written Chinese characters.

Slide 69



Baron is the target word. Words can be individually confirmed or per visible hit list as a whole.

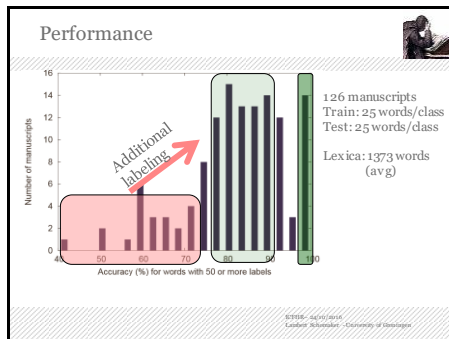
Slide 70



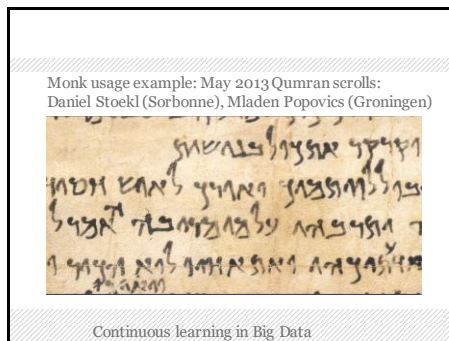
Slide 71



Slide 72



Slide 73



Slide 74

Qumran scrolls: 2400 photographs

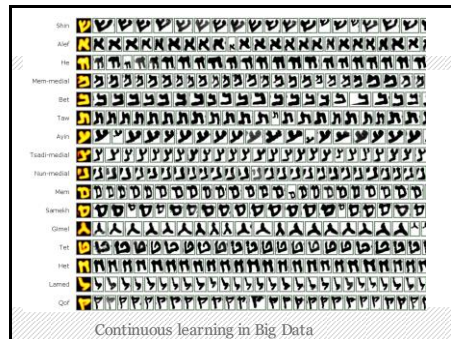
- › Using **Monk** for character labeling
- › With Daniel Stoekl and Mladen Popovics
- › Using its 24/7 machine-learning cycle:
  - Label → Train
    - Label some More → Train
    - Easily label Many
- › **Thousands** of characters 'mined' out of the Qumran collection of photographs in just **two weeks**, with very **little effort** in human labeling

ECTS/2016/04/15  
Ludwig Schröder - University of Duisburg

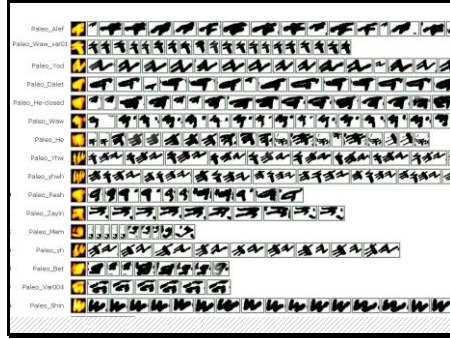
Slide 75



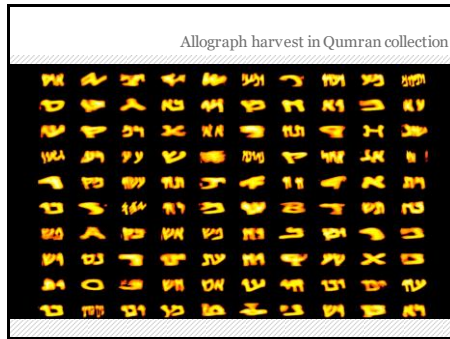
Slide 76



Slide 77



Slide 78



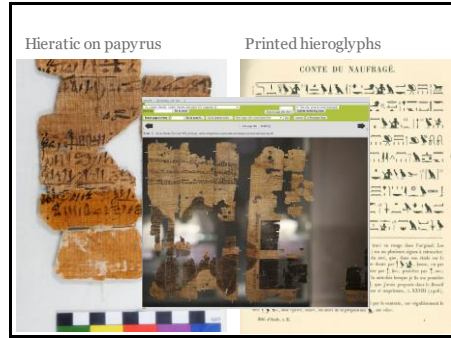
Slide 79

### Conclusions

- > Deep learning is a powerful concept
- > But it is not enough, for building autonomous and intelligent agents
- > The challenge is to design systems that handle unseen problems
- > Part of deep-learning success may be just caused by the amount of data: better comparative evaluation is needed
- > Iterative recognition and ranking works great!
- > Engineered neural-network architectures vs
- > Engineered features: we're still not there, either way!

©2016-2020 DeepLearning.AI  
Lecture 8: Introduction to Deep Learning - University of Washington

Slide 80



We consider hieratic script on papyrus the most difficult material in Monk at the moment. Hieroglyphs are easier to recognize, but for interpretation a stochastic grammar needs to be trained (2D).