Deep Learning in an industrial context: predictive maintenance, inspection and beyond

Data+Service Expert Group Predictive Maintenance, May 10, 2019

Thilo Stadelmann
1. The group
2. Predictive maintenance
3. Industrial quality control (inspection)
4. Face recognition
5. Optical music recognition
6. Lessons learned
1. ZHAW Datalab: Est. 2013

Forerunner
• **One of the first** interdisciplinary data science initiatives in Europe
• One of the first interdisciplinary centers at ZHAW

Foundation
• **People**: ca. 90 researchers from 7 institutes / 3 departments opted in
• Vision: **Nationally leading** and **internationally recognized** center of excellence
• Mission: **Generate projects** through critical mass and mutual relationships
• Competency: **Data product design** with structured and unstructured data

Success factors
• **Lean** organization and operation → geared towards projects
• Years of successful **pre-Datalab collaboration**
1. ML @ Information Engineering Group
Institute of Applied Information Technology, School of Engineering

Machine learning-based Pattern Recognition

- Robust Deep Learning
- Voice Recognition
- Document Analysis
- Learning to Learn & Control
2. Data-driven Condition Monitoring

Situation: Maintaining big (rotating) machinery is expensive, defect is more expensive

Goal: Schedule maintenance shortly before defect is expected, not merely regularly

Challenge: Develop an approach that adapts to each new machine automatically

Solution: Use machine learning approaches for anomaly detection to learn the normal state of each machine and deviations of it purely from observed sensor signals; the approach combines classic and industry-proven features with e.g. deep learning auto-encoders.

2. Data-driven Condition Monitoring: Results

Signal:

Shallow learning methods:

Deep learning methods:

➔ DL and standard methods detect the defect time; DL show less novelty where there is still no defect
3. Industrial quality control

Task

- Reliably sort out faulty balloon catheters in image-based production quality control

Challenges

- Non-natural image source, class imbalance, optical conditions, variation in defect size & shape

3. Industrial quality control – baseline results

Ingredients
- Weighted loss
- Defect cropping
- Careful customization

Interim results
3. Industrial quality control – recent results

- Human performance isn’t flawless
- Tailoring pays off
- Data shortage may be outsmarted

Figure 2: Samples of failure cases in classification. The shown defect samples in the table are not recognized as a defects, and the good images are misclassified as defects.
3. Industrial quality control – future work

Trying to overcome class imbalance and small training set sizes
3. Face matching
3. Face matching – challenges & solutions

4. Music scanning

4. Music scanning – methodology
OMR vs state of the art object detectors

YOLO/SSD-type detectors


R-CNN
- Two-step proposal and refinement scheme
- Very large amount of proposals at high resolution needed
4. Music scanning – methodology (contd.)
The deep watershed detector
4. Music scanning – methodology (contd.)
The (deep) watershed transform
4. Music scanning – methodology (contd.)

Output heads of the deep watershed detector

<table>
<thead>
<tr>
<th></th>
<th>Energy</th>
<th>Class</th>
<th>Bbox</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground truth</td>
<td><img src="image" alt="Ground truth" /></td>
<td><img src="image" alt="Class" /></td>
<td><img src="image" alt="Bbox" /></td>
</tr>
<tr>
<td>Prediction</td>
<td><img src="image" alt="Prediction" /></td>
<td><img src="image" alt="Prediction" /></td>
<td><img src="image" alt="Prediction" /></td>
</tr>
</tbody>
</table>
4. Music scanning – industrialization

Recent results on class imbalance and robustness challenges
1. Added sophisticated data augmentation in every page’s margins

2. Put additional effort (and compute) into hyperparameter tuning and longer training
3. Trained also on scanned (more real-worldish) scores

→ Improved our mAP from 16% (on purely synthetic data) to 73% on more challenging real-world data set (additionally, using Pacha et al.’s evaluation method as a 2nd benchmark: from 24.8% to 47.5%)

5. Lessons learned

Data is key.
- Many real-world projects miss the required quantity & quality of data → even though «big data» is not needed
- Class imbalance needs careful dealing → special loss, resampling (also in unorthodox ways), exploitation of every possible learning signal
- Unsupervised methods need to be used creatively
- Users & label providers need to be trained

Robustness is important.
- Training processes can be tricky → give hints via a unique loss, proper preprocessing and pretraining
5. Lessons learned – model interpretability

Interpretability is required.
- Helps the developer in «debugging», needed by the user to trust
  → visualizations of learned features, training process, learning curves etc. should be «always on»

Schwartz-Ziv & Tishby (2017). «Opening the Black Box of Deep Neural Networks via Information».
Conclusions

• Important for DL in practice, and hence target of applied research: sample efficiency, robustness, interpretability

• Future work will include:
  - Unsupervised and semi-supervised learning approaches
  - Novel object detection approaches for many tiny objects
  - Work on explainable DL

On me:
• Prof. AI/ML, scientific director ZHAW digital, board Data+Service
• thilo.stadelmann@zhaw.ch
• 058 934 72 08
• https://stdm.github.io/

Further contacts:
• Data+Service Alliance: www.data-service-alliance.ch
• Collaboration: datalab@zhaw.ch

➡️ Happy to answer questions & requests.
APPENDIX
6. Speaker clustering

For the 630 training utterances, GMMs with 32 mixtures are built a priori, then an identification experiment is run for the 630 test utterances. It yields a satisfactory 0.5% closed set identification error.

Evaluations typically concentrate on data sets built from broadcast news/shows and meeting recordings, where diarization error rates ranging from 8% to 24% are reported [28][34][45]. These results are confirmed by more recent [34].

The interpretation of our results has shown that it is the stage of modeling that bears the highest potential: the inclusion of temporal context information among feature vectors is what is crucially missing there. Furthermore, the inclusion of a context vector. This corresponds to a syllable length of 130 ms and is found to best capture speaker specific sounds in informal listening experiments over a range of 32–496 ms (in intervals of 16 ms). Our context vector step is one orig-
6. Speaker clustering – exploiting time information

6. Speaker clustering – learnings & future work

«Pure» voice modeling seems largely solved
• RNN embeddings work well (see t-SNE plot of single segments)
• RNN model robustly exhibits the predicted «sweet spot» for the used time information
• Speaker clustering on clean & reasonably long input works an order of magnitude better (as predicted)
• Additionally, using a smarter clustering algorithm on top of embeddings makes clustering on TIMIT as good as identification (see ICPR’18 paper on dominant sets)

Future work
• Make models robust on real-worldish data (noise and more speakers/segments)
• Exploit findings for robust reliable speaker diarization
• Learn embeddings and the clustering algorithm end to end

7. Learning to cluster
7. Learning to cluster – architecture & examples

7. Learning to cluster – loss

**Probability of** two instances $i, j$ being **in the same cluster** $\ell$ (of $k$ clusters):

$$P_{ij}(k) = \sum_{\ell=1}^{k} P(\ell \mid x_i, k) P(\ell \mid x_j, k).$$

**Probability of two instances** $i, j$ **being in the same cluster** $\ell$ **in general**:

$$P_{ij} = \sum_{k=1}^{k_{\text{max}}} P(k) \sum_{\ell=1}^{k} P(\ell \mid x_i, k) P(\ell \mid x_j, k).$$

**Cluster assignment loss** (with $y_{ij} = 1$ *iff* the two instances are from the same cluster, 0 otherwise):

*Weighted binary cross entropy* (weights account for imbalance due to more dissimilar pairs)

$$L_{\text{ca}} = \frac{-2}{n(n-1)} \sum_{i<j} \left( \varphi_1 y_{ij} \log(P_{ij}) + \varphi_2 (1 - y_{ij}) \log(1 - P_{ij}) \right)$$

**Number of cluster loss**:

*Categorical cross entropy*

$$L_{\text{cc}} = -\log(P(k))$$

**Total loss**:

$$L_{\text{tot}} = L_{\text{cc}} + \lambda L_{\text{ca}}$$
The Swiss Alliance for Data-Intensive Services provides a significant contribution to make Switzerland an internationally recognized hub for data-driven value creation.

In doing so, we rely on cooperation in an interdisciplinary expert network of innovative companies and universities to combine knowledge from different fields into marketable products and services.
What is Data Science?

Enables Data Products ➔ Applied Science ➔ Interdisciplinary

Data Science := “Unique blend of skills from analytics, engineering & communication aiming at generating value from the data itself [...]”

(ZHAW Datalab)
**Overview**

**Partners**

ARGUS der Presse AG
- Switzerland's leading media monitoring and information provider
- Experience of more than 100 years

ZHAW Datalab
- Interdisciplinary research group at Zurich University of Applied Sciences
- Combining the knowledge of different fields related to machine learning

**The Project**

**Who are we**

ARGUS der Presse AG
- Switzerland’s leading media monitoring and information provider
- Experience of more than 100 years

ZHAW Datalab
- Interdisciplinary research group at Zurich University of Applied Sciences
- Combining the knowledge of different fields related to machine learning

**What do we do**

Goal
- Real Time Print Media Monitoring
- Extraction of relevant articles from newspaper pages
- Delivering articles to customers

Problem
- Fully automated article segmentation
- Identification of article elements (e.g. title, subtitle, etc.)

**ARGUS der Presse AG**
- Switzerland’s leading media monitoring and information provider
- Experience of more than 100 years

**ZHAW Datalab**
- Interdisciplinary research group at Zurich University of Applied Sciences
- Combining the knowledge of different fields related to machine learning

**Most Successful Approach [3]**

**References**

