Developing Multilingual OCR and Handwriting Recognition at Google

Observations and Reflections

Ashok Popat
Research Scientist, Google Inc.

IAPR Summer School, Jaipur: Jan 23 2017
Joint work with

Jon Baccash
Marcos Calvo
Victor Cărbune
Thomas Deselaers
Karel Driesen
Sandro Feuz
Yasuhisa Fujii
Philippe Gervais
Pedro Gonnet
Patrick Hurst
Henry Rowley
Li-Lun Wang
Optical Character Recognition
OCR in Google Products

- Google books
- Google Keep - notes and lists
- Google Drive
- Google Translate
Google Handwriting Input

on-device recognition
> 80 languages + emoji
Google Translate for Android

Translate 2.3

enabled by default only for CJ

Translate 2.4+

enabled for all supported lang.
Handwrite for Mobile Search

write your search right on the Google homepage

available on Google.com from smartphone or tablet

can be activated or disabled in mobile search settings
Other Applications

Other input methods for Android

Input tools

... and more
Outline

- Multilingual OCR and On-line handwriting systems
- Research at Google
- Personal observations, reflections
Part 1a: A multilingual OCR system
Examples from Google Books

Multiple scripts / languages on a page:
Examples from Google Books (cont.)

Per-word script and language variation:

Dieser Familie nähern sich die Slawonischen Worte лана, лопасть und vorzüglich лопата (lapa, lopast und lopata, Pfote, Flügel einer Mütze, Schaufel), mit welcher ein Blatt durch seine Fläche eine grosse Aehnlichkeit hat. Daher nennen auch wir die grossen Blätter einiger Gewächse лапушник (lapuschnik, Klette). Wie sehr sich aber auch die letztern Worte
§ 6. Rezeichen:

1. virâma, „Ruhezeichen“ ein unten rechts an den Konsonanten angesetzter Strich, nimmt dem Konsonanten den a-Laut, mit welchem er nach § 1 und § 4 an und für sich zu sprechen ist. Der Virâma findet sich namentlich am Ende eines Wortes und manchmal anstatt einer Ligatur; z. B. यत् yat, आसीत् āsīt, अगमद् agamad देवी devī;

2. avagraha, „Apostroph“ steht am Anfang eines Wortes zur Bezeichnung eines vorn abgesunkenen a; z. B. दुःखितो duḥkhitō 'bhavat.
Some of the 26 scripts of interest
Starting point: Markov-model-based approaches

- Document image decoding [Kopec and Chou, 1994]
  - Explicit model of typesetting process: seek to invert
  - Influenced by speech recognition methods
  - Extremely high accuracy when models match the data

- BBN Byblos system [Schwartz et al., 1996]
  - Treat text line like a speech waveform
  - Built on existing speech recognition system
  - First successful Arabic OCR
Generalization of the noisy channel model

- Speech approach

\[ m' = \arg \max_m P(m)^\alpha P(x|m) \]
\[ = \arg \max_m \alpha \log P(m) + \log P(x|m) \]

- Generalize to multiple feature functions

\[ m' = \arg \max_m \sum_i \lambda_i h_i(x, m) \]

- Learn \( \{\lambda\} \) via minimum error-rate training [Macherey et al. ‘08, Och ‘03]
Minimum Error Rate Training

Training flow

Text data

Rendering w/ degradation

Optical model training

HMM

Labeled data

MERT

OCR system

Confidence filtering

Evaluation

Packaging

Language model training

LM

Decode

Self-labeled data

Unsupervised data

Text data
Technical evolution

- **Optical model**
  - GMM -> DNN
  - DNN -> LSTM
  - Sequential discriminative training of DNN/LSTM

- **Language model**
  - N-gram -> RNN-LM

- **Decoding**
  - Pruning algorithms designed for OCR
  - Automatic decoding parameter optimization
  - Fujii et al., ICDAR’15
Script ID (Li et al., 2015)
Regions not covered
Part 1b: A multilingual handwriting recognition system
Segment and Decode

Hidden Markov Models

neural network variants:
Recurrent, Time-Delay,
Long Short-term Memory
[Jaeger 2001], [Graves 2009], ...

Apple Newton [Yaeger 1996]
Microsoft Tablet PC / Vista [Pittman 2007]
Segment and Decode 1: Creating a segmentation lattice
Segment and Decode 2: recognizing character hypotheses

0
1:1.2 e:3.8 ⊗:3.5
1
s:0.2 n:4.2 ⊗:6.2
2
:1.3 ,:1.4 ⊗:1.2
0,1
a:7.1 u:8.2 ⊗:0.9
0,2
i:0.1 é:4.7 ⊗:5.3
1,2
j:4.8 ź:6.5 ⊗:0.4
0,1,2
m:3.7 ú:9.1 ⊗:0.2

Google
Segment and Decode 3: Decoding
Feature Function Weights

Label "i"
Feature functions values:
0.1 – character score
0.9 – language model score
2.3 – relative size to neighbors
0.2 – cut score

Label "é"

[...]

determine edge score as weighted sum

\[ c = \sum_{f} \gamma f C_f \]
Features: Per character hypothesis

- Histograms of point features (3210 dimensional)
- Bitmap features: 3x8x8 pixels (192 dimensional)
- Simple statistics (384 dimensional)
- Water reservoir features (64 dimensional)
- Stroke direction (180 dimensional)
- Quantized stroke direction maps (512 dimensional)
More feature functions

- string length
- character prior
- segmenter cut features
- relative size
Part 2: Research at Google
Google’s Hybrid Approach to Research

Spector, Norvig, Petrov ‘12 Comm. of the ACM

- Pattern 2: Small research team builds a system that gets deployed.

“This pattern applies best when continuing research can further improve and extend the resulting products.”
Enablers

- Single code base, wide range of library functions
- Infrastructure
- Expertise and skills of other teams
- Data
Enablers (cultural)

- Transparency and cooperation
- Peer review
- Respect and psychological safety
- Team- and personal-level pace and execution
- Data-centrism
Software engineering

- Respected and valued
- If it’s not checked in, it doesn’t exist
- Toy prototypes versus production-quality code
- A day in the life: 80/20
Part 3: Observations and Reflections
Translation quality: Franz Och et al., NIST’06

Improvement over time (May - July): Arabic
Rapid real progress

- Multiple contributors, one system
- Industry folks at NIST’06 meeting were startled
- Incentive: get a real gain, check it in quickly
- From each according to ability
- Data is important; eval data is paramount
Keeping it real

- Working, deployed system that solves a whole problem
- Tight feedback loop
- Everything that matters gets measured
Pedestrian approaches versus cutting edge

- Translate: world-beating and obsolete
- Data versus Syntax
- Language modeling: “Stupid Backoff” (Brants et al., 2007)
- When and how to invest in promising researchy approaches?
Reward and recognition

- Cleverness, independence, origination of new ideas?
- Cooperation, generosity, communication, productivity, risk taking?
- Imposter syndrome
- Happiness
Summary: what’s worked for me?

- Work on real systems
- Measure what matters
- Incent the right things
- Keep aware of new research while investing conservatively
Then and now
Thank you!