A New Smoothing Method for Lexicon-based Handwritten Text Keyword Spotting

Joan Puigcerver, Alejandro H. Toselli, Enrique Vidal
{joapuipe,ahector,evidal}@prhlt.upv.es

Pattern Recognition and Human Language Technology Research Center
Departament de Sistemes Informàtics i Computació
Universitat Politècnica de València

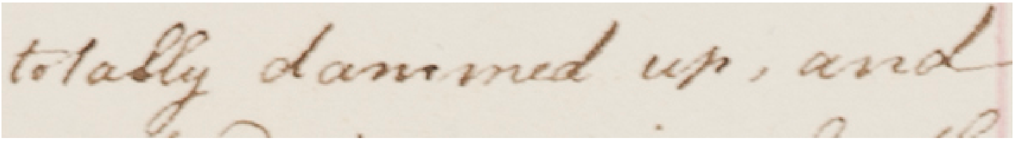
June 17, 2015

Presentation Outline

Introduction

- Keyword spotting (KWS) aims to determine whether a given query is present in a given image or image region.

- We focus on line-level Query-by-String: i.e. the query is just a text keyword and we must retrieve segmented lines containing it (not the exact location within the line).



Is “damned” written in this line?

- Different methods have been used in the past:
  - Lexicon-free: HMM-Filler [1], BLSTM [2]
  - Lexicon-based: Word-Graph (WG) [5]

Introduction

- Lexicon-based (WG) approach provides much faster lookup-time than the lexicon-free alternatives (orders of magnitude).
- Typically, it also works better than HMM-Filler and similarly to BLSTM.
- However, performance is greatly affected by out-of-vocabulary words (i.e. words that were not in the LM used to produce the WG). They are unable to *find* those on images.
- **Goal**: Benefit from lexicon information and mitigate the problem of OOV queries.
- **Smoothing**: Use information of similar in-vocabulary keywords to approximate information about OOV.

Keyword Spotting Framework

- We try to model the probability of a word v being written in image line $\mathbf{x}$: $P(R \mid \mathbf{x}, v)$

- Here, R is a binary random variable modeling the event "word v is written in line $\mathbf{x}$".

- To approximate $P(R \mid \mathbf{x}, v)$, we use the *frame-level word posterior*, $P(v \mid \mathbf{x}, i)$: the probability that horizontal position (frame) i is part of word v in line $\mathbf{x}$.

$$P(R \mid \mathbf{x}, v) \approx \begin{cases} \max_{1 \leq i \leq n} P(v \mid \mathbf{x}, i) & R = 1 \\ 1 - \max_{1 \leq i \leq n} P(v \mid \mathbf{x}, i) & R = 0 \end{cases} \quad (1)$$

- $P(v \mid \mathbf{x}, i)$ can be very efficiently computed using the WG of $\mathbf{x}$, [5].

Keyword Spotting Framework: Example

$$P(R = 1 \mid \mathbf{x}, v = \text{“lindo”}) = 0.45$$

$$P(R = 1 \mid \mathbf{x}, v = \text{“volteo”}) = 0.6$$

$$P(R = 1 \mid \mathbf{x}, v = \text{“rojo”}) = 0.8$$

Out-of-Vocabulary Queries

- What if v was not part of the vocabulary V used to build the WG of $\mathbf{x}$?
- Then, $P(v \mid \mathbf{x}, i) = 0, 1 \leq i \leq n$. Which implies: $P(R = 1 \mid \mathbf{x}, v) = 0$.
- Previous approaches to overcome the OOV problem in KWS:
  - Line Max [4]: $P(R = 1 \mid \mathbf{x}, v) \approx \max_{v' \in V} P(R \mid \mathbf{x}, v') \cdot \exp(-\alpha\ d(v, v'))$
  - Posteriorgram [4]: Smooth $P(v \mid \mathbf{x}, i)$ instead of $P(R \mid \mathbf{x}, v)$.
  - WG + HMM-Filler [3]: Not a smoothing technique. Use a lexicon-based (WG) system when $v \in V$, and a lexicon-free (HMM-Filler) when $v \notin V$.

New Proposed Line-Level Smoothing

- Main idea: Marginalize $P(R \mid \mathbf{x}, u)$ for a given $u \notin V$ among all words $v \in V$.

$$P(R \mid \mathbf{x}, u) = \sum_{v \in V} P(R, v \mid \mathbf{x}, u) = \sum_{v \in V} P(R \mid \mathbf{x}, u, v) \cdot P(v \mid \mathbf{x}, u) \tag{2}$$

- Two assumptions: $P(v \mid \mathbf{x}, u) \approx P(v \mid u)$ and $P(R \mid \mathbf{x}, u, v) \approx P(R \mid \mathbf{x}, v)$.

$$P(R \mid \mathbf{x}, u) \approx \sum_{v \in V} P(R \mid \mathbf{x}, v) \cdot P(v \mid u) \tag{3}$$

- $P(v \mid u)$ is a distribution over a finite set V. We define it from the Levenshtein distance between two strings, $d(u, v)$.

$$P(v \mid u) \stackrel{\text{def}}{=} \frac{\exp(-\alpha \; d(u, v))}{\sum_{v' \in V} \exp(-\alpha \; d(u, v'))} \tag{4}$$

Experiments

- Two datasets were used:
  - Cristo-Salvador (CS), XIXc. Spanish manuscript, single writer, 2K-word lexicon, 497 test lines, 1671 queries (1051 OOV, 31% relevant events).
  - IAM database (IAM), modern English, multiple writers, 20K-word lexicon, 929 test lines, 2209 queries (437 OOV, 14% relevant events).

- Different metrics were measured:
  - Accuracy: Average Precision (AP) and Mean Average Precision (mAP).
  - Speed: seconds required to serve an OOV query.

- Same experimental setup than the publications we were comparing to: HMM-GMM based system, bi-gram LM, WG max-input-degree (40), same data partitions and validation procedure to adjust hyper-parameters.

Experiments: Results

Method	CS			IAM		
	AP	mAP	Qtime	AP	mAP	Qtime
No smoothing	55.6	29.0	—	69.1	68.8	—
Line Max [4]	57.8	45.0	0.44	69.8	76.0	8.78
Posteriorgram [4]	58.8	46.7	27.21	70.2	76.1	42.48
WG + HMM-Filler [3]	72.5	76.6	177.10	76.9	82.2	58.16
This work	59.5	46.0	0.52	71.3	76.0	9.96

- New proposal improves AP of previous smoothing methods, mAP comparable to the highest of them. Worse than *WG + HMM-Filler*, but orders of magnitude faster.

- Smoothing methods are affected by different factors: vocabulary size, num. of lines, LM/WG perplexity. Thus, the relative differences in Qtime.

Conclusions

- New smoothing method mitigates the problem of OOV queries in lexicon-based KWS.
- A detailed comparison between the new proposal and previous published alternatives was conducted.
- Our proposal gives higher AP and similar mAP than previous smoothing alternatives, and it's fast.
- Not as good as combining a lexicon-based and a lexicon-free systems, but orders of magnitude faster.
- Offers a trade-off between speed and accuracy to the user of the KWS system.

References

[1] A. Fischer, A. Keller, V. Frinken, and H. Bunke. Lexicon-free handwritten word spotting using character HMMs. *Pattern Recognition Letters*, 33(7):934 – 942, 2012. Special Issue on Awards from ICPR 2010.

[2] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke. A Novel Word Spotting Method Based on Recurrent Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2):211 –224, Feb. 2012.

[3] J. Puigcerver, A. H. Toselli, and E. Vidal. Word-Graph and Character-Lattice Combination for KWS in Handwritten Documents. In *14th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 181–186, 2014.

[4] J. Puigcerver, A. H. Toselli, and E. Vidal. Word-Graph-based Handwriting Keyword Spotting of Out-of-Vocabulary Queries. In *22nd International Conference on Pattern Recognition (ICPR)*, pages 2035–2040, 2014.

[5] A. H. Toselli, E. Vidal, V. Romero, and V. Frinken. Word-Graph Based Keyword Spotting and Indexing of Handwritten Document Images. Technical report, Universitat Politcnica de Valncia, 2013.