

Improving Back-off Models with Bag of Words and Hollow-grams

Benjamin Lecouteux, Raphaël Rubino, Georges Linarès

▶ To cite this version:

Benjamin Lecouteux, Raphaël Rubino, Georges Linarès. Improving Back-off Models with Bag of Words and Hollow-grams. Interspeech, 2010, Makuhari, Japan. hal-02088832

HAL Id: hal-02088832 https://hal.science/hal-02088832

Submitted on 3 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





Introduction

- We propose a simple and efficient model based on word co-occurrences and a new hollow-gram model.
- Our approaches are applied to traditional modified Kneser-Ney back-off language models.
- We decide to take into account the short context around *n*-grams to tackle the issue of **sparse training data**.
- A slight improvement on word error rate (WER) is reached, with and without acoustic adaptation.

Experimental Framework

- Experiments are carried out by using the **LIA broadcast news system** [?], which relies on the HMM-decoder **LIA SPEERAL**. []
- A classical 3-gram model built on *Le Monde* French Newspaper and Gigaword corpus (1.3G words).
- Training and development parts of the data set are based on the **ESTER-2** [?] evaluation campagn corpus (100h).

Hollow-gram model

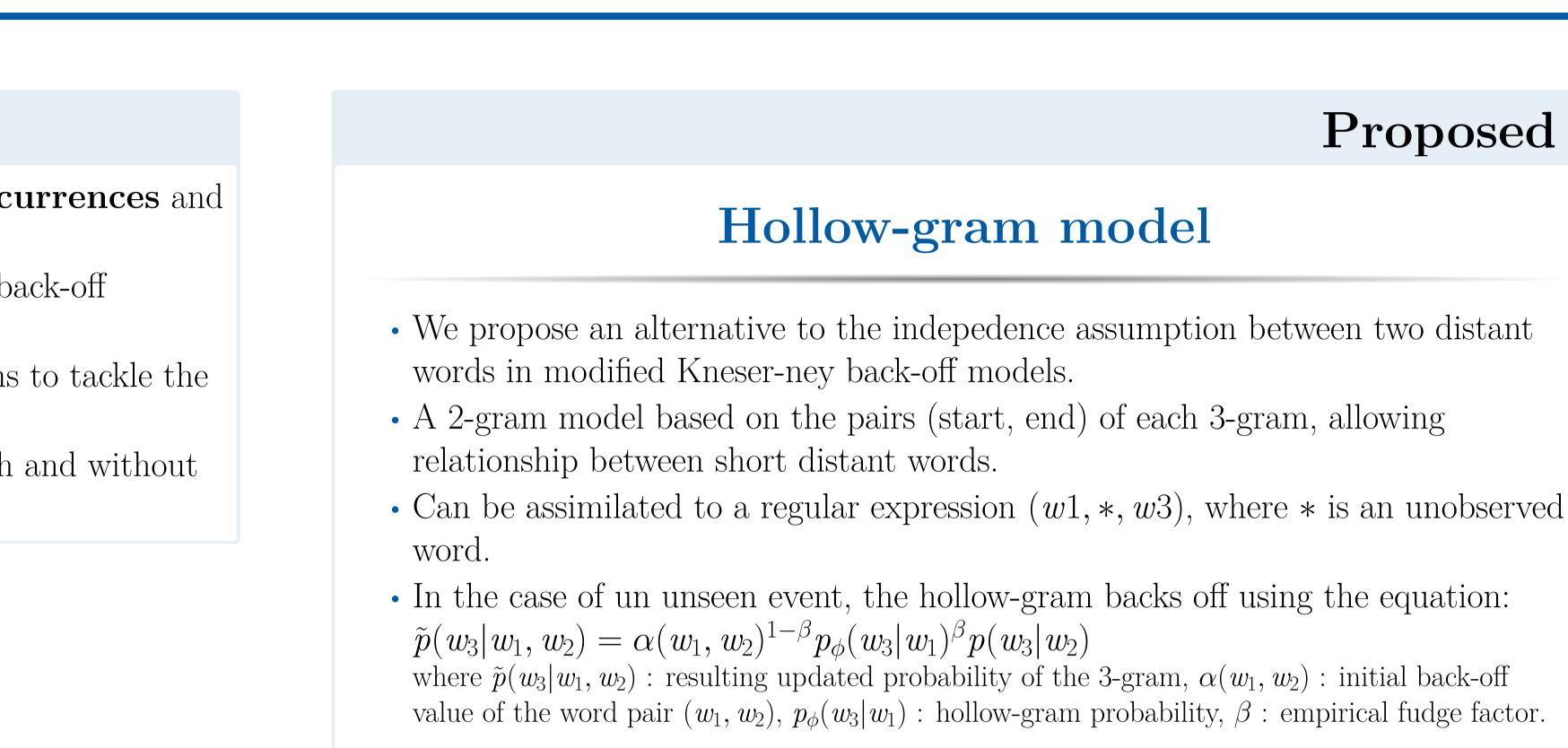
Shows	WER	SER	CWR
Inter (4h)	32.8 (-0.3)	74.5 (-0.5)	70.5(+0.8)
TVME $(1h)$	31.3(0.0)	67.0 (-0.3)	71.7 (+0.5)
RFI (1h)	18.5(-0.2)	65.7 (-0.9)	84.4 (+0.4)
GLOBAL	-0.2	-0.5	+0.7

- Table: WER, SER, CWR using the hollow-gram based backoff mode.

- The **hollow-gram model** can be extended to *n*-grams.

Improving Back-off Models with Bag of Words and Hollow-grams

Benjamin Lecouteux, Raphaël Rubino, Georges Linarès Laboratoire Informatique d'Avignon (LIA), University of Avignon, France {benjamin.lecouteux, raphael.rubino, georges.linares}@univ-avignon.fr



	WER	SER	CWR
Inter (4h)	33.1	75.0	69.7
TVME (1h)	31.3	67.6	71.2
RFI (1h)	18.7	66.6	84

Resul	tc
Resul	US

Word co-occurence model

Shows	WER	SER	CWR	Shows	WER	SER	CWR	Shows	WER	SER	CWR
				Inter (4h)	32.4 (-0.7)	74.6 (4)	70.8(+1.1)	Inter (4h)	32.7(-0.3)	74.8(-0.2)	70.3(+0.6)
				TVME (1h)	30.9(-0.4)	67.2(4)	72.0(+0.8)	TVME (1h)	31.1(-0.2)	67.5(-0.1)	71.6(+0.3)
TVME $(1h)$			71.8 (+0.6)	RFI(1h)			84.5 (+0.5)				84.2(+0.2)
RFI(1h)	18.4 (-0.3)	65.6(-1)	84.5 (+0.5)		\ /				\ /		/
GLOBAL	-0.4	-0.6	+0.7	GLOBAL	-0.6	-0.5	+0.9	GLOBAL	-0.25	-0.3	+0.5

Conclusion

• A simple back-off values reordering can improve a Kneser-Ney based model for a 0.6% absolute gain of WER and 0.9% of CWR. • Our compact version leads to a slight improvement of the classical back-off model, with a very **low memory consumption**.

• We plan to extend the **co-occurence model** to more sophisticated heuristics and algorithms, smoothing values with word distances.

Proposed Approaches

Word co-occurence model

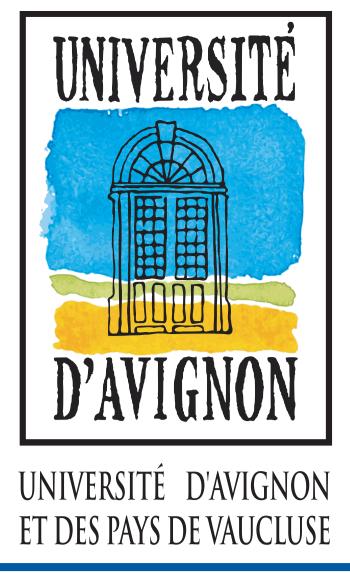
- n-gram language model.
- word co-occurrences.
- model, on 2-gram and 1-gram.
- In the case of a full back-off behavior, $\widetilde{p}(w_3|w_1,w_2) = lpha(w_1,w_2)^{1-eta} p_\psi(w_1,w_3)^eta lpha(w_2)^{1-eta} p_\psi(w_2,w_3)^eta p(w_3)$ smoothing function based on word co-occurrences.

Experiments

• Four series of experiments: for each approach separately, for the combination and for a compact model. • For each experiment, the initial back-off value is re-estimated according to: $\tilde{\alpha}(w_{i-n},..,w_i) = \alpha(w_{i-n},..,w_{i-1})^{1-\beta} p_{\phi}(w_{i-n},w_i)^{\beta}$ with $\tilde{\alpha}$: updated back-off value, α : initial back-off value, $p_{\phi}(w_{in}, w_i)$: smoothing function • The back-off based hollow grams has two advantages: it is a regular expression based model, it can capture hollow-grams events into the training corpora. • For the **back-off based on word co-occurrences**, a word co-occurrence symmetric matrix is built on the whole training corpora, counting word pairs with a window size of five words. • The compact model depicts the binary possibility of a back-off. This binary possibility is computed from the co-occurrence matrix: if the value is

not null, we consider as true the possibility of the combination.

Combination



• We combine a word association score [?] to the back-off based on classical

• It eliminates word frequency effects and emphasizes relations between significant

• Our model can be interpolated with an initial modified Kneser-ney back-off

where $\alpha(w_1, w_2)$: initial back-off value of the words, β : empirical fudge factor, p_{ψ} : back-off

Compact model

Ref