Lateen EM: Unsupervised Training with Multiple Objectives, Applied to Dependency Grammar Induction

VALENTIN I. SPITKOVSKY, HIYAN ALSHAWI AND DANIEL JURAFSKY

vals@stanford.edu, hiyan@google.com and jurafsky@stanford.edu

AIM

Mitigate *local optima* and *slow convergence* in unsupervised training, by using additional imperfect objectives.

SIMPLE LATEEN EM

Alternate ordinary "soft" and "hard" EM algorithms: switching objectives when stuck helps escape local optima.

EARLY-STOPPING LATEEN EM

Use one objective to validate moves proposed by the other: stop if the secondary objective gets worse.

FACTS

- EM guarantees to improve likelihood at every step
- EM tends to begin with large steps in a parameter space

- EM takes disproportionately many (and ever-smaller) steps to approach a likelihood's fixed point

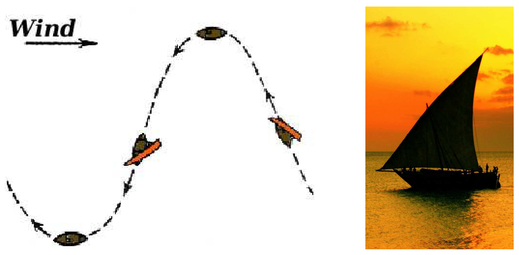
- these fixed points are almost invariably local optima
- moreover, underlying unsupervised likelihood objectives are, at best, loosely correlated with extrinsic performance

⟶ thus, we are justified in (occasionally) deviating from a prescribed training course...

INTUITION

Use *two* objectives (a primary and a secondary).

As a captain can't count on favorable winds, so an unsupervised learner can't rely on co-operative gradients. Lateen strategies de-emphasize fixed points, e.g., by tacking around local attractors, in a zig-zag fashion.



A triangular sail atop a traditional Arab sailing vessel, the *dhow* (right). Older square sails permitted sailing only before the wind. But the efficient *lateen* sail worked like a wing (with high pressure on one side and low pressure on the other), allowing a ship to go almost directly into a headwind. By *tacking*, in a zig-zag pattern, it became possible to sail in any direction, provided there was some wind at all (left). For centuries seafarers expertly combined both sails to traverse extensive distances, greatly increasing the reach of medieval navigation.

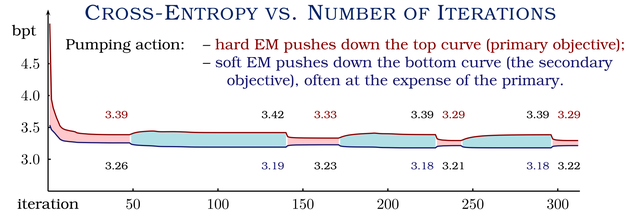
MORE PERFORMANT

5.5% higher accuracy, on average, compared to Viterbi EM, for dependency grammar induction tasks with CoNLL data

EXAMPLE:

Italian grammar induction improves from 41.8% to 56.2% directed dependency accuracy, after three lateen alternations.

CROSS-ENTROPY VS. NUMBER OF ITERATIONS



Cross-entropies, in bits per token (bpt), for the CoNLL 2007 Italian data set, initialized uniformly and trained on sentences up to length 45.

METHODOLOGY

- FACTORIAL EXPERIMENTAL DESIGN: controls for
  1. language (e.g., Arabic, Basque);
  2. initial objective (hard vs. soft EM);
  3. data size/complexity (sentence lengths);
  ... etc.

- MULTI-LINEAR REGRESSIONS: *joint* testing of hypotheses
  1. after all controls, do lateen strategies affect accuracy?
  2. and how do they affect running times?

ACKNOWLEDGMENTS

Partially funded by Air Force Research Laboratory (AFRL), under prime contract no. FA8750-09-C-0181, and by NSF, via award #IIS-0811974. We thank Angel X. Chang, Spence Green, David McClosky, Fernando Pereira, Slav Petrov and anonymous reviewers, for helpful comments on draft versions of the paper, and Andrew Ng, for a stimulating discussion.

MORE EFFICIENT

30% faster, on average, than either standard EM, for dependency grammar induction tasks with CoNLL data

EXAMPLE:

Training runs that terminate early are nearly twice as fast, and only two score slightly lower than standard training:

CoNLL Year & Language	standard *DDA*	standard *iters*	lateen *DDA*	lateen *iters*
Arabic 2006	28.4	180	28.4	118
Bulgarian '06	39.1	253	**39.6**	131
Chinese '06	49.4	268	49.4	204
Dutch '06	21.3	246	**27.8**	35
Hungarian '07	17.1	366	**17.4**	213
Italian '07	39.6	194	39.6	164
Japanese '06	56.6	113	56.6	93
Portuguese '06	37.9	180	*37.5*	102
Slovenian '06	30.8	234	**31.1**	118
Spanish '06	33.3	125	*33.1*	73
Average:	35.4	216	36.1	125

Accuracies and iteration counts for training runs affected by early termination with soft EM, using shorter sentences and ad-hoc initialization.

SUMMARY

- EARLY STOPPING: faster, with same accuracy, for both EMs — could be used to more fairly compare learners with radically different objectives, requiring quite different numbers of steps (or magnitude changes in cross-entropy) to converge.

- HARD EM: lateen strategies consistently improve accuracy — once stuck, the longer we follow the secondary objective to dig ourselves out (e.g., for just one step, until the primary suffers, or to convergence), the better we score in the end.

- SOFT EM: lateen strategies do not affect accuracy — hence, it seems best simply to terminate early.