IBM 1, 1 iteration with the toy sentences : (Notation & algorithm identical to the one used in [1])

Sentence 1: maison blue - blue house Sentence 2: maison - house

Uniform Initialization

$house \mid maison = 0.5$ $\quad$ $house \mid blue = 0.5$

$bleu \mid maison = 0.5$ $\quad$ $bleu \mid blue = 0.5$

Expectation Step - Mapper

Sentence 1:

$s\_total(e) += t(e|f)$

$s\_total(house) = 0.5 + 0.5 = 1.0$

$s\_total(bleu) = 0.5 + 0.5 = 1.0$

$count(e|f) += \frac{t(e|f)}{s\_total(e)}$

$count(house|maison) = \frac{0.5}{1.0} = 0.5$

$count(bleu|maison) = \frac{0.5}{1.0} = 0.5$

$count(house|bleu) = \frac{0.5}{1.0} = 0.5$

$count(bleu|bleu) = \frac{0.5}{1.0} = 0.5$

Sentence 2:

$s\_total(house) = 0.5$

$count(house|maison) = \frac{0.5}{0.5} = 1.0$

Expectation Step - Reducer (Assumption: the foreign words f were the keys sent to Reducer)

Finish summing up: $count(e|f) += \frac{t(e|f)}{s\_total(e)}$

$count(house|maison) = 0.5 + 1.0 = 1.5$

$count(bleu|maison) = 0.5$ (no change)

$count(house|bleu) = 0.5$ (no change)

$count(bleu|bleu) = 0.5$ (no change)

$total(f) += \frac{t(e|f)}{s\_total(e)}$

$total(bleu) = 0.5 + 0.5 = 1.0$

$total(maison) = 0.5 + 0.5 + 1.0 = 2.0$

Maximization Step - Reducer

$t(e|f) = \frac{count(e|f)}{total(f)}$

$t(house|maison) = \frac{1.5)}{2.0} = 0.75$ $\quad$ $t(house|bleu) = \frac{0.5)}{1.0} = 0.5$

$t(bleu|maison) = \frac{0.5)}{2.0} = 0.25$ $\quad$ $t(bleu|bleu) = \frac{0.5)}{1.0} = 0.5$

IBM 1 log space, 1 iteration with the toy sentences : (Notation & algorithm identical to the one used in [1], a * after an operator symbolises the operation happens in log space $\rightarrow *$ becomes $+$, $/$ becomes $-$, $+$ becomes $a + log(1 + e^{b-a})$ with $a \geq b$, else reverse a and b in the equation)

Sentence 1: maison blue - blue house Sentence 2: maison - house

Uniform Initialization

$house \mid maison = log(0.5) = -0.6931$

$bleu \mid maison = log(0.5) = -0.6931$

$house \mid blue = log(0.5) = -0.6931$

$bleu \mid blue = log(0.5) = -0.6931$

Expectation Step - Mapper

Sentence 1:

$s\_total(e) += t(e|f)$

$s\_total(house) = -0.6931 +^* -0.6931 = 0$

$s\_total(bleu) = -0.6931 +^* -0.6931 = 0$

$count(e|f) += \frac{t(e|f)}{s\_total(e)}$

$count(house|maison) = -0.6931 /^* 0 = -0.6931$

$count(bleu|maison) = -0.6931 /^* 0 = -0.6931$

$count(house|bleu) = -0.6931 /^* 0 = -0.6931$

$count(bleu|bleu) = -0.6931 /^* 0 = -0.6931$

Sentence 2:

$s\_total(house) = -0.6931$

$count(house|maison) = -0.6931 /^* - 0.6931 = 0$

Expectation Step - Reducer (Assumption: the foreign words f were the keys sent to Reducer)

Finish summing up: $count(e|f) += \frac{t(e|f)}{s\_total(e)}$

$count(house|maison) = -0.6931 +^* 0 = 0.4054$

$count(bleu|maison) = -0.6931$ (no change)

$count(house|bleu) = -0.6931$ (no change)

$count(bleu|bleu) = -0.6931$ (no change)

$total(f) += \frac{t(e|f)}{s\_total(e)}$

$total(bleu) = -0.6931 +^* -0.6931 = 0$

$total(maison) = -0.6931 +^* 0.4054 + 1.0 = 0.6931$

Maximization Step - Reducer

$t(e|f) = \frac{count(e|f)}{total(f)}$

$t(house|maison) = 0.4054 /^* - 0.6931 = -0.2877$

$t(bleu|maison) = -0.6931 /^* - 0.6931 = -1.3862$

$t(house|bleu) = -0.6931 /^* 0 = -0.6931$

$t(bleu|bleu) = -0.6931 /^* 0 = -0.6931$

HMM, 1 iteration with the toy sentences : (Notation & algorithm identical to the one used in [2])

Sentence 1: maison blue - blue house Sentence 2: maison - house

Translation Probability from 1 Iteration of IBM 1, Rest Uniform Initialization

$translation\ \ table/emission\ \ probability\ \ given\ \ state$

$house \mid maison/0 = 0.5$

$bleu \mid maison/0 = 0.5$

$initial\ \ state\ \ probability : \pi = [0.5 \quad 0.5]$

$house \mid blue/1 = 0.5$

$bleu \mid blue/1 = 0.5$

$transition\ \ matrix\ \ A_{iq}$:

Forward algorithm - Mapper

Sentence 1:

$\alpha[0][0] = \pi[0] * B(bleu|0) = 0.5 * 0.25 = 0.125$

$\alpha[0][1] = \pi[0] * B(bleu|1) = 0.5 * 0.5 = 0.25$

$sum[1][0] = \alpha[0][0] * A_{00} + \alpha[0][1] * A_{10} = 0.125 * 0.5 + 0.25 * 0.5 = 0.1875$

$sum[1][1] = \alpha[0][0] * A_{01} + \alpha[0][1] * A_{11} = 0.125 * 0.5 + 0.25 * 0.5 = 0.1875$

$\alpha[1][0] = sum[1][0] * B(house|0) = 0.1875 * 0.75 = 0.140625$

$\alpha[1][1] = sum[1][1] * B(house|1) = 0.1875 * 0.5 = 0.09375$

Sentence 2: $\alpha[0][0] = \pi[0] * B(house|0) = 0.5 * 0.25 = 0.125$

Backward algorithm - Mapper

Sentence 1:

$\beta[1][0] = 1.0$

$\beta[1][1] = 1.0$

$\beta[0][0] = \beta[1][0] * A_{00} * B(house|0) + \beta[1][1] * A_{01} * B(house|1) = 1.0 * 0.5 * 0.75 + 1.0 * 0.5 * 0.5 = 0.625$

$\beta[0][1] = \beta[1][0] * A_{10} * B(house|0) + \beta[1][1] * A_{11} * B(house|1) = 1.0 * 0.5 * 0.75 + 1.0 * 0.5 * 0.5 = 0.625$

Sentence 2:

$\beta[0][0] = 1.0$

Initial State Probability - Mapper

Sentence 1:

$I(0) = \alpha[0][0] * \beta[0][0] = 0.125 * 0.625 = 0.0781$

$I(1) = \alpha[0][1] * \beta[0][1] = 0.25 * 0.625 = 0.1563$

Sentence 2:

$I(0) = \alpha[0][0] * \beta[0][0] = 0.375 * 1.0 = 0.375$

Emissions - Mapper

Sentence 1:

$O(0)(house) = \alpha[1][0] * \beta[1][0] = 0.1406 * 1.0 = 0.1406$

$O(1)(house) = \alpha[1][1] * \beta[1][1] = 0.09375 * 1.0 = 0.0938$

$O(0)(blue) = \alpha[0][0] * \beta[0][0] = 0.125 * 0.625 = 0.0781$

$O(1)(blue) = \alpha[0][1] * \beta[0][1] = 0.25 * 0.625 = 0.1563$

Sentence 2:

$O(0)(house) = \alpha[0][0] * \beta[0][0] = 0.375 * 1.0 = 0.375$

Transitions - Mapper

Sentence 1:

$T(0)(0) = \alpha[0][0] * A_{00} * B(house|0) * \beta[1][0] = 0.125 * 0.5 * 0.75 * 1.0 = 0.0469$

$T(0)(1) = \alpha[0][0] * A_{01} * B(house|1) * \beta[1][1] = 0.125 * 0.5 * 0.5 * 1.0 = 0.0313$

$T(1)(0) = \alpha[0][1] * A_{10} * B(house|0) * \beta[1][0] = 0.25 * 0.5 * 0.75 * 1.0 = 0.0938$

$T(1)(1) = \alpha[0][1] * A_{11} * B(house|1) * \beta[1][1] = 0.25 * 0.5 * 0.5 * 1.0 = 0.0625$

Sentence 2: no transitions

Summing up - Reducer (only changes are listed)

$I(0) = 0.78125 * +0.0781 = 0.4531$

$O(0)(house) = 0.1406 + 0.375 = 0.5156$

Final Parameters - Reducer

$z(emitfrommaison/0) = 0.0781 + 0.5156 = 0.5937$

$blue|maison = 0.0781/0.5937 = 0.1315$ $house|maison = 0.5156/0.5937 = 0.8685$

$z(emitfromblue/1) = 0.1563 + 0.0938 = 0.25$

$blue|blue = 0.1563/0.25 = 0.625$ $house|blue = 0.0938/0.25 = 0.375$

$z(transitfrom0) = 0.0469 + 0.0313 = 0.0781$

$A_{00} = 0.0469/0.0781 = 0.6$ $A_{01} = 0.0313/0.0781 = 0.4$

$z(transitfrom1) = 0.0938 + 0.0625 = 0.1563$

$A_{00} = 0.0938/0.1563 = 0.6$ $A_{01} = 0.0625/0.1563 = 0.4$

$z(initialstateprobability) = 0.4531 + 0.1563 = 0.6094$

$\pi[0] = 0.4531/0.6094 = 0.7436$ $\pi[1] = 0.1563/0.6094 = 0.2564$

Viterbi Algorithm after 1 Iteration of IBM 1 and HMM Respectively

(Notation & algorithm identical to the one used in [2])

Sentence 1:

$\gamma[0][0] = \pi[0] * B(blue|0) = 0.7436 * 0.1315 = 0.0977 \ \ , bp = -1$

$\gamma[0][1] = \pi[1] * B(blue|1) = 0.7436 * 0.625 = 0.1603 \ \ , bp = -1$

$\gamma[1][0] : argmax\{$

$\gamma[0][0] * A_{00} * B(house|0) = 0.0977 * 0.6 * 0.8685 = 0.0509,$

$\gamma[0][1] * A_{10} * B(house|0) = 0.1603 * 0.6 * 0.8685 = 0.0835\}$

$\rightarrow \gamma[1][0] = 0.0835 \rightarrow bp = 1$

$\gamma[1][1] : argmax\{$

$\gamma[0][0] * A_{01} * B(house|1) = 0.0977 * 0.4 * 0.375 = 0.0147,$

$\gamma[0][1] * A_{11} * B(house|1) = 0.1603 * 0.4 * 0.375 = 0.024\}$

$\rightarrow \gamma[1][1] = 0.024 \rightarrow bp = 1$

Best sequence for "blue house": maison(house) blue(blue) with probability 0.0835

Sentence 2 (trivial):

$\gamma[0][0] = \pi[0] * B(house|0) = 0.7436 * 0.8685 = 0.6458$

Best sequence for "house": maison(house) with probability 0.6458

References

[1] Philipp Koehn. *Statistical Machine Translation*. Cambridge University Press, 2010.

[2] Jimmy Lin and Chris Dyer. *Data-Intensive Text Processing with MapReduce*. Morgan & Claypool Publishers, 2010.