

# Decompose Restaurant(DR) 2

Donglai Wei

2010.6.20

## 1) Algorithm (DR)

GOAL: Maximize log Probability  $P = \log p(x, z | \lambda, \alpha, \gamma)$

- (i) Make Restaurant j into one table t1 where customers following uniform distribution
- (ii) Iterate until no customers are left in this uniform table t1:
  - (a) For each dish k, propose to form a new table<sup>1\*</sup> out of t1 with dish k and calculate the change  $\Delta P$
  - (b) Sample<sup>2\*</sup> a proposal according to the weight and make the new table
- (iii) TKM:Local search Table/Dish(allowing new dish)+Merge dish
- (iv) Decision<sup>3\*</sup>

**Remarks::**

Sampling weight:  $e^{r\Delta P}$ :  $r(>0)$  is the tuning parameter(the more decrease of  $\Delta P$  the less weight)

	Previous Approach	Current Approach
<sup>1*</sup> Form new table	Greedily assign overlapped words(threshold/not)	sampling word assignment $\sim (\frac{1}{W}, \frac{n_{..k}^w + \phi}{n_{..k} + W\phi})$
<sup>2*</sup> Sample proposal	Hard Assignment ( $r_{accept} = \infty$ )	Soft Assignment: $r_{proposal} \in [0, 1]$
<sup>3*</sup> Decision:	Accept all( $r_{accept} = 0$ )/Accept&Reject( $r_{accept} = \infty$ )	Sample $r_{accept} \in [0, 1]$

## Decompose Restaurants

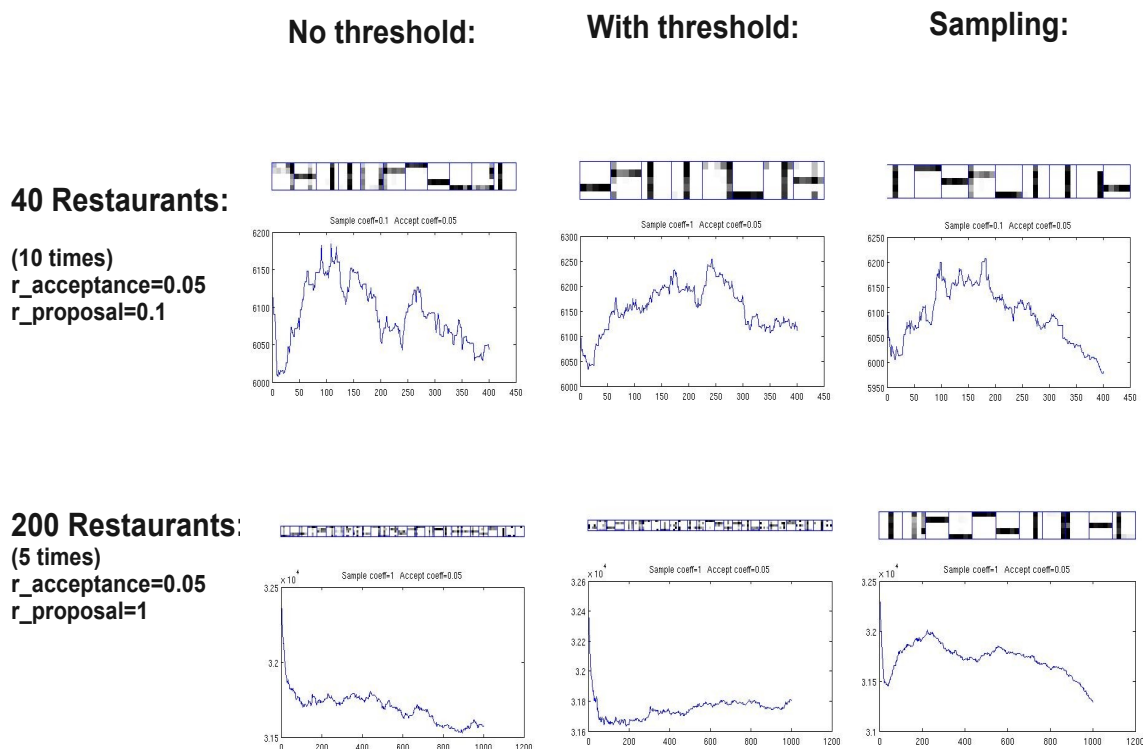


Figure 1: With same set of parameters, sampling new table assignment (right column) is much better than deterministic (greedy) ones (left, mid column)

## 2) Experiment

### i) Sampling new table proposal

<sup>1\*</sup> (Sampling new tables instead of Deterministic assignment) has **GREAT** improvement!!! (No need for Decompose Dish any more)

- 1) Hard assignment may seduce dishes to keep their mixture of bars since it has less variability.
- 2) Sampling new table tends to try out more possibilities, which may finally break down the mixture of bars after accumulation.

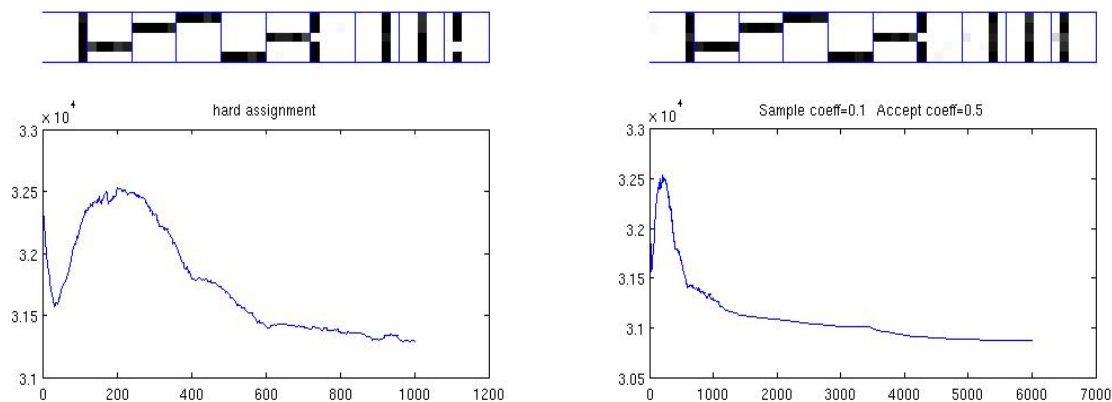


Figure 2: (left) is the initial 5 DR strategy a) moves; (right) More DR strategy a) runs with Accept/Reject

## ii) Comparison of 2 strategies

(Below, algorithms all use Sampling new table assignment in <sup>\*1</sup>)

a) Peaky new table Proposal( $r_{proposal} \sim 1$ ) + More Acceptance( $r_{accept} < 1$ )

### Good:

1) Quickly shrink the number of dishes:

In the beginning, every restaurant is its own dish. Peaky proposal greedily try out one searching direction instead of wasting time wandering around. Decompose restaurant tends to interpret the restaurant with more dishes which may decrease  $\Delta P$  to vary degrees. Thus we should be more tolerant during Initialization.

### Bad:

- 1) Maybe vulnerable to local optima, since it's lack of variability
- 2) Don't know when to stop

### Further Remarks:

- 1) I also go to extremes that: Find the best new table Proposal + Accept All, which works equally well. So I guess the variability in <sup>\*1</sup>(sampling) is enough and the greedy decision may not be too bad to be rejected.
- 2) Though it finds 10 bars quickly,  $P$  is still a little lower than some of the noisy configurations. (More DR runs with Accept/Reject needed, guaranteed to decrease)

## Strategy 1: Peak Proposal+More Acceptance

40 Restaurants:(10 times)    200 Restaurants:(5 times)    1000 Restaurants:(5 times)

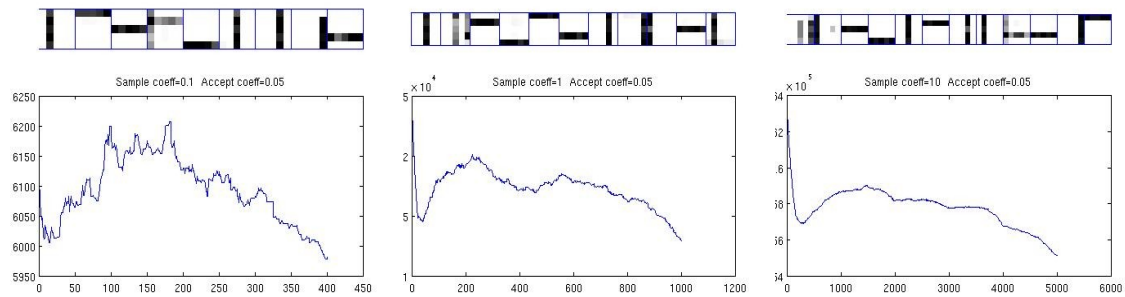


Figure 3: A set of parameters that Peak new table proposal+More Acceptance work

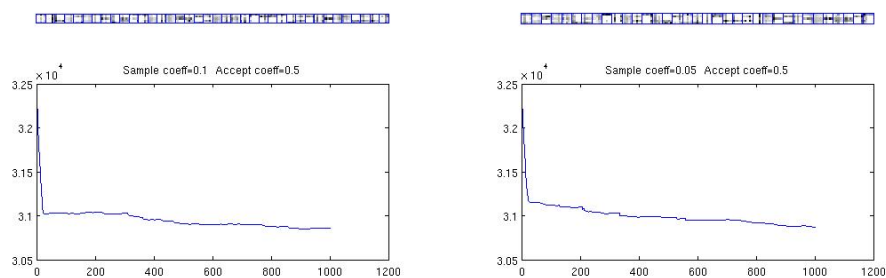


Figure 4: Flatter Proposal+Less Acceptance: the  $P$  is really good

b) Flatter Proposal( $r_{proposal} \ll 1$ )+Less Acceptance( $r_{accept} \sim 1$ )

**Good:**

1) More variability during the search:

Since the chance to get "worse" (in terms of long-term searching direction) config increases, we need sharper Acceptance Decision.

**Bad:**

1) Too Slow to reduce the number of dishes, since it wastes lots of time wondering around.

2) Tend to have noisy dishes: Since the Acceptance Decision is sharper, many of the potential "better" config(tendency to break down the noisy dish) are rejected, leaving dishes nasty.

**Further Remarks:**

1) Though it is hopeless to find 10 bars(I tried more iterations), it has equally good  $P$ . Bigger tables and Noisy dishes maynot be a bad interpretation in terms of  $P$