CMPS 561 Assignment #3, Fall 2010

Vijay V Raghavan

Assigned: November 16, 2017

Due: November 29, 2017

Note:

1. All details of work for each question must be submitted.

2. Staple the question and answer sheet together

3. Make a cover with Name, CLID

4. Number all pages and give an index to each question.

*5. Most importantly, any sort of cheating will **NOT** be tolerated. More information can be found on class Web page on cheating policy.*

Q1. **30 Points**

In the Vector Space Model, the relationships among different terms can be expressed as a term-term matrix G_t, which is called the Grammarian matrix. The term-document relationship is shown in Table 1.

	d_1	d_2	d_3	d_4
t_1	3	0	1	3
t_2	0	1	3	0
t_3	3	0	0	2
t_4	2	0	1	3

Table 1

(a). If $G_t = I$, and that a query is given by $q = 2\vec{t}_1 + \vec{t}_3$, calculate the RSVs for d_1 through d_4 with respect to q.

(b). Repeat part a) if $G_t = \begin{bmatrix} t_1.t_1 & t_1.t_2 & t_1.t_3 & t_1.t_4 \\ t_2.t_1 & t_2.t_2 & t_2.t_3 & t_2.t_4 \\ t_3.t_1 & t_3.t_2 & t_3.t_3 & t_3.t_4 \\ t_4.t_1 & t_4.t_2 & t_4.t_3 & t_4.t_4 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 & -0.4 & 0.5 \\ 0.1 & 1 & -0.3 & 0.2 \\ -0.4 & -0.3 & 1 & 0.1 \\ 0.5 & 0.2 & 0.1 & 1 \end{bmatrix}$

(c). Pick a document for which RSV in part (b) is greater than that in part (a) and explain which element(s) (term relationships) from G_t cause this change.

(d). Can the RSV of a document become smaller when G_t is incorporated into the RSV computation? If yes, explain what the characteristics of term-term relationships matrix are that will cause this effect.

(e). Compute the RSVs for d_1 through d_4 with respect to q, assuming the GVSM model is employed. In this case, no assumption about G_t is needed.

Q2. **40 Points**

Answer the questions based on the following table shown in the figure

	t1	t2	t3	t4		Relevance
d1	0	2	0	2		REL
d5	1	3	1	0		REL
d7	0	3	1	2		NREL
d3	1	3	0	0		REL
d6	0	2	1	1		NREL

Fig 1: W matrix for training.

(a). Use the gradient descent approach based on (standard) Perception Criterion discussed in Class and determine the optimum query.

(b). Uses the gradient descent approaches based on Generalized Perception Criterion discussed in class and determine the optimum query. We have:

$$b1 = d1 - d7, b2 = d1 - d6, b3 = d5 - d7, b4 = d5 - d6, b5 = d3 - d7, b6 = d3 - d6$$

(c). Use Generalized Perception—learning by sample discussed in Class and determine the optimum query. We have:

$$b1 = d1 - d7, b2 = d1 - d6, b3 = d5 - d7, b4 = d5 - d6, b5 = d3 - d7, b6 = d3 - d6$$

Only learn from b_i s in the forward direction, by following the order, as listed next: $\bar{b}1, \bar{b}2, \bar{b}3, \bar{b}4, \bar{b}5, \bar{b}6$

(d). Obtain the optimal query based on Rochhio's method.

Q3. **10 Points**

Use the following W' as test instances set to compare the above four methods (use the three optimal queries derived by them). Use R_{norm} measure for evaluation.

	t1	t2	t3	t4		Relevance
d2	0	2	0	2		REL
d4	2	3	1	0		REL
d8	0	1	1	0		NREL
d9	1	2	0	0		REL
d10	1	3	2	1		NREL

Fig 2: W' matrix for testing