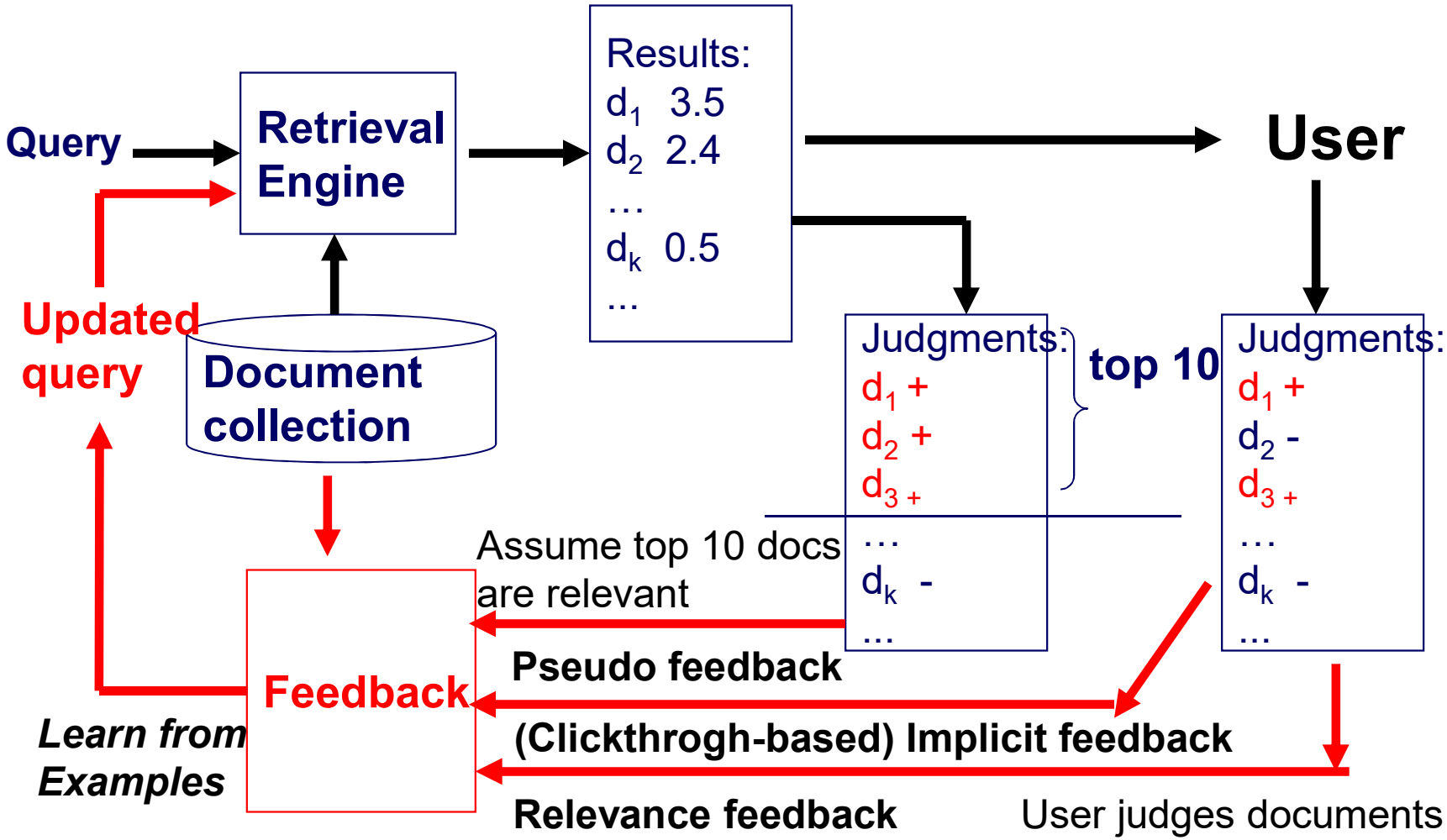
Relevance Feedback for Vector Space Model

Quach Dinh Hoang

Slides are obtained from [Zhai and Massung, 2016]

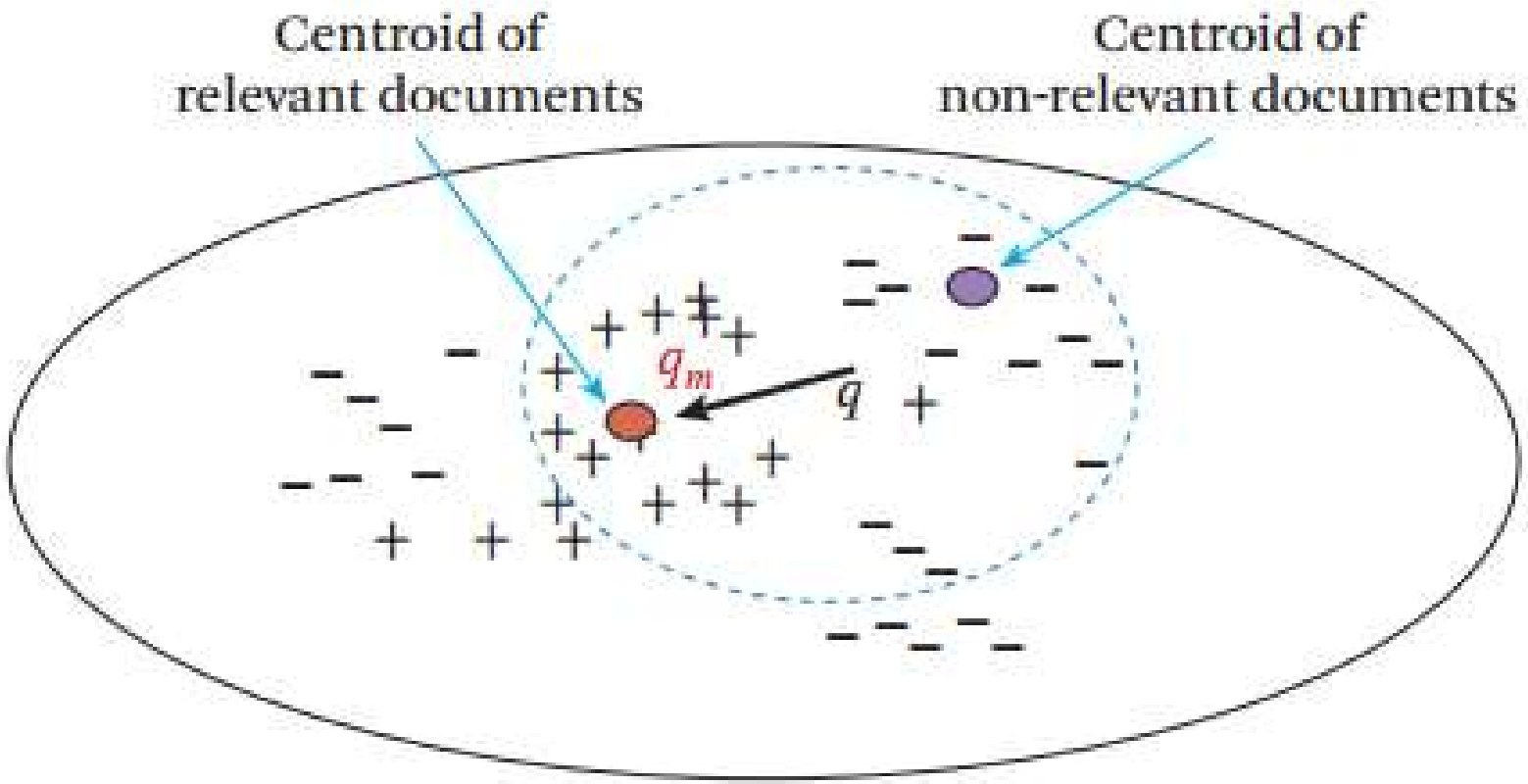
Feedback in Information Retrieval



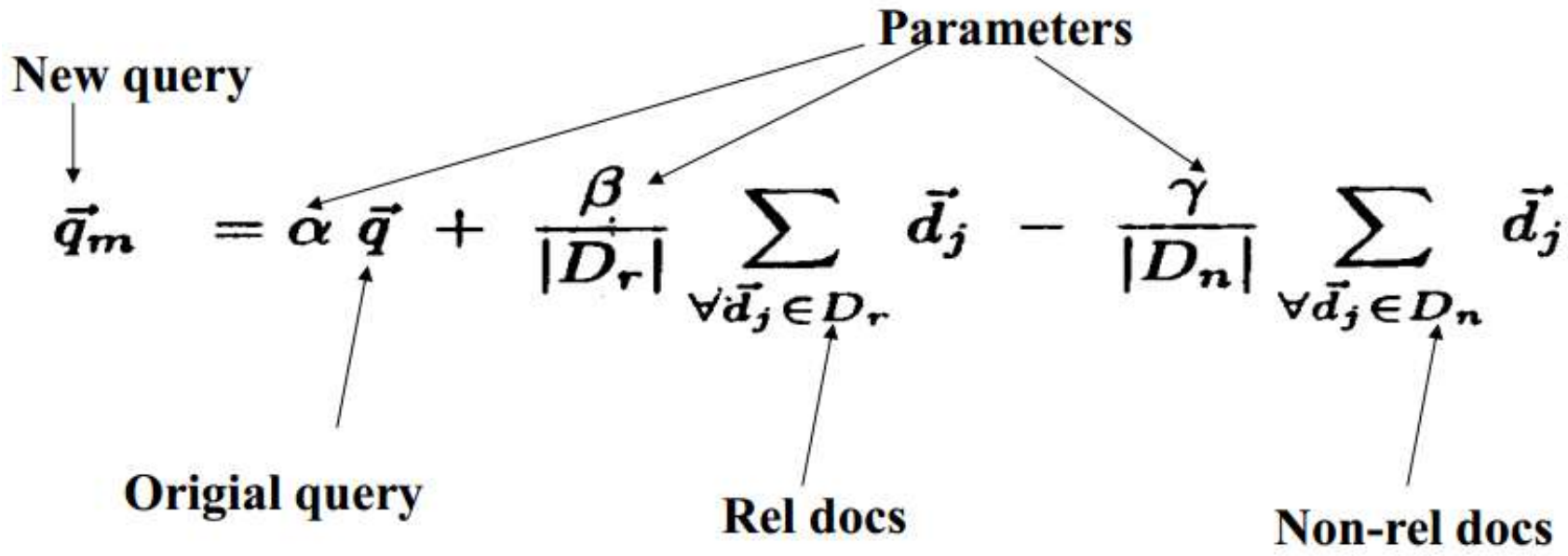
Feedback in Vector Space Model

- How can a TR system learn from examples to improve retrieval accuracy?
  - Positive examples: docs known to be relevant
  - Negative examples: docs known to be non-relevant
- General method: query modification
  - Adding new (weighted) terms (query expansion)
  - Adjusting weights of old terms

Rocchio Feedback: Illustration



Rocchio Feedback: Formula



Example of Rocchio Feedback

$$V = \{news, about, presidential, campaign, food, text\}$$

$$\vec{q} = \{1, 1, 1, 1, 0, 0\}$$

		news	about	pres.	campaign	food	text
−	d_1	1.5	0.1	0.0	0.0	0.0	0.0
−	d_2	1.5	0.1	0.0	2.0	2.0	0.0
+	d_3	1.5	0.0	3.0	2.0	0.0	0.0
+	d_4	1.5	0.0	4.0	2.0	0.0	0.0
−	d_5	1.5	0.0	0.0	6.0	2.0	0.0

		news	about	pres.	campaign	food	text
+	C_r	$\frac{1.5+1.5}{2}$	0.0	$\frac{3.0+4.0}{2}$	$\frac{2.0+2.0}{2}$	0.0	0.0
−	C_n	$\frac{1.5+1.5+1.5}{3}$	$\frac{0.1+0.1+0.0}{3}$	0.0	$\frac{0.0+2.0+6.0}{3}$	$\frac{0.0+2.0+2.0}{3}$	0.0

$$\vec{q}_m = \alpha \cdot \vec{q} + \beta \cdot C_r - \gamma \cdot C_n$$

$$= \{\alpha + 1.5\beta - 1.5\gamma, \alpha - 0.067\gamma, \alpha + 3.5\beta, \alpha + 2\beta - 2.67\gamma, -1.33\gamma, 0\}$$

Rocchio in Practice

- Negative (non-relevant) examples are not very important (why?)
- Often truncate the vector (i.e., consider only a small number of words that have highest weights in the centroid vector) (efficiency concern)
- Avoid “over-fitting” (keep relatively high weight on the original query weights) (why?)
- Can be used for relevance feedback and pseudo feedback (β should be set to a larger value for relevance feedback than for pseudo feedback)
- Usually robust and effective

Summary of Feedback in Information Retrieval

- Feedback = learn from examples
- Three major feedback scenarios
  - Relevance, pseudo, and implicit feedback
- Rocchio for VSM

References

- ChengXiang Zhai and Sean Massung, *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*, ACM Books, 2016.
  - Chapter 7, Section 7.1 (Feedback in the Vector Space Model)