



北京交通大学

BEIJING JIAOTONG UNIVERSITY



基于隐主题马尔科夫模型的多特征自动文摘

刘江鸣，徐金安，张玉洁

北京交通大学





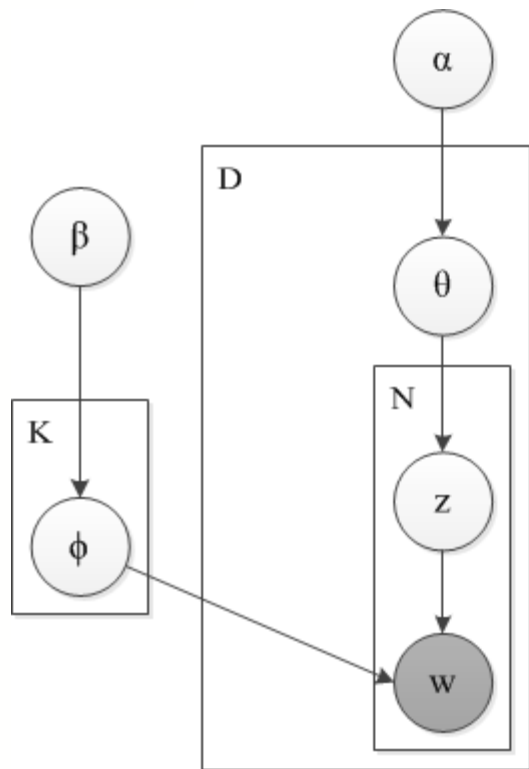
- 多文档自动摘要
- 方法

隐主题马尔科夫模型 (HTMM) ——缓解主题独立假设

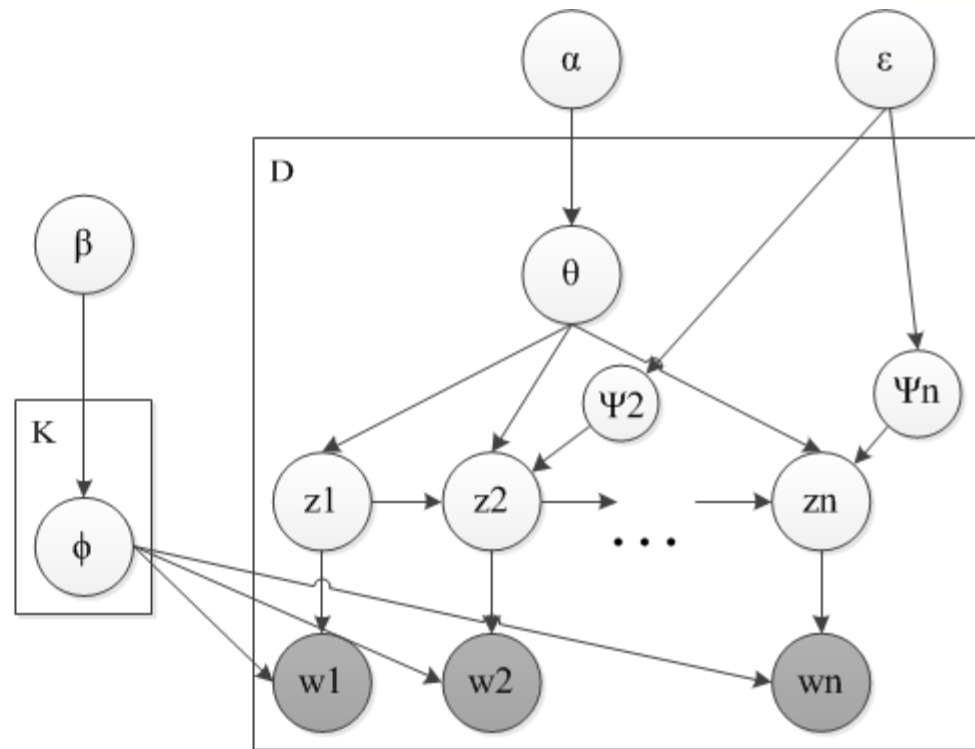
结合文档其他有效特征 ——主要内容



LDA & HTMM



(a)



(b)



北京交通大学

BEIJING JIAOTONG UNIVERSITY

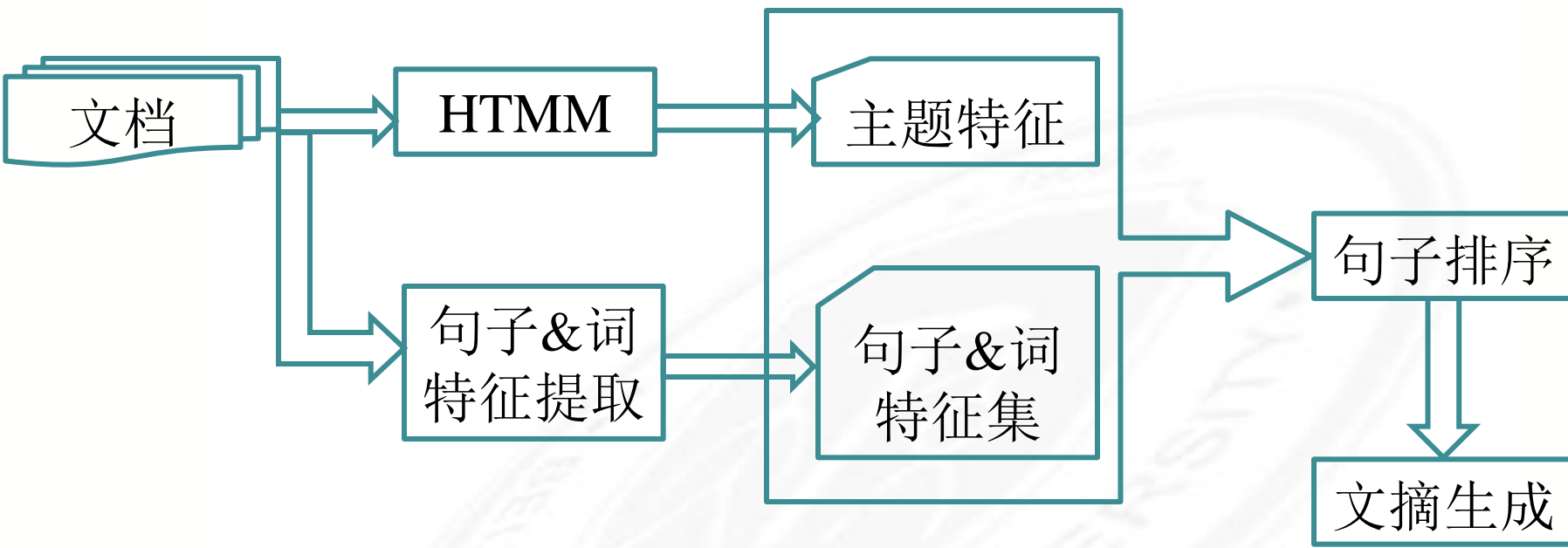
文档特征

- 句子级特征
位置、长度、标题相似度、命名实体
- 词级特征
位置、词频





系统框架

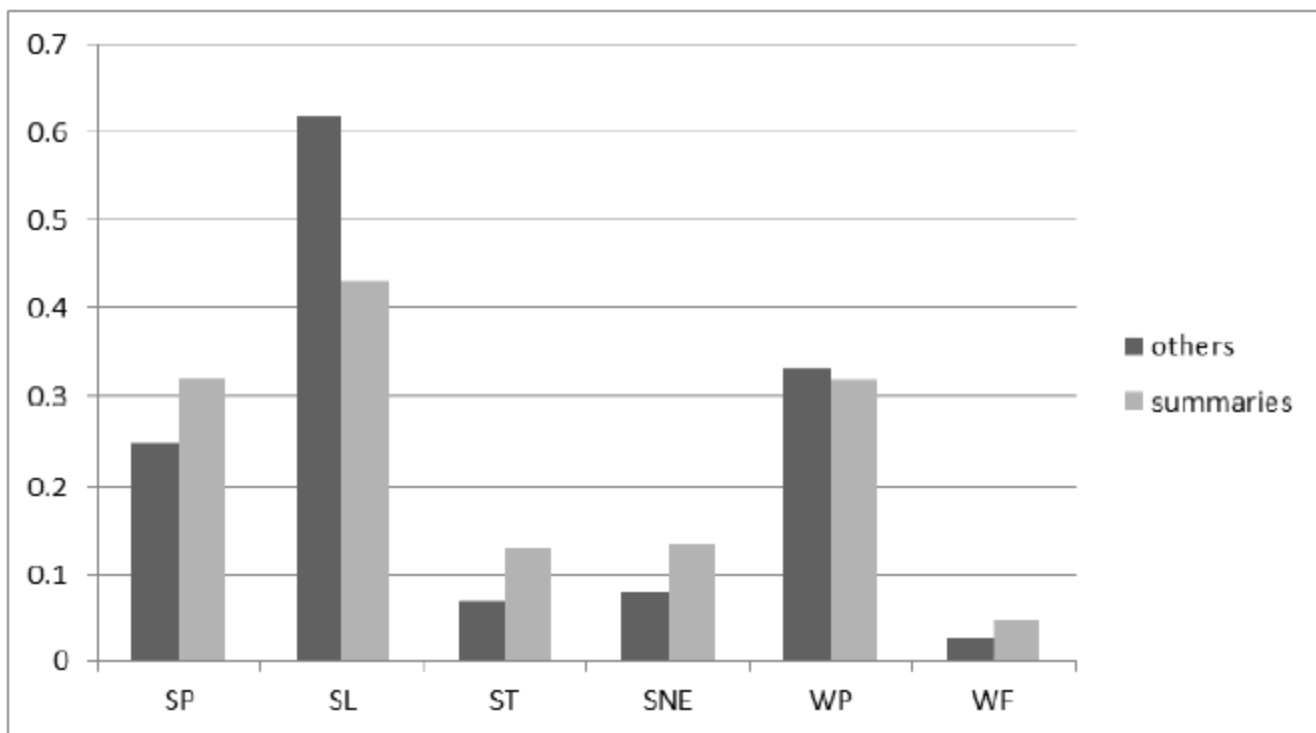




• DUC2007

	ROUGE-1	ROUGE-2	ROUGE-3	ROUGE-4	ROUGE-L	ROUGE-SU4
Features	0.32272	0.05056	0.01309	0.00582	0.29022	0.10133
LDA	0.26861	0.02523	0.00343	0.00130	0.24962	0.07454
LDA+Features	0.27142	0.02767	0.00406	0.00139	0.25165	0.07626
HTMM	0.36076	0.06505	0.01652	0.00626	0.32947	0.11843
HTMM+Features	0.37125	0.07686	0.02231	0.00940	0.34096	0.12672

文摘句与非文摘句
的特征表现





北京交通大学

BEIJING JIAOTONG UNIVERSITY



谢谢

