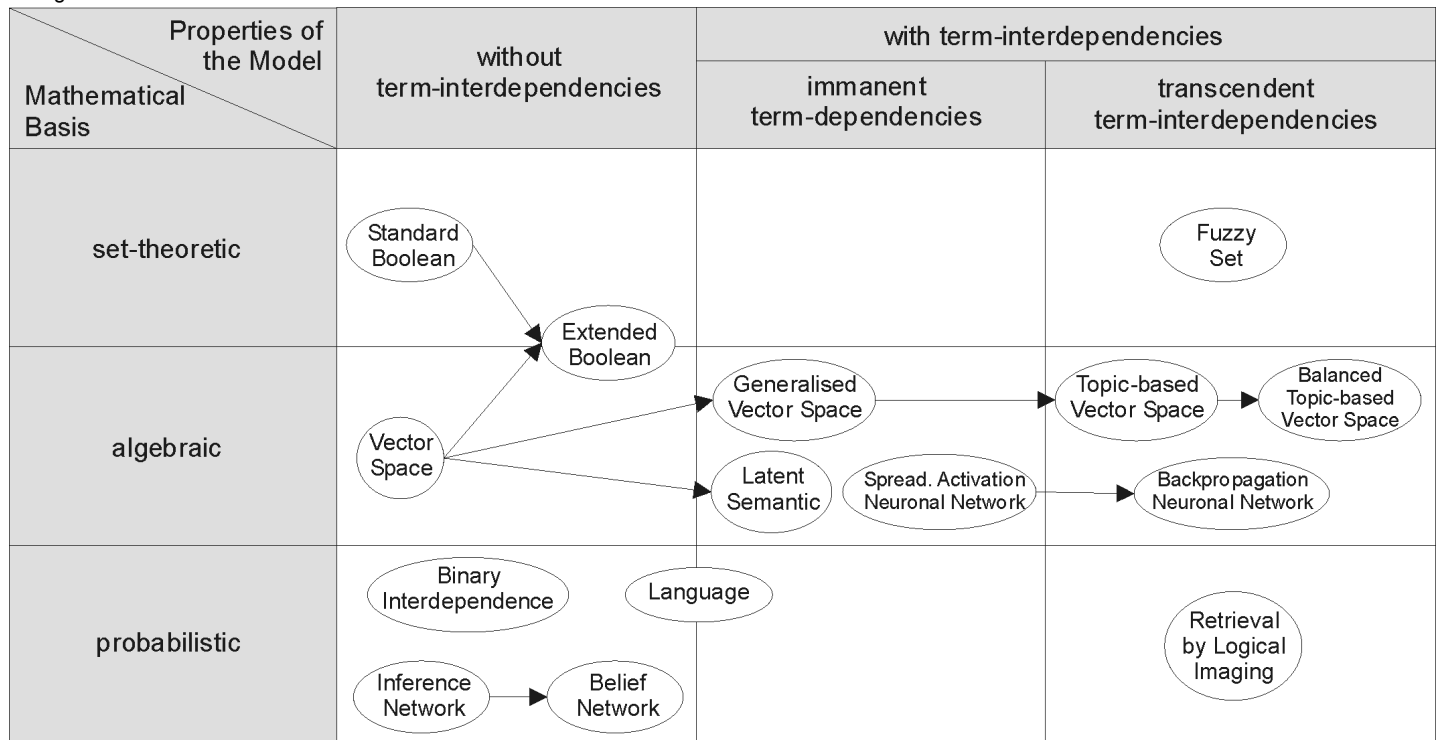


Project dataset-sts: paper collection

Basic concepts

Information Retrieval [wiki](#)

- In the 1960s, the first large information retrieval research group was formed by Gerard Salton at Cornell.
- Categorization of IR-models



- Performance and correctness measure
 - Precision (除以检索文本)
 - Recall (除以相关的文本, 即答案)
 - Fall-out
 - F-score/F-measure
 - **Average precision**
 - Precision at K
 - R-precision
 - **Mean Average precision** the mean of the average precision scores for each query
 - Discounted cumulative gain
 - **Mean reciprocal rank** - The mean reciprocal rank is the average of the reciprocal ranks of results for a sample of queries Q: the rank position of the first relevant document for the i-th query.

Results

Dataset Size

Dataset	Train pairs	Val. pairs	Test pairs	Val.-Test r	Ev. # s_0	Ev. # s_1 per s_0
wang	44648	1149	1518	-0.078	178	34.9 ±131%
yodaqa/large2470	220846	55052	120069	0.348	1100	159.2 ±100%
wqmprop	407465	137235	277509	0.836	3430	118.753 ±85%
Ubuntu Dialogue v2	1M	195600	189200	0.884	38480	10

[Ubuntu-dialog](#)

- Train set : 1M
- Val set : 195600
- Test set : 189200

Model	Time	valMRR	val-2R@1	val-10R@2	testMRR	test-2R@1	test-10R@2	settings
avg		0.636035	0.805922	0.623185	0.639989	0.804449	0.626524	inp_e_dropout=0 dropout=0
		±0.001673	±0.001594	±0.001892	±0.001760	±0.002325	±0.002013	
DAN	2500s/epoch eta = 56 hr							inpedropout=0 inpwdropout=1/3 deep=2 "pact='relu'"
rnn	19043s/epoch eta = 423 hr							sdim=1 pdim=1 ptscorer=B.dot_ptscorer
cnn	2060747s/epoch eta = 45790 hr							pdim=1 ptscorer=B.dot_ptscorer
rnnCNN	901966s/epoch eta = 20040hr							sdim=1/2 pdim=1 ptscorer=B.dot_ptscorer
attn1511	198683s/epoch eta = 4420 hr							sdim=1/2 cdim=1/2 ptscorer=B.dot_ptscorer

Wang

The classic academic standard stems from the TREC-based dataset originally by Wang et al., 2007, in the form by Yao et al., 2013

- Train set : 44648
- Val set : 1149
- Test set : 1518

Model	Time	trainAllMRR	devMRR	testMAP	testMRR	settings
termfreq TF-IDF #w		0.714169	0.725217	0.578200	0.708957	<code>freq_mode='tf'</code>
termfreq BM25 #w		0.813992	0.829004	0.630100	0.765363	(defaults)
Tan (2015)				0.728	0.832	QA-LSTM/CNN+attention; state-of-art 2015
dos Santos (2016)				0.753	0.851	Attentive Pooling CNN; state-of-art 2016
Wang et al. (2016)				0.771	0.845	Lexical Decomposition and Composition; state-of-art 2016
----- -----	-----	-----	-----	-----	-----	
avg	20s/epoch eta = 0.44 hr					(defaults)
DAN	28s/epoch eta = 0.62 hr					<code>inp_e_dropout=</code> <code>inp_w_dropout=1/3</code> <code>deep=2</code> <code>pact='relu'</code>
----- -----	-----	-----	-----	-----	-----	
rnn	2100s/epoch eta = 46.66 hr					(defaults)
cnn	26138s/epoch eta = 580hr					(defaults)
rnnCNN	30101s/epoch eta = 668.1hr					(defaults)
attn1511	15333s/epoch eta = 341hr					(defaults)

Accumulate

- immanent 天生的，内在的 term-dependencies
- transcendent 超然的 term-interdependencies
- fuzzy set 毛茸茸的，含混不清的
- wiggle 摇动
- interpolated 插入的