TF-IDF Uncovered

Thomas Roelleke
Jun Wang, Hengzhi Wu, Hany Azzam
Queen Mary University of London

Yahoo Labs Barcelona
January 19th 2012

	TREC-2		TREC-3		TREC-8		WT2G		Blog06	
	MAP	P@10	MAP	P@10	MAP	P@10	MAP	P@10	MAP	P@10
$LM_{Dir,\mu=2000}$	18.02	41.20	22.87	48.20	21.48	40.00	29.85	46.2	29.21	60.80
$LM_{JM,\lambda=0.7}$	14.70	32.4	20.80	40.20	21.81	39.4	23.11	33.80	21.04	45.60
$TF_{b=0.25,k1=1.2} \cdot IDF$	18.90	42.2	25.0	50.0	**22.39**	**40.60**	**31.76**	48.2	**30.46**	**63.8**
$TF_{TF=1} \cdot IDF$	09.19	17.00	11.53	22.00	11.20	09.40	14.00	15.20	05.51	11.80
$TF_{TF=tf\_d} \cdot IDF$	02.78	06.20	03.98	05.2	04.34	07.80	07.96	13.00	22.37	48.20
$BM25_{b=0.25,k1=1.2}$	**18.90**	**42.80**	**25.05**	**50.20**	22.3	40.2	31.41	**49.20**	30.27	63.40

See also: http://barcelona.research.yahoo.net/dokuwiki/doku.php?id=baselines

	TREC3 MAP	TREC8A MAP	TREC8B MAP	WT2G MAP
BM25	20.64	24.39	32.33	32.33
Tfidf				26.15
LM-JM				24.96
LM-Dir				30.87

Credits to Hany Azzam

What is our IR-driven mathematical framework (tool box) to investigate theoretically — to fully understand — why which model is better when?

Definition (Binomial Probability)

$$P_{\text{Binomial},N,p_t}(n_t) := \binom{N}{n_t} \cdot p_t^{n_t} \cdot (1-p_t)^{(N-n_t)} \tag{1}$$

$P($4 sunny days in a week (n=7) ≈ 0.2734 for $p_{\text{sunny}} = 45/90$
$P($4 "sunny" in d (dl=500) ≈ 0.00157 for $p_{\text{sunny}} = 1{,}000/1{,}000{,}000$

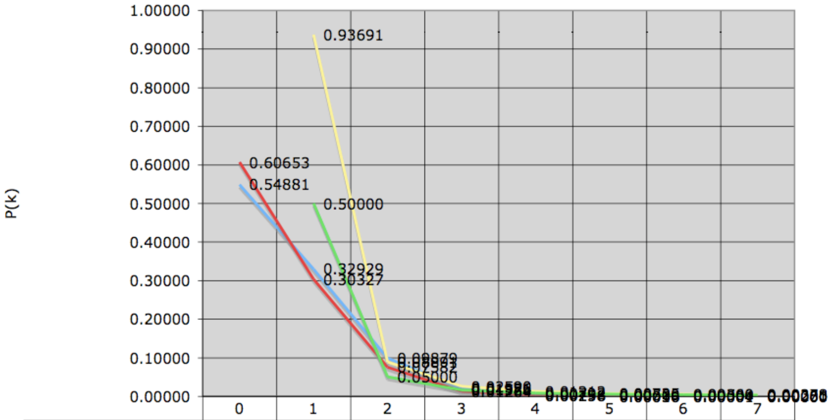
Definition (Poisson Probability)

$$P_{\text{Poisson},\lambda_t}(n_t) := \frac{\lambda_t^{n_t}}{n_t!} \cdot e^{-\lambda_t} \tag{2}$$

Definition (Independent Events)

$$P(e_1, \ldots, e_n|h) = \prod_{e_i} P(e_i|h) \tag{3}$$

The world according to Binomial/Poisson Prob and Independent Events



	0	1	2	3	4	5	6	7
Poisson Prob t_1	0.54881	0.32929	0.09879	0.01976	0.00296	0.00036	0.00004	0.00000
Poisson Prob t_2	0.60653	0.30327	0.07582	0.01264	0.00158	0.00016	0.00001	0.00000
semi-sub prob t_1		0.93691	0.08891	0.02590	0.01212	0.00725	0.00500	0.00378
semi-sub prob t_2		0.50000	0.05000	0.01581	0.00792	0.00500	0.00360	0.00281

TF-IDF

$$\mathrm{RSV}_{\text{TF-IDF}}(d,q,c) := \sum_t \mathrm{TF}(t,d) \cdot \mathrm{TF}(t,q) \cdot \mathrm{IDF}(t,c) \quad (4)$$

TF “normalisation"

$$\mathrm{TF}(t,d) := \frac{\mathrm{tf}_d}{\mathrm{tf}_d + k_1 \cdot \left(b \cdot \frac{\mathrm{dl}}{\mathrm{avgdl}} + (1-b)\right)} \qquad \text{Semantics?} \quad (5)$$

IDF “normalisation"

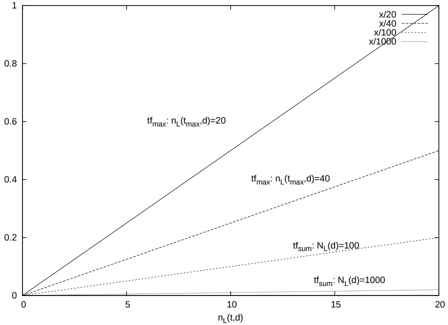
$$\mathrm{pidf}(t,c) := \frac{\mathrm{idf}(t,c)}{\mathrm{maxidf}} \qquad 0 \le \mathrm{pidf} \le 1 \qquad \text{Semantics?} \quad (6)$$

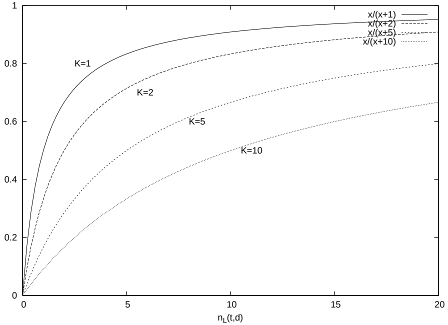
TF Variants

$$\mathrm{TF}(t,d) := \begin{cases} \mathrm{tf}_d & \text{total tf count} \\ \frac{\mathrm{tf}_d}{\mathrm{dl}} & P_{\mathrm{sum}}(t|d) \\ \frac{\mathrm{tf}_d}{\mathrm{maxtf}_d} & P_{\mathrm{max}}(t|d) \\ \frac{\mathrm{tf}_d}{\mathrm{tf}_d + K} & \text{parameter } K \propto \text{pivdl} \\ \frac{\mathrm{tf}_d}{\mathrm{tf}_d + k_1 \cdot \left(b \cdot \frac{\mathrm{dl}}{\mathrm{avgdl}} + (1-b)\right)} & K \text{ set in BM25-like way} \\ b + (1-b) \cdot \frac{\mathrm{tf}_d}{\mathrm{dl}} & \text{lifted tf; e.g. b=0.5} \\ \frac{\mathrm{tf}_d}{K} & \text{“pivoted" tf} \end{cases}$$

- [ ] total tf count
- [ ] $P_{\mathrm{sum}}(t|d)$
- [ ] $P_{\mathrm{max}}(t|d)$
- [ ] parameter $K \propto$ pivdl
- [ ] K set in BM25-like way
- [ ] lifted tf; e.g. b=0.5
- [ ] “pivoted" tf

TF Variants: Graphical Illustration



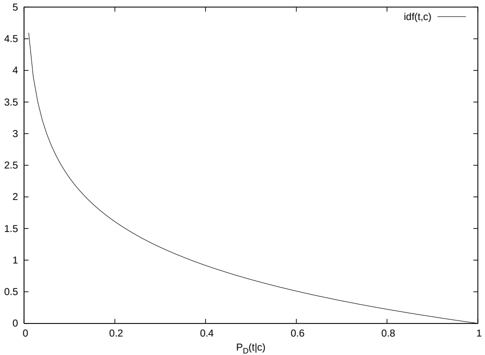


$$\frac{\mathrm{tf}_d}{\mathrm{dl}} \text{ and } \frac{\mathrm{tf}_d}{\mathrm{maxtf}_d}$$

$$\frac{\mathrm{tf}_d}{\mathrm{tf}_d + K}$$

IDF Variants

$$
\mathrm{IDF}(t,c) = \begin{cases}
-\log \frac{\mathrm{df}(t,c)}{N_D} & \text{is } -\log P_D(t|c) \quad \square \\
-\log \frac{\mathrm{df}(t,c)+0.5}{N_D+1} & \text{Laplace-like correction} \quad \square \\
-\log \frac{\mathrm{df}(t,c)}{N_D-\mathrm{df}(t,c)} & \text{BIR/BM25} \quad \square \\
-\log \frac{\mathrm{df}(t,c)+1}{N_D-\mathrm{df}(t,c)+0.5} & \text{RSJ/BM25} \quad \square
\end{cases}
$$



$P(q|d)$

$$P(q|d,c) = \prod_t P(t|d,c)^{\mathrm{TF}(t,q)}$$

$$\log P(q|d,c) = \sum_t \mathrm{TF}(t,q) \cdot \log\left(\lambda \cdot P(t|d) + (1-\lambda) \cdot P(t|c)\right)$$

TF-IDF and LM

$P(q|d)$: semantics of LM.
$P(d|q)$: ??? Semantics of TF-IDF???

Motivation

Before we engage with math to assign semantics to TF and IDF, the question is:

Why should we care?

Motivation continued

What people say (common beliefs):

- “We used *STANDARD* TF-IDF ..."
- “LM is $P(q|d)$ - good. TF-IDF is *HEURISTIC* - bad."
- “LM and BM25 are the main baselines; TF-IDF is out ..."
- “It’s clear why TF-IDF works; not clear why LM works."

What we would like to know (research challenges):

1. Can we improve (the retrieval quality of) existing models, or have we reached a ceiling?
2. Are there other models out there? One model per decade?

   VSM/TF-IDF mid 60s, probabilistic retrieval (BIR/RSJ weight) mid 70s, LSI and BM25 80s/90s, LM late 90s, FooBar 2010+ ???

Roger Penrose describes in the opening of his book “Shadows of the Mind" a scene where dad and daughter enter a cave.

- “Dad, that boulder at the entrance, if it comes down, we are locked in."
- “Well, it stood there the last 10,000 years, so it won't fall down just now."
- “Dad, will it fall down one day?"
- “Yes."
- “So it is more likely to fall down with every day it did not fall down?"

$$
\begin{aligned}
P(\text{boulder falls}) &\quad ?=? \quad n(\text{boulder fell})/N \\
P(\text{boulder falls}) &\quad ?=? \quad 1 - n(\text{boulder stood})/N \\
P(\text{x}) &\quad ?=? \quad n(\text{x})/N
\end{aligned}
$$

independent events

$$P(\text{information} \wedge \text{theory} \wedge \text{theory}) = P(\text{information}) \cdot P(\text{theory})^2$$

. . . . how about . . . .

multiple occurrence of same term: dependent events

$$P(\text{information} \wedge \text{theory} \wedge \text{theory}) = P(\text{information}) \cdot P(\text{theory})^{(2 \cdot \frac{2}{2+1})}$$

At roulette, you observe 1 $\times$ black followed by 17 $\times$ red.
Where do you place your tokens?

Pythagorean (a,b,c) triplets

(3, 4, 5), (5, 12, 13), (7, 24, 25), ...

$$a^2 + b^2 = c^2 \qquad 9 + 16 = 25$$

Fermat's last theorem

There are no three positive integers

$$a^n + b^n = c^n \qquad \text{for } n > 2$$

How long did it take to prove the theorem?

math4physics: Physics inspired math, math inspired physics.
math4IR: ???
Do we IR-ler have the "away-time" to engage with math4IR?

Definition

TF-IDF retrieval status value $\mathrm{RSV}_{\text{TF-IDF}}$:

$$\mathrm{RSV}_{\text{TF-IDF}}(d, q, c) := \sum_t w_{\text{TF-IDF}}(t, d, q, c) \tag{7}$$

Inserting the TF-IDF term weight yields the decomposed form:

$$\mathrm{RSV}_{\text{TF-IDF}}(d, q, c) = \sum_t \mathrm{TF}(t, d) \cdot \mathrm{TF}(t, q) \cdot \mathrm{IDF}(t, c) \tag{8}$$

What is the probabilistic semantics of

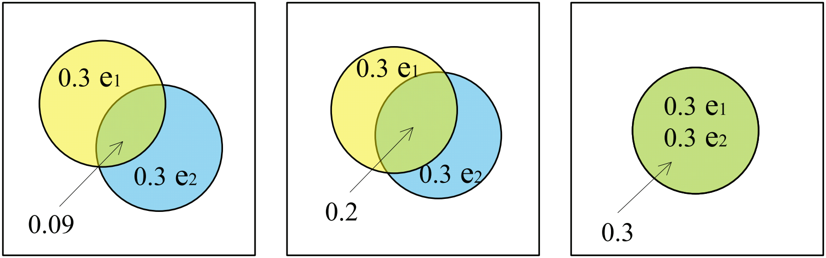
Definition

BM25 TF

$$\mathrm{TF}_{\mathrm{BM25}}(t,d) := \frac{\mathrm{tf}_d}{\mathrm{tf}_d + K_d} \tag{9}$$

$$K_d := k_1 \cdot \left(b \cdot \frac{\mathrm{dl}}{\mathrm{avgdl}} + (1-b) \right) \tag{10}$$

$\mathrm{pivdl} := \mathrm{dl} \,/\, \mathrm{avgdl}$.



independent semi-subsumed subsumed

Credits to Hengzhi Wu

Example

For the two events e_1 and e_2, the combined probabilities are:

$$
\begin{array}{ll}
0.3^2 = 0.09 & \text{independent} \\
0.3^{\left(2\cdot\frac{2}{2+1}\right)} \approx 0.2008 & \text{semi-subsumed} \\
0.3^1 & \text{subsumed}
\end{array}
$$

Independence-Subsumption Triangle

	independent	semi-subsumed	subsumed
1		$\frac{1}{2/2}$	
2	$\frac{2}{1}$	$\frac{2}{3/2}$	$\frac{2}{2}$
3	$\frac{3}{1}$	$\frac{3}{4/2}$	$\frac{3}{3}$
4	$\frac{4}{1}$ $\frac{4}{2}$	$\frac{4}{5/2}$	$\frac{4}{3}$ $\frac{4}{4}$
5	$\frac{5}{1}$ $\frac{5}{2}$	$\frac{5}{6/2}$	$\frac{5}{4}$ $\frac{5}{5}$
...	...	...	...
n	$\frac{n}{1}$ $\frac{n}{2}$ $\frac{n}{3}$	$\frac{n}{(n+1)/2}$	$\frac{n}{n-2}$ $\frac{n}{n-1}$ $\frac{n}{n}$

Note: Gaussian sum $1+2+...+n = n \cdot (n+1)/2$.

The story: Gauss as a school kid faced “time-spending" task by his teacher: add the numbers 1 to 100. Gauss answered within a minute: 5050. The famous formula: (1+100) + (2+99) + ... + (50+51) = 50 x 101.

Independence-Subsumption Triangle

d	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	
n							**a=2n/(n+d+1)**							n/a
1							**1.00**							1.00
2						**2.00**	**1.33**	**1.00**						1.50
3					**3.00**	2.00	**1.50**	1.20	**1.00**					2.00
4				**4.00**	2.67	2.00	**1.60**	1.33	1.14	**1.00**				2.50
5			**5.00**	3.33	2.50	2.00	**1.67**	1.43	1.25	1.11	**1.00**			3.00
6		**6.00**	4.00	3.00	2.40	2.00	**1.71**	1.50	1.33	1.20	1.09	**1.00**		3.50
7	**7.00**	4.67	3.50	2.80	2.33	2.00	**1.75**	1.56	1.40	1.27	1.17	1.08	**1.00**	4.00
N						…	2N/(N+d+1)	…						

Independence-Subsumption Triangle: embeds the BM25 TF into probability theory.

$$P(\text{theory} \wedge \text{theory}) = P(\textit{theory})^{(2 \cdot \text{TF}_{\text{BM25}})} = P(\textit{theory})^{(2 \cdot \frac{2}{2+1})} = P(\textit{theory})^{(1.33)}$$

	ind	semi-sub	sub
prob	2	1.33	1
0.001	0.000001	0.0001	0.001

What is the probabilistic semantics of

Definition

Probability of being informative (probabilistic idf):

$$\text{maxidf}(c) := -\log \frac{1}{N_D(c)} = \log N_D(c) \tag{11}$$

$$P(t \text{ informs}|c) := \text{pidf}(t, c) := \frac{\text{idf}(t, c)}{\text{maxidf}(c)} \tag{12}$$

Definition

BIR term weight w_{BIR}:

$$w_{\text{BIR}}(t, r, \bar{r}) := \log \frac{P(t|r)}{P(t|\bar{r})} \cdot \frac{P(\bar{t}|\bar{r})}{P(\bar{t}|r)} \tag{13}$$

A simplified form considers term presence only:

$$w_{\text{BIR,F1}}(t, r, \bar{r}) := \log \frac{P(t|r)}{P(t|\bar{r})} \tag{14}$$

$$\log w_{\text{BIR}}(t, r, \bar{r}) = \log \frac{P(t|r)}{1 - P(t|r)} - \log \frac{P(t|\bar{r})}{1 - P(t|\bar{r})} \approx -\log \frac{n_t}{N - n_t} \approx \text{IDF}(t, c)$$

Here, the log is a mathematical transformation; no information-theoretic or probabilistic meaning associated to IDF.

See also: [Croft and Harper, 1979], “prob models without relevance information"

Proof: Probability of Being Informative

Euler's number/convergence:

$$\lim_{N \to \infty} \left(1 - \frac{\lambda}{N}\right)^N = e^{-\lambda} \tag{15}$$

$\lambda := \mathrm{idf}(t, c)$, $N := \mathrm{maxidf}(c)$.

Theorem

Occurrence-Informativeness-Theorem: The probability that a term t occurs is equal to the probability that the term is not informative in maxidf *trials.*

$$P(t\ occurs|c) = (1 - P(t\ informs|c))^{\mathrm{maxidf}(c)} \tag{16}$$

Moreover, for the probability to be not informative:

$$1 - P(t\ informs|c) = \frac{\log n_D(t, c)}{\log N_D(c)} \tag{17}$$

Does this help to estimate $P(\text{boulder falls})$?

Definition

Poisson Bridge: Let x be a set of documents (e.g. the collection, set of relevant documents, set of retrieved documents).

$$\text{avgtf}(t,x) \cdot P_D(t|x) = \lambda(t,x) = \text{avgdl}(x) \cdot P_L(t|x) \tag{18}$$

Example

Poisson bridge: For a collection, let a term t (“sailing") occur in $n_L(t,\text{toy}) = 2,000$ of $N_L(\text{toy}) = 10^9$ *Locations*, and $n_D(t,\text{toy}) = 1,000$ of $N_D(\text{toy}) = 10^6$ *Documents*. The Poisson bridge is:

$$\frac{2,000}{1,000} \cdot \frac{1,000}{10^6} = \frac{2,000}{10^6} = \frac{10^9}{10^6} \cdot \frac{2,000}{10^9}$$

Note: Which averages are “useful"?

Credits to Theodora Tsikrika and Gabriella Kazai, “Notation in General Matrix Framework"

LM semantics: conventional

$$P(q|d,c) = \prod_t P(t|d,c) \qquad \text{Semantics for LM} \tag{19}$$

Conventional mixture:
$P(t|d,c) = \lambda \cdot P(t|d) + (1-\lambda) \cdot P(t)$

TF-IDF semantics: non-conventional

$$P(d|q,c) = \prod_t P(t|q,c) \qquad \text{Semantics for TF-IDF} \tag{20}$$

“Extreme" mixture:
$t \in q : P(t|q,c) = P(t|q)$, otherwise, $P(t|q,c) = P(t|c)$.

Total Probability

$$P(q|d) = \sum_t P(q|t) \cdot P(t|d) \tag{21}$$

$$P(d,q) = \sum_t P(d|t) \cdot P(q|t) \cdot P(t) \tag{22}$$

$$\frac{P(d,q)}{P(d) \cdot P(q)} = \sum_t P(t|d) \cdot P(t|q) \cdot \frac{1}{P(t)} \tag{23}$$

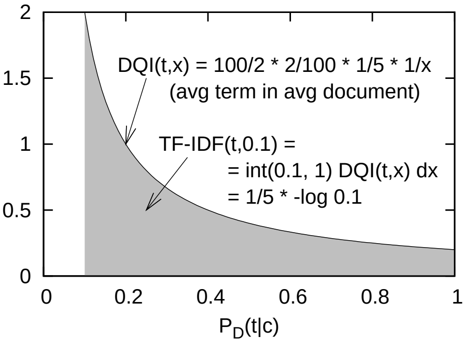
Relationship between total prob and TF-IDF??? And LM???
Option PIN's (probabilistic inference networks, [Turtle and Croft, 1990]):

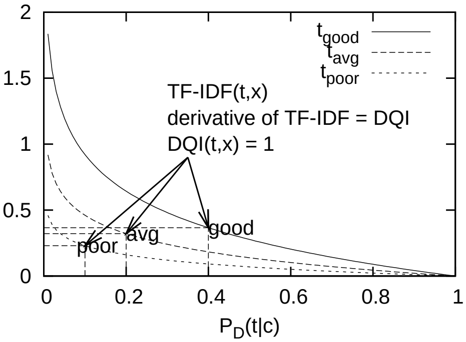
$$\sum_t \frac{P(q|t)}{\sum_{t'} P(q|t')} \cdot P(t|d) \propto \sum_t \frac{\mathrm{IDF}(t)}{\sum_{t'} \mathrm{IDF}(t')} \cdot \mathrm{TF}(t,d) \tag{24}$$

Indefinite and Definite Integral

$$\int \frac{1}{x}\, dx = \log x \tag{25}$$

$$\int_{P(t)}^{1} \frac{1}{x}\, dx = \log 1 - \log P(t) = -\log P(t) \tag{26}$$





Credits to Jun Wang

- TF
  - BM25 TF corresponds to semi-subsumed events
  - this relationship opens up pathways to new — IR-driven — probability theory, applicable in contexts beyond IR
- IDF
  - Poisson bridge: relates $P_D(t|c)$ (IDF) and $P_L(t|c)$ (LM): pathways to relate IDF/BIR to LM
  - normalisation pidf $= \text{idf}(t)/\text{maxidf}$: is sound
- $P(q|d)/P(q)$ and $P(d|q)/P(d)$: conjunctive
  - symmetric relationship between LM and TF-IDF
  - positions IR models; clarifies the $P(q|d)$ vs $P(r|d,q)$ issue
- $P(q|d)/P(q)$ and $P(d|q)/P(d)$: disjunctive
  - $\int \frac{1}{x}\,dx$: relationship between total prob and TF-IDF
- TF-IDF uncovered — TF-IDF is not heuristic anymore.

Outlook

- A unifying framework to derive all models from?
- A formal framework to prove ranking equivalences/differences?
- A “new" model?
- “New" math (probability theory) inspired by IR results but applicable in other domains?

Background: References

- IDF: deviation from Poisson, [Church and Gale, 1995]
- Information-theoretic explanation of TF-IDF, [Aizawa, 2003]
- Understanding IDF, [Robertson, 2004]
- Event Spaces, [Robertson, 2005]
- On Event Spaces and Rank Equivalences, [Luk, 2008]
- A Probabilistic Justification for TF-IDF, [Hiemstra, 2000]
- Understanding Relationships between Models, [Aly and Demeester, 2011]
- DFR, [Amati and van Rijsbergen, 2002]
- TF-IDF Uncovered, [Roelleke and Wang, 2008]
- Semi-subsumed Events: A Probabilistic Semantics of BM25 TF, [Wu and Roelleke, 2009]
- Probability of Being Informative, [Roelleke, 2003]
- Axiomatic Approach to IR Models, [Fang and Zhai, 2005]
- Bayesian extension to the language model for ad hoc information retrieval, [Zaragoza et al., 2003], ‘integral over model parameters"

Binomial Prob: $\rightarrow$ Poisson Prob $\rightarrow$ 2-Poisson Prob

- 2-Poisson is motivation for BM25 TF: [Robertson and Walker, 1994]

Event Spaces:

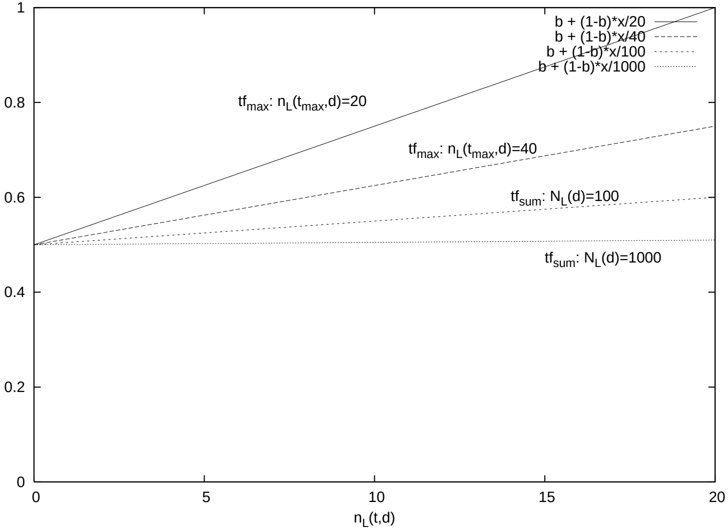
- $\{0,1\}$: BIR (and TF-IDF?)
- $\{0,1,2,\ldots\}$: Poisson (and TF-IDF?)
- $\{t_1,t_2,\ldots\}$: LM (and TF-IDF?)

Document-Query-(In)dependence: $\mathrm{DQI} = \frac{P(d,q)}{P(d)\cdot P(q)}$

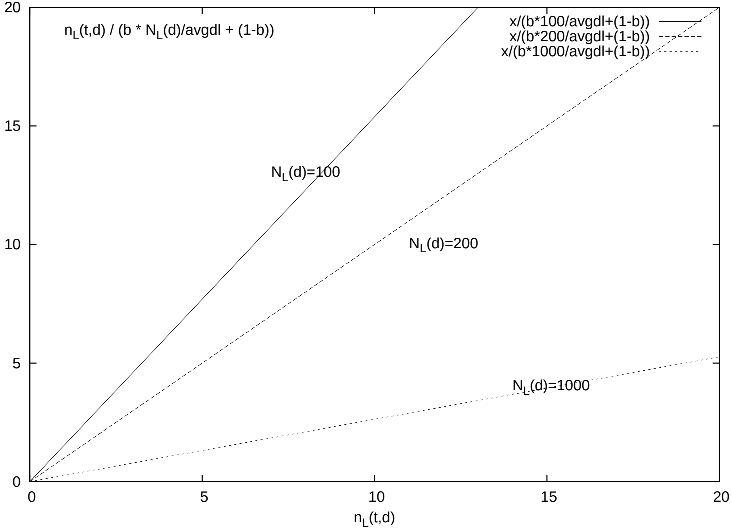
Burstinesss (avgtf): Given avgdl $= 100$. Given d.

- t_1: occurs in 1,000 locations, 500 docs. avgtf $= 2$. $\mathrm{tf}_d = 2$. Term is “average".
- t_2: occurs in 1,000 locations, 999 docs. avgtf ≈ 1. $\mathrm{tf}_d = 2$. Term is “good"; however: $\mathrm{IDF}(t_2) < \mathrm{IDF}(t_1)$.

tf_{lifted}: $b = 0.5$



$\mathrm{tf}_{\mathrm{pivoted}}$: $b = 0.7$, avgdl $= 200$



TF pivoted (in SIGIR BM25 tutorial by Hugo/Stephen, tf'_d)

Definition

TF pivoted

$$\mathrm{TF_{piv}}(t,d) := \frac{\mathrm{tf}_d}{K_d} \tag{27}$$

Move from BM25 TF to semi-subsumed in probability theory

$$2 \cdot \mathrm{TF_{BM25}} = 2 \cdot \frac{\mathrm{TF_{piv}}}{\mathrm{TF_{piv}} + 1}$$

$$P(\mathrm{theory} \wedge \mathrm{theory}) = P(\mathit{theory})^{(2 \cdot \mathrm{TF_{BM25}})}$$

References

Aizawa, A. (2003).
An information-theoretic perspective of tf-idf measures.
Information Processing and Management, 39:45–65.

Aly, R. and Demeester, T. (2011).
Towards a better understanding of the relationship between probabilistic models in ir.
In *Advances in Information Retrieval Theory: Third International Conference, Ictir 2011, Bertinoro, Italy, September 12-14, 2011, Proceedings*, volume 6931, pages 164–175. Springer-Verlag New York Inc.

Amati, G. and van Rijsbergen, C. J. (2002).
Probabilistic models of information retrieval based on measuring the divergence from randomness.
ACM Transaction on Information Systems (TOIS), 20(4):357–389.

Church, K. and Gale, W. (1995).
Inverse document frequency (idf): A measure of deviation from Poisson.
In *Proceedings of the Third Workshop on Very Large Corpora*, pages 121–130.

Croft, W. and Harper, D. (1979).
Using probabilistic models of document retrieval without relevance information.
Journal of Documentation, 35:285–295.

Fang, H. and Zhai, C. (2005).
An exploration of axiomatic approaches to information retrieval.
In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 480–487, New York, NY, USA. ACM.

References

Hiemstra, D. (2000).

A probabilistic justification for using tf.idf term weighting in information retrieval.

International Journal on Digital Libraries, 3(2):131–139.

Luk, R. W. P. (2008).

On event space and rank equivalence between probabilistic retrieval models.

Inf. Retr., 11(6):539–561.

Robertson, S. (2004).

Understanding inverse document frequency: On theoretical arguments for idf.

Journal of Documentation, 60:503–520.

Robertson, S. (2005).

On event spaces and probabilistic models in information retrieval.

Information Retrieval Journal, 8(2):319–329.

Robertson, S. E. and Walker, S. (1994).

Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval.

In Croft, W. B. and van Rijsbergen, C. J., editors, *Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 232–241, London, et al. Springer-Verlag.

Roelleke, T. (2003).

A frequency-based and a Poisson-based probability of being informative.

In *ACM SIGIR*, pages 227–234, Toronto, Canada.

References

Roelleke, T. and Wang, J. (2008).
TF-IDF uncovered: A study of theories and probabilities.
In *ACM SIGIR*, pages 435–442, Singapore.

Turtle, H. and Croft, W. B. (1990).
Inference networks for document retrieval.
In Vidick, J.-L., editor, *Proceedings of the 13th International Conference on Research and Development in Information Retrieval*, pages 1–24, New York. ACM.

Wu, H. and Roelleke, T. (2009).
Semi-subsumed events: A probabilistic semantics for the BM25 term frequency quantification.
In *ICTIR (International Conference on Theory in Information Retrieval)*. Springer.

Zaragoza, H., Hiemstra, D., and Tipping, M. (2003).
Bayesian extension to the language model for ad hoc information retrieval.
In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval*, pages 4–9, New York, NY, USA. ACM Press.