Temporal Information retrieval

Shrinivas Kane

Agenda

- Introduction
- Characteristics of temporal data
- Types of temporal information
- Temporal expressions
- Recent Research work
- Case study : Learning Recurrent Event Queries for Web Search
- My interests
- Future work
- References

Introduction

- What is Temporal Information retrieval?
- Why we need it?
- What is temporal data?
- From where temporal data can be extracted?

Characteristics of temporal data

- Well defined
  - Assuming we know start point and end points relation between two temporal data can be identified easily
- Easy to normalize
  - Temporal data can be normalized with reference to specific semantics in specified format.
- Easy to cluster
  - Hierarchical clustering is possible using level(s) of granularity.

Grouping temporal information

- Date
  - Exact Date
- Time
  - Exact Time
- Duration
  - Time period
- Set
  - Recurrence of event(s).

Temporal Expressions

- Implicate
  - Single or multiples values of this data can be mapped to single
- Explicit
  - Date or time specified explicitly
- Relative
  - Time information in context with other Implicate or explicit expression(s)

Recent research work

- Microblogging and real time search
- Temporal summaries
  - Provide snippet based results
- Temporal queries
  - Temporal constraint along with text query.

Case Study : Learning Recurrent Event Queries for Web Search

Problem

- Classifying queries as recurrent (REQ) and non recurrent(NON-REQ)
  - queries for event which occur at regular time interval
- Re-ranking documents based on temporal data

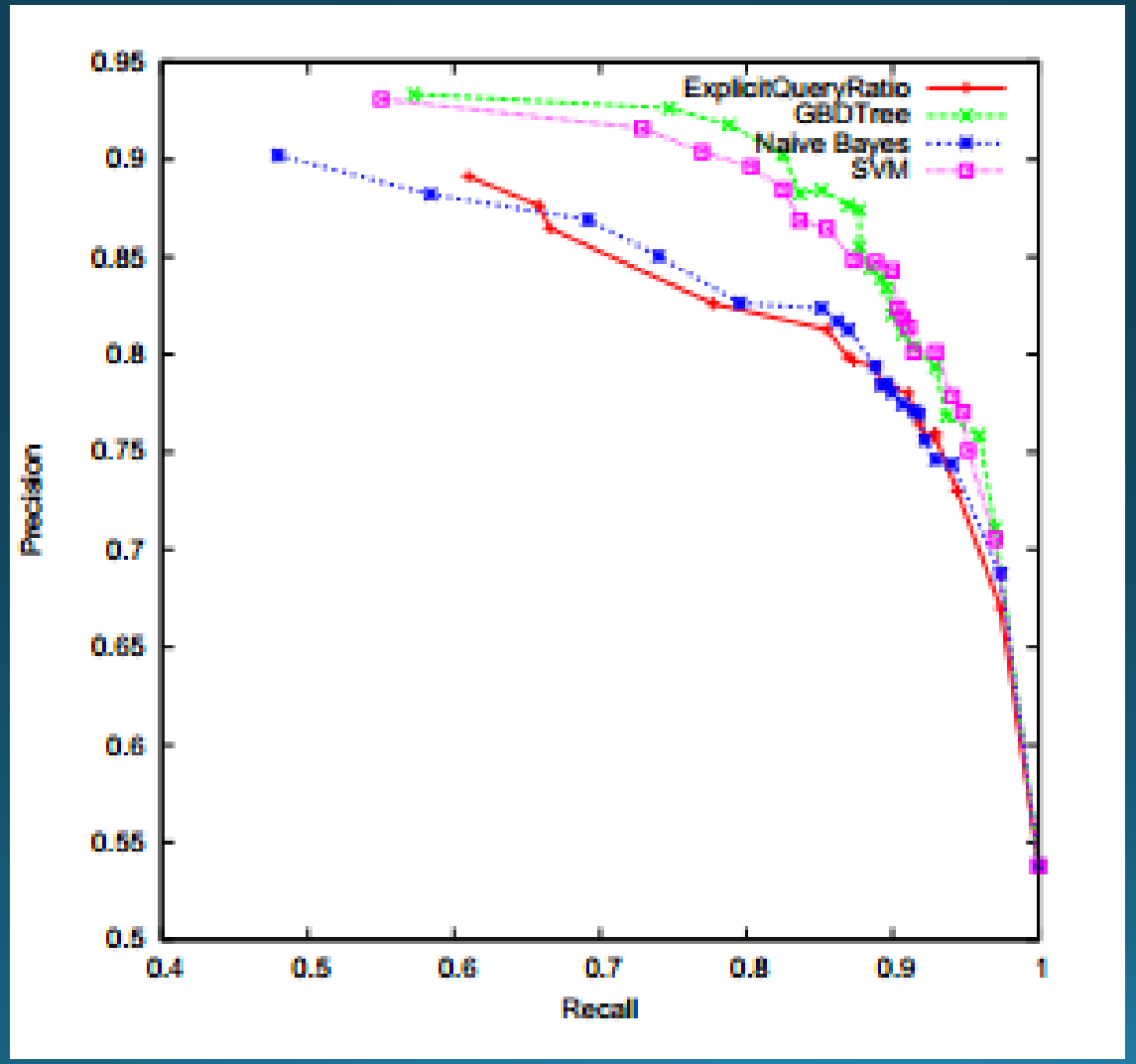
Feature generation

- Extracting time
  - Used Implicate, explicit and no timestamp queries
- Query log analysis
  - Explicit query ratio
- Query reformation
- Click log analysis
- Search engine result set
- Time series analysis
- Recurrent event(s) term list

Machine learning approach for Classification of queries

- Naive based method
  - This treats all features are independent
- SVM
- Gradient boosted decision tree algorithm

ML evaluation



Re-ranking documents

- Page score is boosted if query is REQ
- Ranking function
  - $F(q,d) = R(q,d) + [e(d_o,d_n)+k]e^{\lambda\acute{\alpha}\,(q)}$

    Where R(q,d) base scoring function

    $\lambda\acute{\alpha}\,(q)$ is confidence score of REQ query

    $[e(d_o,d_n)+k]e^{\lambda}$ is difference between oldest page and newest page

Ranking evaluation

- Evaluation metrics :
- Discounted cumulative gain at rank k is:

$$DCG@k = \sum_{i=1}^{k} \frac{(2^{r(i)}-1)}{\ln(1+1)}$$

- Where r(i) is relevance grade of i^{th} ranked document

		DCG@5			DCG@1		
bucket	#(query)	Organic	Our's	% over Organic	Organic	Ours	% over Organic
[0.0,0.1]	59	6.87	6.96	1.48(-2.3)	4.08	4.19	2.69(-1.07)
[0.1,0.2]	76	5.86	6.01	2.52(0.98)	2.88	2.91	1.14(1.69)
[0.2,0.3]	85	6.33	6.41	1.24(2.12)	3.7	3.7	0.0(0.8)
[0.3,0.4]	75	5.18	5.24	1.18(-0.7)	2.92	2.95	1.14(1.37)
[0.4,0.5]	78	4.96	4.82	-2.84(-1.35)	2.5	2.42	-3.06(0)
[0.5,0.6]	84	5.4	5.37	-0.45(-0.3)	2.82	2.85	1.05(-1.5)
[0.6,0.7]	78	4.78)	5.19)	8.42(3.64)	2.56	2.83	10.75(4.1)
[0.7,0.8]	80	4.45	4.60	3.41(3.19)	2.21	2.26	1.98(2.8)
[0.8,0.9]	78	4.81	4.96	3.15(4.79)	2.32	2.33	0.55(0.65)
[0.9,1.0]	107	5.08	5.50	**8.41***(4.41)	2.64	3.09	**16.78***(1.36)
[0.0,1.0]	800	5.33	5.47	**2.74***(2.17)	2.83	2.93	**3.6***(1.26)

Table 5: REQ learner improves search engine organic results. The numbers in the brackets are by Zhang's methods. Direct comparison with Zhang's method is valid only in the last line, using all queries. A sign "∗" indicates statistical significance (p-value<0.05)

My interests

- Using machine learning approach to optimize index cache based on temporal query analysis
- Plotting Document improvements on timeline

Future Work

- How to provide timeline?
- How to calculate lifespan of event?
- Detecting trending events

References

- Ruiqiang Zhang ,Yuki Konda ." Learning Recurrent Event Queries for Web Search". Oct 11, 2010 - cO2010 Association for Computational Linguistics
- James Allan, Rahul Gupta. "Temporal summaries of new topics" 2001 – SIGIR
- Omar Alonso , Jannik Strotgen. " Temporal information retrieval challenges and opportunities " 2011 – TWAW
- J F Allen. "Maintaining knowledge about temporal intervals" communications of ACM 1983
- F. Schilder and C Habel "From temporal expressions to temporal information" 2001-TASIP