A Formalization of Logical Imaging for IR using QT

Guido Zuccon, Leif Azzopardi, Keith van Rijsbergen

September 1st, 2008

guido@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk/∼guido

Information Retrieval Group
University of Glasgow

How could QT be useful to IR?

- In the last years there have been some attempts to model classical systems using QT:
  - Cognitive processes such as concepts formation and combination [AG05]
  - Modeling of semantics [BKNM08, BW08, CCS08]
  - Modeling IR processes and techniques [vR04, AvR07, HRAvR08]
- QT trait d'union between logics, probability and geometry
- QT naturally models the contextual behaviour of complex systems [Kit08]
- Inspires the invention of new models and techniques

How IR can be translated in QT

Correspondences between concepts in IR and QT:

- Information space $\rightarrow$ Hilbert space $\mathcal{H}$
- Document d $\rightarrow$ vector $|\underline{d}\rangle$
- Relevance $\rightarrow$ an observable, that is a Hermitian operator ($\mathbf{R}$)

The value of an observable is given by the eigenvalues of the operator (always real).
A state of the Quantum system corresponds to a vector; we can associate a density operator to each state, e.g. $\rho_{|\underline{x}\rangle} = |\underline{x}\rangle\langle\underline{x}|$

In this section:

A formalization of Logical Imaging using QT

- What is Logical Imaging?
- Our formalization of Logical Imaging in QT
- The Kinematics Operator
- Considerations and Future directions

Motivation for using Logical Imaging in IR

- How do we estimate the probability of relevance of a document given a query? $P(R|q,d) \sim P(d \rightarrow q)$
- *LUP by van Rijsbergen*:

  > *"Given any two sentences x and y; a measure of the uncertainty of $y \rightarrow x$ relative to a given data set, is determined by the MINIMAL EXTENT to which we have to add information to the data set, to establish the truth of $y \rightarrow x$."[van89]*

- Logical Imaging (LI) addresses such minimal belief revision

Logical Imaging: a technique for belief revision

Then, what is LI?

- LI is a technique for belief revision, based on Possible Worlds Semantics (PWS)
- Introduced by Stalnaker [HSP81], extended and refined by Lewis [Lew73, Lew76], investigated by Gärdenfors [Gär88], applied to IR by Crestani [CvR95a, CvR95b] and Amati [AK92]

The intuition behind LI

- A probability is assigned to each term,
- There is a measure of similarity between terms (e.g. EMIM, cosine, NGD, ...)
- Probability mass is moved between similar terms: this generates a *kinematics* of probabilities

A first LI model

The LI model for IR by Crestani [CvR95a, CvR95b]:

- Each term has a probability
- For every document in the collection, we evaluate $P(d \rightarrow q)$ by imaging on the document
- Uses EMIM as similarity measure

The kinematics:

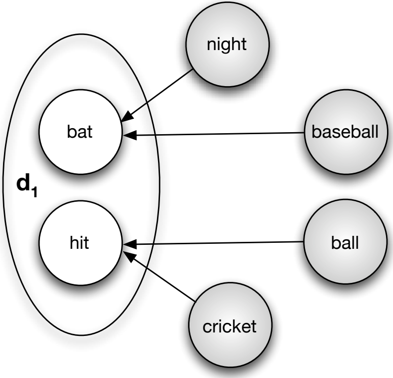
- The probabilities associated to a *term* NOT IN the *document* are transfered to the *most similar term* IN the *document*

Formal definition of the model (1/2)

A formal definition of Crestani's model:

- $|T|$ terms t_i, probability $P(t_i)$ associated, $\sum_{t_i \in T} P(t_i) = 1$
- $P(d \rightarrow q) = \sum_T P(t_i)\tau(t_d, q)$
- t_d is the most similar term to t_i for which d is true; τ is a truth function

Formal definition of the model (2/2)



Variations on the theme

The variations affect the kinematics of probabilities:

- Transfer of probabilities towards more than a single term in the document:
  - proportional
  - opinionated
- The transfer leaves a small probability mass associated to the term not in the document: Jeffrey's conditionalization

Logical Imaging in QT

- Represent a term t_i with a vector of document

$$\text{e.g. } |\underline{t}_i\rangle = [2\ 0\ 1]^T$$

- $P(t)$ be a probability distribution over the set T of terms (thus, $\sum_{i_1}^{k} P(t_i) = 1$ and $0 \leq P(t_i) \leq 1$)

Note: We use the Dirac notation, where the following correspondences hold:

column vector $\underline{x} \longrightarrow |\underline{x}\rangle$
inner product $\underline{y} \cdot \underline{x} \longrightarrow \langle \underline{y} | \underline{x} \rangle$

row vector $\underline{x}^* \longrightarrow \langle \underline{x}|$
outer product $\underline{y} \times \underline{x} \longrightarrow |\underline{y}\rangle\langle \underline{x}|$

Logical Imaging in QT

- A document d is represented by $|\underline{d}\rangle$, a query q by $|\underline{q}\rangle$.

  e.g. $|\underline{d}_i\rangle = \sum_{t_i \in W_d} \lambda_i |\underline{t}_i\rangle$

  where W_d is the set of terms which are present in d
- Projector associated to document: $\mathbf{P}_d = \vee_{t_i \in d} |\underline{t_i}\rangle\langle\underline{t_i}|$
- Projector associated to query: $\mathbf{P}_d = \vee_{t_i \in q} |\underline{t_i}\rangle\langle\underline{t_i}|$

Density operator and its evolution

- We can associate a density operator to document d:

$$\rho_{\underline{d}} = \sum_{t_i \in d} \alpha_i (\sum_{j=1}^{n} \lambda_{i,j}^2 |\underline{e}_j\rangle\langle\underline{e}_j|)$$

- Thus, the density operator associated to d after imaging is:

$$\rho'_{\underline{d}} = \sum_{t_i \in W_d} \alpha'_i (\sum_{j=1}^{n} {\lambda_{i,j}}^2 \left|\underline{e}_j\right\rangle \left\langle\underline{e}_j\right|),$$

Note: To simplify the notation, we assume $P(t_i) = \alpha_i$

Computing $P(d \rightarrow q)$

We can now compute the probability $P(d \rightarrow q)$ after Imaging, or, more precisely, the probability induced by $[\![\mathbf{P}_d \rightarrow \mathbf{P}_q]\!]$.
In order to do so we make use of the Gleason's Theorem:

> *"Let $\mathcal{H}$ be a Hilbert space with dimension n (finite or countably infinite) larger than 2. Let P be a probability measure defined upon the closed subspaces of $\mathcal{H}$ that is additive for a finite or infinite number of mutually orthogonal subspaces. Then there exists a trace–class positive operator ρ with unit trace such that $P(\mathcal{L}) = tr(\rho \mathbf{P}_{\mathcal{L}})$, $\mathbf{P}_{\mathcal{L}}$ being the projector upon a closed subspace $\mathcal{L}$."*

Computing $P(d \rightarrow q)$

Thanks to Gleason's Theorem, $P(d \rightarrow q) = tr(\rho'_{\underline{d}} \mathbf{P}_{\underline{R}})$, where $\mathbf{P}_{\underline{R}}$ is the projector associated to $[\![\mathbf{P}_d \rightarrow \mathbf{P}_q]\!]$

Note, in the case that $\mathbf{P}_d$ and $\mathbf{P}_q$ are not compatible (they do not commute, $\mathbf{P}_d \mathbf{P}_q \neq \mathbf{P}_q \mathbf{P}_d$):

$$\mathbf{P}_R = \mathbf{I} - \lim_{n\rightarrow\infty}((\mathbf{I} - \mathbf{P}_d)(\mathbf{I} - \lim_{n\rightarrow\infty} \mathbf{P}_d \mathbf{P}_q \mathbf{P}_d)^n (\mathbf{I} - \mathbf{P}_d))^n$$

The Kinematics Operator (1/2)

How do we move the probabilities?

- Let us assume that the probability associated to each term is stored in an entry of a diagonal matrix, eg:

$$\begin{bmatrix} 0.3 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.4 \end{bmatrix}$$

  and $t_1 \longrightarrow t_3$

- Applying transformation $\mathbf{A}' = \mathbf{K}^T\mathbf{A}\mathbf{K}$ with $\mathbf{K} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ ...

The Kinematics Operator (2/2)

- ...we obtain

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0.4 \end{bmatrix}$$

  remember $t_1 \longrightarrow t_3$

- The Kinematics operator corresponds to the evolution operator **U** of the Schrödinger Picture.
- The evolution of the density operator **D** from time t_1 to time t_2 is governed by the equation:

$$\mathbf{D}(t_2) = \mathbf{U}^\dagger(t_2)\mathbf{D}(t_1)\mathbf{U}(t_2)$$

Recap

- LI moves probabilities from terms not in the document to similar terms into the document
- QT is capable to model LI and its kinematics, by means of a particular operator, the Kinematics Operator

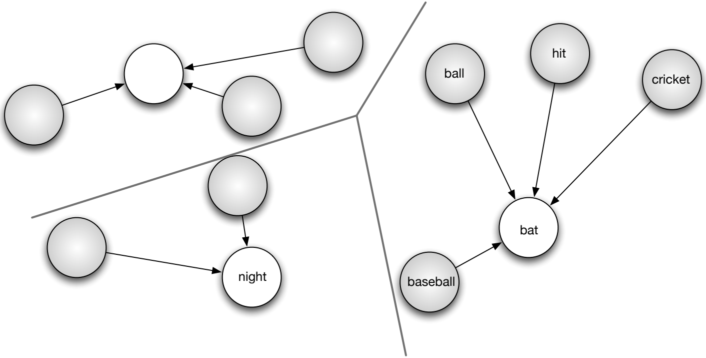
Considerations and Future Directions (1/2)

- This formulation of LI is not effective: MAP and P@R are lower than baselines values obtained with simple weightings as TF–IDF or IDF on standard TREC collections (e.g. WSJ, AP).
- The current formulation of LI is more effective than simple baselines on small collections, e.g. CACM, NPL.
- The LI theory is soundness, but the current formulation is poor (higher rank to documents that contain not discriminative information)

Considerations and Future Directions (2/2)

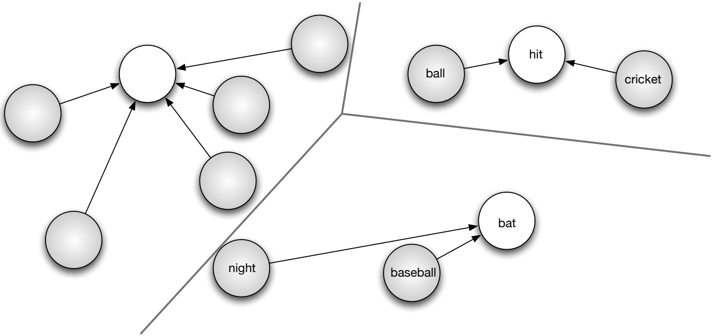
- We are investigating a reformulation of LI considering an intra-document kinematics
- We also want to take into account the context where a term appears: Context Based Logical Imaging (CBLI) using Voronoi Tessellation in the QT framework

A Voronoi Tessellation of a 2D space, and its use to guide the Kinematics



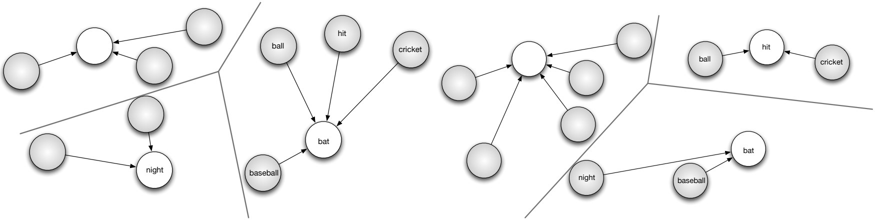
Voronoi Tessellation generated by $d = \{night, bat, \ldots\}$

A Voronoi Tessellation of a 2D space, and its use to guide the Kinematics



Voronoi Tessellation generated by $d = \{bat, hit, \ldots\}$

A Voronoi Tessellation of a 2D space, and its use to guide the Kinematics



The difference in the Kinematics produced by two different Voronoi Tessellations

Bibliography I

[AG05] D. Aerts and L. Gabora.
A Theory of Concepts and their Combinations I.
Kybernetes, 34:151–175, 2005.

[AK92] G. Amati and S. Kerpedjev.
An Information Retrieval Model: Implementation and Experiments.
Technical Report 5B04892, Fondazione Ugo Bordoni, Roma, IT, 1992.

[AvR07] S. Arafat and C. J. van Rijsbergen.
Quantum Theory and the Nature of Search.
In *Proc. of the 1st AAAI Quantum Interaction Symposium*, pages 114–121, Stanford, CA, USA, 2007.

[BKNM08] P. Bruza, K. Kitto, D. Nelson, and K. McEvoy.
Entangling Words and Meaning.
In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 118–124, Oxford, UK, 2008.

Bibliography II

[BW08] P. Bruza and J. Woods.
Quantum Collapse in Semantic Space: Interpreting Natural Language Argumentation.
In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 141–147, Oxford, UK, 2008.

[CCS08] S. Clark, B. Coecke, and M. Sadrzadeh.
A Compositional Distributional Model of Meaning.
In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 133–140, Oxford, UK, 2008.

[CvR95a] F. Crestani and C. J. van Rijsbergen.
Information Retrieval by Logical Imaging.
Journal of Documentation, 51(1):3–17, Mar 1995.

[CvR95b] F. Crestani and C. J. van Rijsbergen.
Probability Kinematics in Information Retrieval.
Proc. ACM SIGIR '95, pages 291–299, Jan 1995.

Bibliography III

[Gär88] P. Gärdenfors.
Knowledge in Flux. Modeling the Dynamics of Epistemic States.
The MIT Press, Cambridge, MA, USA, 1988.

[HRAvR08] A. Huertas-Rosero, L. A. Azzopardi, and C.J. van Rijsbergen.
Characterising Through Erasing – A Theoretical Framework for Representing Documents Inspired by Quantum Theory.
In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 160–163, Oxford, UK, 2008.

[HSP81] W. L. Harper, R. Stalnaker, and G. Pearce, editors.
Ifs. Conditionals, Belief, Decision, Chance and Time.
Reidel, Dordrecht, NL, 1981.

[Kit08] K. Kitto.
Why Quantum Theory?
In *Proc. of the 2nd AAAI Quantum Interaction Symposium*, pages 11–18, Oxford, UK, 2008.

Bibliography IV

[Lew73] D. K. Lewis.
Counterfactuals.
Basil Blackwell, Oxford, UK, 1973.

[Lew76] D. K. Lewis.
Probabilities of Conditionals and Conditional Probabilities.
The Philosophical Review, 85:297–315, 1976.

[van89] C. J. van Rijsbergen.
Towards an Information Logic.
In *Proc. ACM SIGIR '89*, pages 77–86, Cambridge, MA, USA, 1989.

[vR04] C. J. van Rijsbergen.
The Geometry of Information Retrieval.
CUP, 2004.

Questions?

- Thanks for the attention
- Any question?
- For further discussion do not hesitate to contact me on email (guido@dcs.gla.ac.uk)