Conditional probabilistic reasoning without conditional logic

Fabrizio Sebastiani
Istituto di Elaborazione dell'Informazione
Consiglio Nazionale delle Ricerche
Via S. Maria, 46 - 56126 Pisa (Italy)
E-mail: fabrizio@iei.pi.cnr.it

Abstract

Imaging is a class of non-Bayesian methods for the revision of probability density functions originally proposed as a semantics for conditional logic. Two of these revision functions, Standard Imaging and General Imaging, have successfully been applied to modelling information retrieval (IR). Due to the problematic nature of a"direct" implementation of Imaging revision functions, we propose their alternative implementation by representing the semantic structure that underlies them, in the language of a probabilistic (Bayesian) logic. Recasting these models of information retrieval in such a general-purpose knowledge representation (KR) tool, besides showing the potential of this "Bayesian" tool for the representation of non-Bayesian revision functions, paves the way to a possible integration of these models with other more KR-oriented models of IR, and to the exploitation of general purpose domain-knowledge.

1 Introduction

Researchers have recently devoted an increasing effort to the specification of models of information retrieval (IR) along the so-called *logical approach* [1]. Although there are various interpretations of this approach, by and large we may take it to say that the relevance of documents to user queries may be viewed in terms of the validity of the formula $d \rightarrow q$ of a logical language, where d is a formula representing the document, q is a formula representing the query and "$\rightarrow$" is the conditional ("implication") connective of the chosen logic[a]. However, the impossibility of finding *perfect* (i.e. absolutely faithful) representations of the information content of documents and queries calls for a probabilistic treatment of this conditional sentence: researchers agree that a realistic approach to the IR problem must rather rely on the evaluation of the real-valued term $P(d \rightarrow q)$, where $P(\alpha)$ stands for "the probability that α".

A number of researchers have recently taken up these ideas, and proposed logics and logic-based models of IR based on them. Among these, of particular

[a] For a discussion why we think that *validity*, rather than *truth*, of $d \rightarrow q$ is the notion to consider, see [2].

interest to the present paper are the models of IR based on "Imaging" [3] (hereafter called *Standard* Imaging) and "General Imaging" [4] by Crestani and van Rijsbergen [5,6,7]. Standard and General Imaging are *density revision functions* (DRFs – see Section 2) originally proposed as a semantics for *conditional logic*, the branch of logic that addresses the "if ... then ..." notion of natural language. The experimental results presented in [5,6] show a definite improvement of performance over standard approaches to IR, thus supporting the conjecture that Imaging methods capture some fundamental intuition underlying IR.

A full-blown implementation of Imaging methods is, unfortunately, problematic. The reason is that implementation techniques for DRFs (of which belief networks are a primary example) have so far concentrated on the Bayesian case, i.e. *Bayesian conditionalisation.* To our knowledge, no technique has been developed yet for non-Bayesian DRFs such as Imaging, and no theorem proving technique has been developed for Imaging-based conditional logics. In this paper we propose an alternative method for implementing Imaging methods. Essentially, the idea is to represent the semantic structure that underlies Imaging-based conditional logics in the language of a probabilistic (Bayesian!) logic. This process of *abstraction* (i.e. of transfer from the realm of semantics to that of syntax) is conceptually not dissimilar from the so-called "standard translation" (see e.g. [8]) of modal propositional logic into first order logic (FOL), whereby modal propositional reasoning is reduced to FOL reasoning by simulating within FOL the possible worlds semantics of modal propositional logic.

We have shown that Halpern's $\mathcal{L}_1$ [9] logic, a FOL extended with features for *objective* probability, is powerful enough to accommodate Standard and General Imaging, but also generalizations of them such as "Proportional Imaging". For reasons of space, in this paper we confine our discussion to Standard Imaging only; in an extended version of this paper [10], besides dealing with the General and Proportional Imaging cases, we also show that an extension of $\mathcal{L}_1$ with features for *subjective* probability (called $\mathcal{L}_3$) can further accommodate "Jeffrey Imaging", a variant of Imaging obtained by combining (any variant of) Imaging and *Jeffrey Conditionalisation* [11], which seems a promising tool for the analysis of non-binary "relevance feedback" in IR [12]. Our implementation of Imaging (and variations thereof) on top of $\mathcal{L}_1$ shows then that Bayesian revision tools can be seen as convenient and powerful toolkits for fast prototyping of non-Bayesian models of IR. Quite obviously, recasting these models of IR in such a general purpose knowledge representation (KR) and reasoning tool paves the way to a possible integration of these models with other more KR-oriented models of IR (such as e.g. [13,14]), and to the exploitation of general-purpose domain-knowledge.

The paper is organised as follows. While in Section 2 we briefly review

Standard Imaging, in Section 3 we look at the main features of the $\mathcal{L}_1$ logic, the main tool we will use in this work. In Section 4 we show a $\mathcal{L}_1$ implementation of a model of IR based on Standard Imaging. Section 5 discusses both some theoretical underpinnings and the practical consequences of our work by comparing it with related work.

2 The Bayesian model of epistemic states

The notion of Imaging (together with its variations) assumes that the epistemic state of a cognitive agent is represented by a (subjective) *probability function* P defined on the set of sentences of a language L and that complies with the standard axioms of probability. If A is a sentence of L, then $P(A)$ is meant to represent the *degree of confidence* (or certainty, or belief) that the agent has in the truth of A: if $P(A) = 1$ the agent is certain of the truth of A, if $P(A) = 0$ the agent is certain that A is false, while if $0 < P(A) < 1$ the agent is unsure whether A is the case or not. From now on, we will take L to be the language of propositional logic defined on a finite number of propositional letters.

In real life agents may change their mind as a result of the acquisition of new evidence; e.g. the agent may come to believe true facts that she believed probably false. In order to model this, one needs a mechanism to change the probability value associated to a sentence and change the values of other semantically related sentences accordingly. This is called a probability *revision* function. The standard probability revision function is *Bayesian conditionalisation*, according to which if an agent comes to firmly believe in the truth of a sentence A (which she believed at least possibly true — i.e. $P(A) > 0$), her new epistemic state must be described by a new probability function $P(-|A)$ (also indicated as P_A) which, for all sentences B is defined as:

$$P_A(B) \stackrel{def}{=} P(B|A) \stackrel{def}{=} \frac{P(A \wedge B)}{P(A)} \tag{1}$$

Note that P_A is such that A is (correctly) deemed true, i.e. $P_A(A) = 1$. An alternative, semantically oriented but equivalent way of characterizing epistemic states is to assume that there is a *density function* (also called a probability *distribution*) μ on the set W of the 2^n *possible worlds* (or simply *worlds*) on which the propositional language is interpreted (where n is the number of propositional letters in the language)[b]; i.e. μ is such that $\sum_{\{w \in W\}} \mu(w) = 1$. The

[b]This characterisation of possible worlds should not be confused with the Hintikka-Kripke notion, according to which for giving semantics to modal logic we group possible worlds in so-called "Kripke structures", and each of them is considered "possible" by an explicitly (rather than implicitly) represented cognitive agent. This latter notion implies a notion of

degree of confidence of the agent in sentence A is defined as the sum of the probabilities of the worlds that satisfy A ("A-worlds"), i.e.

$$P(A) \stackrel{def}{=} \sum_{\{w \in W \ | \ w \models A\}} \mu(w) \tag{2}$$

In the possible worlds view, instead of specifying a probability revision function one specifies a *density revision function* (i.e. a functions mapping a density function into another density function) on possible worlds that induces the desired probability revision function through (2). Viewed as a DRF, Bayesian conditionalisation then amounts to eliminating from consideration the worlds that do not satisfy A ("$\neg A$-worlds"), and creating a new density function μ' obtained from μ by redistributing to the A-worlds the probability originally assigned to the $\neg A$-worlds, where the redistribution is proportional to the probability originally assigned to the A-worlds:

$$\mu'(w) = \begin{cases} \mu(w) \cdot (1 + \frac{P(\neg A)}{P(A)}) & \text{if } w \models A \\ 0 & \text{if } w \not\models A \end{cases} \tag{3}$$

Imaging and its variations are DRFs alternative to Bayesian conditionalisation. They differ from it in that they are based on the idea that the probability of $\neg A$-worlds is *not* redistributed proportionally to the original probability of the A-worlds. The underlying assumption is that there is a measure S of *similarity* defined on W such that $0 \leq S(w, w') \leq 1$ measures, for every pair $\langle w, w' \rangle \in W^2$, how similar to w w' is (the higher $S(w, w')$, the more similar to w w' is). According to Imaging DRFs, only worlds sufficiently similar to the $\neg A$-worlds receive some probability; how similar they need to be is what differentiates the various forms of Imaging.

Standard Imaging (first introduced in [3]) is based on the assumption that, for all satisfiable sentences A and for all $\neg A$-worlds w, a most similar A-world $w' = \sigma(A, w) \stackrel{def}{=} max\{S(w, w') \mid w' \models A\}$ always exists and is unique; it is to w' that the probability $\mu(w)$ is transferred. Imaging thus sanctions that:

$$\mu'(w') = \begin{cases} 0 & \text{if } w' \not\models A \\ \mu(w') + \sum_{\{w \in W \ | \ w' = \sigma(A, w)\}} \mu(w) & \text{if } w' \models A \end{cases} \tag{4}$$

belief located in the object language (with explicit "$\mathbf{Bel}_i$" operators, where "$\mathbf{Bel}_i(\alpha)$" means "agent i believes that α"), rather than in the metalanguage as the view we discuss.

Obviously, the results of applying Imaging depend on the choice of the S function. In general, however, for no choice of the similarity function the results of Bayesian conditionalisation coincide with those of Imaging.

3 The $\mathcal{L}_1$ probabilistic logic

The $\mathcal{L}_1$ probabilistic logic is a FOL for reasoning about (objective) probabilities [9]. Probability values can explicitly be mentioned in the language: rather than mapping non-probabilistic formulae on the real interval $[0, 1]$, probabilistic formulae are mapped on the standard truth values *true* and *false*. The logic allows the expression of real-valued terms of type $w_{\langle x_1,\ldots,x_n\rangle}(\alpha)$ (where α is any $\mathcal{L}_1$ formula), with the meaning "the probability that random individuals $x_1, \ldots, x_n$ verify α". It also allows their comparison by means of standard numerical binary operators, resulting in formulae that can be composed by the standard sentential operators of FOL. The semantics of $\mathcal{L}_1$ is given by assuming the existence of a discrete probability structure on the domain; a formula such as $w_{\langle x_1,\ldots,x_n\rangle}(\alpha) \geq r$ is true in an interpretation iff the probability assigned to the individuals that verify α sums up to at least r[c].

The semantics of $\mathcal{L}_1$ can be specified by means of *type 1 probabilistic structures* (PS_1), i.e. triples $M = \langle D, \pi, \mu\rangle$, where D is a domain of individuals, π is an assignment of n-ary relations on D to n-ary predicate symbols and of n-ary functions on D to n-ary function symbols ($\langle D, \pi\rangle$ is then a FOL interpretation), and μ is a discrete density function (DDF) on D. The numerical value $\mu(d)$ may be interpreted as "the probability that, if a random individual has been picked from the domain D, it is d". In what follows, we will use $\mu(D')$ (where $D' \subseteq D$) as a shorthand for $\sum_{d\in D'} \mu(d)$. Also, given a DDF μ on D, μ^n is defined as the DDF on D^n such that $\mu^n(\langle d_1, \ldots, d_n\rangle) = \mu(d_1) \times \ldots \times \mu(d_n)$.

A *valuation* is a mapping v of object variables (i.e. variables denoting individuals of the domain, indicated by the subscript ^{o}) into D and numerical variables (i.e. variables denoting real numbers, indicated by the subscript ^{c}) into $\mathcal{R}$. Three semantic notions can now be defined:

- the *(numerical) value* $\|t^c\|_{\langle M,v\rangle}$ of a numerical term t^c in $\langle M, v\rangle$, with values in the real interval $[0, 1]$;
- the *(object) value* $[t^o]_{\langle M,v\rangle}$ of an object term t^o in $\langle M, v\rangle$, with values in D;

[c] It follows that, if x does not occur free in α, the term $w_{\langle x\rangle}(\alpha)$ may evaluate to 0 or 1 only, depending whether α evaluates to *false* or *true*, respectively. Given a closed formula α, the term $w_{\langle x\rangle}(\alpha)$ plays then the role of its characteristic function.

- the *truth* $\langle M, v\rangle \models \alpha$ of a formula α in $\langle M, v\rangle$, with values in $\{true, false\}$.

The semantics of the logic is more formally described by the semantic clauses that follow. In these, "*mathop*" is an operator in the set $MATHOP=\{+,-,\cdot,\div\}$, and "*relop*" is an operator in the set $RELOP=\{=,\neq,\geq,\leq,<,>\}$; **mathop** and **relop** are the corresponding operations on real numbers.

$$
\begin{array}{rcl}
[x^o]_{\langle M,v\rangle} & = & v(x^o) \\
[f_i^n(t_1^o,\ldots,t_n^o)]_{\langle M,v\rangle} & = & \pi(f_i^n)([t_1^o]_{\langle M,v\rangle},\ldots,[t_n^o]_{\langle M,v\rangle}) \\
\|x^c\|_{\langle M,v\rangle} & = & v(x^c) \\
\|k^c\|_{\langle M,v\rangle} & = & \mathbf{k} \\
\|t_1^c \; mathop \; t_2^c\|_{\langle M,v\rangle} & = & \|t_1^c\|_{\langle M,v\rangle} \; \mathbf{mathop} \; \|t_2^c\|_{\langle M,v\rangle} \\
\|w_{\langle x_1^o,\ldots,x_n^o\rangle}(\alpha)\|_{\langle M,v\rangle} & = & \mu^n(\{\langle d_1,\ldots,d_n\rangle | \langle M, v[x_1^o/d_1,\ldots,x_n^o/d_n]\rangle \models \alpha\}) \\
\langle M,v\rangle \models P_i^n(t_1^o,\ldots,t_n^o) & \text{iff} & \langle [t_1^o]_{\langle M,v\rangle},\ldots,[t_n^o]_{\langle M,v\rangle}\rangle \in \pi(P_i^n) \\
\langle M,v\rangle \models \neg\alpha & \text{iff} & \langle M,v\rangle \not\models \alpha \\
\langle M,v\rangle \models \alpha\wedge\beta & \text{iff} & \langle M,v\rangle \models \alpha \text{ and } \langle M,v\rangle \models \beta \\
\langle M,v\rangle \models \forall x^o.\alpha & \text{iff} & \langle M,v[x^o/d]\rangle \models \alpha \text{ for all } d \in D \\
\langle M,v\rangle \models \forall x^c.\alpha & \text{iff} & \langle M,v[x^c/r]\rangle \models \alpha \text{ for all } r \in \mathcal{R} \\
\langle M,v\rangle \models t_1^c \; relop \; t_2^c & \text{iff} & \|t_1^c\|_{\langle M,v\rangle} \; \mathbf{relop} \; \|t_2^c\|_{\langle M,v\rangle} \\
\langle M,v\rangle \models t_1^o = t_2^o & \text{iff} & [t_1^o]_{\langle M,v\rangle}=[t_2^o]_{\langle M,v\rangle}
\end{array}
$$

A formula α is *satisfiable* iff there exists $\langle M, v\rangle$ such that $\langle M, v\rangle \models \alpha$; a formula α is *valid* (in symbols: $\models \alpha$) iff $\langle M, v\rangle \models \alpha$ for all $\langle M, v\rangle$. Validity, the main notion of interest in reasoning contexts, has been shown to be decidable in $\mathcal{L}_1$ when the domain D has a fixed, finite cardinality n (see [9]). Note that, although the syntax of the logic might seem too limited for practical uses, a number of other constructs may be defined as "shorthands" of the above formulae. For instance, the Bayesian conditionalisation operator "$w_{\langle x_1,\ldots,x_n\rangle}(-|-)$" is expressed by considering the formula $w_{\langle x_1,\ldots,x_n\rangle}(\alpha|\beta) = r$ as shorthand for the formula $w_{\langle x_1,\ldots,x_n\rangle}(\alpha \wedge \beta) = r \cdot w_{\langle x_1,\ldots,x_n\rangle}(\beta)$. Similarly, the square root operator "$\sqrt{-}$" is expressed by considering the formula $\sqrt{t^c} = r$ as shorthand for the formula $t^c = r \cdot r$. In an actual implementation of the logic, numerical functions such as "$\sqrt{-}$" can obviously be implemented as calls to appropriate subroutines rather than as expansions into the appropriate axiomatic definitions, which then serve for theoretical purposes only.

4 A representation of Imaging on top of Probabilistic Logic

Crestani and van Rijsbergen's models of IR are based on a somewhat non-standard interpretation of Imaging DRFs, as: (1) the representation language is not that of propositional logic but a language of simple propositional letters, each representing a document or a query; (2) possible worlds are keywords; this means that there are not necessarily 2^n possible worlds, but there are as many possible worlds as there are keywords in the application domain[d]. The propositional letter d_i (resp. q_i) is conventionally taken to be true at world t_j iff the document represented by d_i (resp. the query represented by q_i) is indexed by the keyword represented by t_j.

We now describe a representation of the model of IR of [5] in terms of $\mathcal{L}_1$. Our purpose is to show how the representation of these mechanisms may be accomplished quite easily, thus establishing Bayesian tools as convenient and powerful platforms for fast prototyping of non-Bayesian IR models. In the full paper we go on to show how also the model of [6], plus some generalisations of both, can also be easily represented. In our approach, the whole information retrieval process is modelled as *a proper theory* of $\mathcal{L}_1$, obtained by assembling together various sets of formulae, each representing a class of entities participating in the process.

In order to implement Standard Imaging, a first subset of $\mathcal{L}_1$ formulae is necessary to identify keywords and documents. This is necessary, as the domain of interpretation must be restricted to deal with these types of individuals only, which are the only entities of interest in the revision processes. Assuming that $\{t_1, \ldots, t_n\}$ is the language of keywords by means of which documents are represented, and that $\{d_1, \ldots, d_m\}$ are the documents in our collection, we need formulae

$$Keyword(t_1) \ \wedge \ldots \wedge \ Keyword(t_n) \tag{5}$$

$$Document(d_1) \ \wedge \ldots \wedge \ Document(d_m) \tag{6}$$

$$\forall x.[x = t_1 \vee \ldots \vee x = t_n \vee x = d_1 \vee \ldots \vee x = d_m] \tag{7}$$

$$\forall x.\neg(Document(x) \wedge Keyword(x)) \tag{8}$$

This is a key feature of this approach: documents and keywords are individuals *belonging to* the domain of discourse of a first order interpretation, while in the original approach of [5,6] keywords *are* (propositional) interpretations and documents are propositions. Back to this point in Section 5.

[d]The actual methods with which Crestani and van Rijsbergen have dealt with are (Standard) Imaging (in [5]) and an approximation of the combination of General and Proportional Imaging (in [6]).

The next formulae are the ones that specify keyword occurrence, i.e. which documents are indexed by which keywords. We represent this by formulae $w_x(Occ(t_i, d_j)) = o_{ij} \in \{0, 1\}$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$, where o_{ij} is 1 iff t_i occurs in d_j. This representation is made possible by the fact that, as noted in Footnote (c), the probability operator applied to a closed formula yields the formula's characteristic function.

Next, the probability of each keyword t_i is specified with the set of formulae

$$w_x(x = t_i \mid Keyword(x)) = p_{t_i} \qquad p_{t_i} \in [0, 1] \tag{9}$$

for all $i = 1, \ldots, n$. These formulae account for the case in which we want to input the probability values p_{t_i} from the outside. Alternatively, these values can be computed *within* $\mathcal{L}_1$ from the already available occurrence data, e.g. as their inverse document frequency (IDF). In this case, formulae (9) are substituted by formulae

$$w_x(x = t_i \mid Keyword(x)) = -log(w_y(Occ(t_i, y) \mid Document(y))) \tag{10}$$

The formula $w_y(Occ(t_i, y) \mid Document(y))$ is in fact to be read as "the probability that, by picking a random document y, keyword t_i occurs in y". For (10) to truly represent IDF, though, we must assume that documents are picked with equal probability, which we state by formula

$$\forall xy.(Document(x) \wedge Document(y)) \Rightarrow [w_z(x = z) = w_z(y = z)] \tag{11}$$

Alternatively, one might choose to include both formulae (9), (10) and (11) in the representation. In this way, probability values would be precomputed "externally" and input to the reasoning process through formulae (9), and formulae (10) and (11) would act as *integrity constraints*. In what follows we will use the expression $P(t_i)$ as short for $w_x(x = t_i \mid Keyword(x))$.

The next subset of formulae is the one that specifies the *similarity matrix*, i.e. how similar document d_i is to document d_j for all $1 \leq i, j \leq m$, $i \neq j$:

$$Sim(t_i, t_j) = s_{ij} \qquad 1 \leq i, j \leq m, i \neq j \tag{12}$$

Only similarities between nonequal documents are specified; in fact the case $i = j$ is not interesting for Imaging methods, and its specification would complicate the expression of formulae (15). Values s_{ij} are input from an external source of information. Alternatively, they can be computed from within $\mathcal{L}_1$ from the already available occurrence values; for instance, they may be taken to be equivalent to the degree of coextensionality of the *Occ* predicate and computed by means of the formula:

$$Sim(t_i, t_j) = w_x(Occ(t_i, x) \mid Occ(t_j, x)) \cdot w_x(Occ(t_j, x) \mid Occ(t_i, x)) \tag{13}$$

or else be computed according to some other measure of similarity (e.g. the EMIM measure adopted in [5]). Again, formulae (12) and (13) might coexist, with formulae (13) acting then as integrity constraints. Further integrity constraints might be added, if one's theory of similarity requires one to do so; e.g. similarity may be constrained to be a symmetric relation by means of formula $\forall xy.[Sim(x,y) = Sim(y,x)]$, and/or a triangular relation by means of formula $\forall xy.[Sim(x,y) + Sim(y,z) \geq Sim(x,z)]$.

The following subset of formulae specifies, for each keyword, its most similar keyword:

$$MostSim(t_i, t_{k_i}) \qquad 1 \leq i \leq n \tag{14}$$

Similarly to formulae (9) and (12) these formulae account for the case in which we want to input the "most-similarity" values from the outside. Alternatively, these values can be computed within $\mathcal{L}_1$ from the already available similarity data by means of the formulae

$$MostSim(t_i, t_{k_i}) \quad \Leftrightarrow \quad \neg\exists t_j.[Sim(t_i, t_j) \geq Sim(t_i, t_{k_i})] \tag{15}$$

Again, formulae (14) and (15) may coexist, with formulae (15) acting then as integrity constraints.

Next, we have to show how to calculate the revised probability of keyword t_i by Imaging on document d_j, i.e. how to implement the probability transfer function. The revised probabilities are specified by the following numerical terms, for $1 \leq i \leq n$:

$$\begin{aligned} & w_x(Occ(t_i, d_j)) \cdot [P(t_i) + \\ & \qquad + \sum_{k=1}^{n} [P(t_k) \cdot w_x(\neg Occ(t_k, d_j)) \cdot w_x(MostSim(t_k, t_i))]] \end{aligned} \tag{16}$$

To interpret term (16) remember Footnote (c) and note that all formulae occurring in the context of a w_x operator are closed (w_x-terms thus act here as "guards"). The summation operator $\sum$ is obviously a shorthand for the corresponding expanded numerical term. In what follows we will use the expression $P^{\#}_{d_j}(t_i)$ as a shorthand of expression (16).

In order to compute relevance of documents to the query, we now have to indicate by which keywords the query q is indexed. This is accomplished by the formulae $w_x(Occ(t_i, q)) = o_i$, for $o_i \in \{0, 1\}$. The probability of relevance of document d_j to query q may be then calculated as the value of the numerical term $Rel^{\#}_{d_j}(q) = \sum_{i=1}^{n} w_x(Occ(t_i, q)) \cdot P^{\#}_{d_j}(t_i)$.

5 Related work and discussion

In this work we have discussed an implementation of a non-Bayesian revision method on top of $\mathcal{L}_1$, a (Bayesian) FOL extended with features for reasoning about objective probability; implementations of other related methods are described in the full paper. These implementations have been achieved by representing the semantic structure that underlies Imaging-based conditional logics in the language of $\mathcal{L}_1$. Recasting the Imaging-related models of information retrieval in such a general purpose knowledge representation and reasoning tool, besides showing the potential of this "Bayesian" tool for the representation of non-Bayesian revision functions, paves the way to a possible integration of these models with other, more KR-oriented models of IR, and to the exploitation of general-purpose domain-knowledge in the IR process.

The nature of this work may be discussed more effectively by comparing it with the implementation of the Imaging-based models of IR discussed in [15,16]. These works, instead of a full-blown probabilistic FOL, use Probabilistic Datalog [17], an extension of Stratified Datalog (itself a version of the well-known deductive database language Datalog) by means of features for subjective probability. Both in our work and in [15,16], the entities that participate in the Imaging process (the keywords, their prior probabilities, the similarity values between them, the documents and the queries) are given an explicit representation in the language. Unlike in [15,16], however, in our approach an explicit representation is given also to the formula that computes the prior probabilities of keywords, to the formula that computes the similarities between keywords and to the formula that chooses the recipients of probability transfers and computes the revised probabilities of these recipients; the meaning of all these formulae is definable in terms of just the available keyword occurrence data. This hints to the fact that different formulae encoding different methods of computation of the above features may be experimented with in our approach. In this sense, the whole Imaging DRF is completely modelled as *a proper theory* of $\mathcal{L}_1$. The definitions of [16] and [15] are instead rather partial, as most of the reasoning needed for the implementation of the DRF has to be done by some external process.

Our approach has the advantage of being more self-contained and conceptually attractive, as it requires the minimum amount of data to be provided from outside the reasoning mechanism. Moreover, with a minimal coding effort, different probability kinematics methods may be experimented with and compared, as can be seen in the full paper by the ease with which we have encoded the probability transfer formula of different variants of Imaging in $\mathcal{L}_1$. The price to be paid for this is that of efficiency, as reasoning in Probabilistic

Datalog, a less expressive reasoning tool than $\mathcal{L}_1$, is no doubt computationally easier.

One may wonder why the implementations of [15,16] require the prior probabilities of keywords, the similarities between keywords and the revised probabilities of keywords to be computed externally. We think that the answer does not lie in the fact that Probabilistic Datalog is a less powerful tool than $\mathcal{L}_1$, but in the fact that it is inherently geared towards subjective, and not objective, probability (this character of Probabilistic Datalog can be seen from the fact that its semantics contemplates density functions on possible worlds rather than on the individuals of the domain). This entails the impossibility to represent entities that are inherently of a frequentistic nature, such as the IDF of a keyword (see Equation 10) and the notion of similarity between two keywords as degree of coextensionality (see Equation 13). It also somewhat entails a distortion of the meaning of probabilities. For instance, in Probabilistic Datalog one needs to code keyword prior probabilities by means of sentences of type `0.2 term`(t_1), which literally means "the agent believes, with degree of confidence `0.2`, that t_1 is a keyword". In $\mathcal{L}_1$ one writes instead $w_x(t_1) = 0.2$, which means "the probability that a random pick among keywords yields t_1 is 0.2". The latter is no doubt a more faithful rendition of prior probabilities of keywords.

Acknowledgements

I would like to thank Fabio Crestani, Thomas Rölleke and Keith van Rijsbergen for fruitful discussions on the topics of this paper.

References

1. Cornelis J. van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29:481–485, 1986.
2. Fabrizio Sebastiani. A note on logic and information retrieval. In Ian Ruthven, editor, *Proceedings of MIRO-95, Workshop on Multimedia Information Retrieval*, pp. 1–15, Glasgow, UK, 1996. Springer Verlag, Heidelberg, FRG.
3. David K. Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85:297–315, 1976.
4. Peter Gärdenfors. Imaging and conditionalization. *Journal of Philosophy*, 79:747–760, 1982.
5. Fabio Crestani and Cornelis J. van Rijsbergen. Information retrieval by logical imaging. *Journal of Documentation*, 51:3–17, 1995.

6. Fabio Crestani and Cornelis J. van Rijsbergen. Probability kinematics in information retrieval. In *Proceedings of ACM SIGIR-95*, pp. 291–299, Seattle, WA, 1995.
7. Cornelis J. van Rijsbergen. Towards an information logic. In *Proceedings of ACM SIGIR-89*, pp. 77–86, Cambridge, MA, 1989.
8. Johan F. van Benthem. *Exploring logical dynamics.* Center for the Study of Language and Information, Stanford, CA, 1996.
9. Joseph Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.
10. Fabrizio Sebastiani. Information retrieval, imaging and probabilistic logic. Technical report, Istituto di Elaborazione dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy, 1997. Forthcoming.
11. Richard C. Jeffrey. *The logic of decision.* McGraw Hill, New York, NY, 1965.
12. Fabio Crestani, Fabrizio Sebastiani, and Cornelis J. van Rijsbergen. Imaging and information retrieval: Variations on a theme. In Fabio Crestani and Mounia Lalmas, editors, *Proceedings of the 2nd International Workshop on Logic and Uncertainty in Information Retrieval*, pp. 48–49, Glasgow, UK, 1996.
13. Carlo Meghini, Fabrizio Sebastiani, Umberto Straccia, and Costantino Thanos. A model of information retrieval based on a terminological logic. In *Proceedings of ACM SIGIR-93*, pp. 298–307, Pittsburgh, PA, 1993.
14. Fabrizio Sebastiani. A probabilistic terminological logic for modelling information retrieval. In *Proceedings of ACM SIGIR-94*, pp. 122–130, Dublin, IRL, 1994.
15. Fabio Crestani and Thomas Rölleke. Issues in the implementation of general imaging on top of Probabilistic Datalog. In Mounia Lalmas, editor, *Proceedings of the 1st International Workshop on Logic and Uncertainty in Information Retrieval*, Glasgow, UK, 1995.
16. Thomas Rölleke. Does Probabilistic Datalog meet the requirements of imaging? In *Proceedings of ACM SIGIR-95*, p. 374, Seattle, WA, 1995.
17. Norbert Fuhr. Probabilistic Datalog: a logic for powerful retrieval methods. In *Proceedings of ACM SIGIR-95*, pp. 282–290, Seattle, WA, 1995.