Preliminary Experiments using Subjective Logic for the Polyrepresentation of Information Needs

Christina Lioma
Computer Science
University of Copenhagen
Denmark
c.lioma@diku.dk

Birger Larsen
Royal School of Library and
Information Science
Copenhagen Denmark
blar@iva.dk

Peter Ingwersen[*]
Royal School of Library and
Information Science
Copenhagen Denmark
pi@iva.dk

ABSTRACT

According to the principle of *polyrepresentation*, retrieval accuracy may improve through the combination of multiple and diverse information object representations about e.g. the context of the user, the information sought, or the retrieval system [9, 10]. Recently, the principle of polyrepresentation was mathematically expressed using *subjective logic* [12], where the potential suitability of each representation for improving retrieval performance was formalised through degrees of belief and uncertainty [15]. No experimental evidence or practical application has so far validated this model.

We extend the work of Lioma et al. (2010) [15], by providing a practical application and analysis of the model. We show how to map the abstract notions of belief and uncertainty to real-life evidence drawn from a retrieval dataset. We also show how to estimate two different types of polyrepresentation assuming either (a) independence or (b) dependence between the information objects that are combined. We focus on the polyrepresentation of different types of context relating to user information needs (i.e. work task, user background knowledge, ideal answer) and show that the subjective logic model can predict their optimal combination prior and independently to the retrieval process.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Theory

Keywords

Polyrepresentation, Subjective Logic, Opinion Fusion

---

[*]Second affiliation: Oslo University College, Norway

1. INTRODUCTION

The principle of *polyrepresentation* [9, 10] posits that information retrieval (IR) effectiveness may improve through the consideration of multiple and diverse representations of information objects or processes, such as IR systems, models, or context about the documents and queries and their mediation to the user. This consideration is usually implemented experimentally as combination of diverse representations with respect to some criterion, more commonly retrieval effectiveness. Whereas in principle, polyrepresentation aims to use different information object representations that may enhance IR performance, in practice, each representation is often associated with different degrees of uncertainty regarding the enhancement that it may bring. Therefore, if one representation is weaker (less reliable or accurate) than the others, this should be reflected in the combination process, otherwise effectiveness will suffer. In addition, given the potentially high-dimensional and noisy data contained in information object representations, many different combinations could be produced, fetching various different retrieval results. The process of deciding which combination of all is optimal for retrieval is an open problem, which was recently formalised by Lioma et al. (2010) [15] in the context of polyrepresentation using a type of probabilistic logic called subjective logic [12]. This formalism was not accompanied by experimental evidence or practical applications.

This work can be seen as an extension of the work of Lioma et al. (2010); we provide a practical application of that model and show that it is experimentally validated. We show how to map the abstract notions of belief and uncertainty to real-life evidence drawn from a retrieval test collection, and how to estimate two different types of combinations for polyrepresentation assuming either (a) independence or (b) dependence between the information objects that are combined. We focus on the polyrepresentation of different types of context relating to user information needs (i.e. work task, user background knowledge, ideal answer) and show that the subjective logic model can predict their optimal combination prior and independently to the retrieval process. This finding holds for six different standard evaluation measures and two state of the art retrieval models. Furthermore, we look at non-commutativity in polyrepresentation in order to trace the bias between dependent contextual representations, and experimentally show its impact to retrieval effectiveness.

The remainder of this work is organised as follows. Section

2 overviews related work. Section 3 summarises the model of Lioma et al. (2010). Section 4 presents the experimental evaluation of the model. Section 5 discusses our findings and their limitations. Section 6 concludes this work.

2. RELATED WORK

There exist several applications of the principle of polyrepresentation to IR (see [8, 15] for overviews). Regarding the application of polyrepresentation to user information needs, that is the focus of this work, most studies use polyrepresentation for query expansion. Diriye et al. (2009) [6] apply polyrepresentation for interactive query expansion, i.e. to improve the suggestion of expansion terms to the users during their search in order to enable better retrieval performance. They show how providing supplementary information on expansion terms can address ambiguity and uncertainty issues, and can improve the perceived usefulness of the terms. In a different (non-interactive approach), Efron and Winget (2010) [7] use polyrepresentation to combine so-called query aspects, which they collect by skimming the top k documents retrieved for a given query. They consider these skimmed documents relevant and build pseudorelevance judgments without human intervention.

Our study differs from the above: we do not apply polyrepresentation for query expansion, but rather as a means to combine contextual evidence about the queries. This is more reminiscent of evidence combination or opinion fusion approaches, as explained below. The effect of capturing multiple representations of query context from a single user or from multiple users is a topic that has long attracted interest. According to Croft (2000) [3], McGill et al. (1979) carried out a study of factors affecting retrieval by different users when they were assigned the same information need as a starting point. They found that there was surprisingly little overlap between the documents retrieved by the different users. Saracevic and Kantor (1988) [18] also found that when different users constructed Boolean queries based on the same descriptions of the information need, there was little overlap in the retrieved sets. One of the earliest perhaps retrieval models that explicitly incorporated the notion of multiple query representations was proposed by Turtle and Croft (1991) [22]. Soon after this, Belkin et al. (1993) [1] carried out a systematic study on the effects of query combination and verified that retrieval effectiveness could be substantially improved by query combination, but that the effectiveness of the combination depends on the effectiveness of the individual queries. Queries that provided less evidence about relevance should optimally have lower weights in the combination, because bad query representations could reduce effectiveness when combined with better representations.

This is exactly the topic we address in this work, based on the model proposed in Lioma et al. (2010). How to identify and separate 'good' from 'bad' query representations, so that undesirable combinations can be avoided, considering not only the features of the query representations, but also the way in which they are combined.

3. POLYREPRESENTATION USING SUBJECTIVE LOGIC

This section outlines basic notions in subjective logic and how it has been applied to polyrepresentation in Lioma et

al. (2010). A detailed introduction into subjective logic can be found in Jøsang (2001) [12].

3.1 Subjective Logic Preliminaries

The starting point is a frame of discernment defined over a proposition. Arguments in subjective logic are *opinions* representing the belief that the proposition is true[1]. An opinion is formally defined as ω_x^A, where A is the opinion owner, and x is the proposition to which the opinion ω applies. An opinion can be decomposed into $\omega_x^A = (b, d, u, \alpha)$, where b is A's belief that the proposition is true, d is A's disbelief that the proposition is true[2], u is A's uncertainty about the proposition, and α is an an a priori probability in the absence of committed belief mass. Given a proposition for which we have an opinion, we can estimate the probability expectation that the proposition is true as:

$$E = b + \alpha \cdot u \qquad (1)$$

If there is one opinion only about the proposition, Equation 1 is computed directly from the components of that opinion. If however several opinions exist about a proposition, assessing the truth of the proposition consists in fusing these opinions and estimating Equation 1 from their combined components. Subjective logic describes several fusion operators, suitable for different situations. Two of these are outlined next: *consensus*, which combines independent opinions without bias, and *recommendation*, which combines dependent opinions by modelling the influence of one opinion upon the other.

Consensus.
Let us assume two independent opinions about a proposition x: $\omega_x^A = (b_x^A, d_x^A, u_x^A, \alpha_x^A)$ and $\omega_x^B = (b_x^B, d_x^B, u_x^B, \alpha_x^B)$. Their consensus is $\omega_x^{A \oplus B}$ with components:

$$b_x^{A \oplus B} = \frac{b_x^A u_x^B + b_x^B u_x^A}{\kappa}, \qquad d_x^{A \oplus B} = \frac{d_x^A u_x^B + d_x^B u_x^A}{\kappa} \quad (2)$$

$$u_x^{A \oplus B} = \frac{u_x^A u_x^B}{\kappa}, \qquad \kappa = u_x^A + u_x^B - u_x^A u_x^B \qquad (3)$$

This operation[3] is commutative, associative and assumes that not all of the combined opinions have zero uncertainty.

Recommendation.
Let us assume two opinions that are not independent of each other, but where one influences the other. Let $\omega_x^B = (b_x^B, d_x^B, u_x^B, \alpha_x^B)$ be B's opinion about the proposition, and $\omega_B^A = (b_B^A, d_B^A, u_B^A, \alpha_B^A)$ be A's opinion about B's recommendation (B's influence to A). Their combination by recommendation $\omega_x^{A \otimes B}$ is A's opinion about the proposition as a result of the recommendation from B, with components:

$$b_x^{A \otimes B} = b_B^A b_x^B, \qquad d_x^{A \otimes B} = b_B^A d_x^B \qquad (4)$$

$$u_x^{A \otimes B} = d_B^A + u_B^A + b_B^A u_x^B \qquad (5)$$

---

[1] The opinion space is a subset of the belief space used in the Dempster-Shafer belief theory [20].
[2] This corresponds to *doubt* in Shafer (1976) [19].
[3] The consensus operator is similar to Dempster's rule [5] (see [12], Section 5.3, for a discussion on the difference between the two).

Table 1: Polyrepresentation and subjective logic analogies.

POLYREPRESENTATION	SUBJECTIVE LOGIC
· original query	· proposition
· query context	· opinion
· how well a query context represents the original query	· belief, disbelief & uncertainty of an opinion about the truth of the proposition
· combination of query context representations	· opinion fusion

This operation[4] is associative but not commutative.

3.2 Subjective Logic for Polyrepresentation

The model of Lioma et al. (2010) uses subjective logic to express the polyrepresentation of user information needs and their context. An analogy is made between opinions (from subjective logic) and query context representations (from polyrepresentation). The opinions express beliefs about the truth of a proposition. In this analogy, the proposition is the original query, and the opinions are the query context representations (see Table 1). The opinion of a query context representation can be seen as the extent to which the query context represents the original query. Then, the combination of multiple query context representations can be modelled as fusing opinions about a proposition.

4. PRACTICAL APPLICATION

In this paper, we implement and experimentally test the model of Lioma et al. (2010) summarised in the previous section. Given an IR test collection with rich query context representations, we convert these representations into subjective logic opinions and we produce their combinations using the consensus and recommendation operations. The strength of each combination is computed as a probability using Equation 1, so that the higher the probability, the closer a combination represents the original query. To evaluate this model, we conduct retrieval experiments using the same query context combinations (completely independently from the subjective logic computations). If the best performing combination of query context (according to retrieval performance) corresponds to the combination of subjective logic opinions with the highest probability, we consider the model validated. In this work, we focus only on pairwise combinations of query context representations.

Section 4.1 describes the dataset and settings used in the polyrepresentation and retrieval experiments. Section 4.2 presents the experimental findings.

4.1 Experimental Setup

4.1.1 Dataset and query context representations

We use the iSearch test collection [16], which consists of 46GB of scientific documents from the physics domain. iSearch comes with a set of 65 queries and their relevance assessments, which have been created by 23 lecturers and experienced postgraduate and graduate students from three different university departments of physics. The queries represent real information seeking tasks. The relevance assessment of each query was made by the same user who formulated that query, by examining a pool of documents retrieved for that query.

Each iSearch query contains the following five representations of different aspects of the user's information need and context:

1. user's verbose description of the information sought

2. background of the user's task

3. description of the user's current work task

4. description of the user's ideal answer

5. query terms (keywords) that the user might use in a search engine

In this work, we consider the keywords (representation no. 5) as the original query, and representations 1-4 as expressions of query context. We choose to regard the keywords as the representation closest to original user query for two reasons. First, the keywords are the most similar to what a user might submit to a search engine, e.g. they are most often in the form of a few key terms or phrases, and are the least verbose of the five representations. Second, earlier results on the iSearch collection indicate [21] that the single best performing representation among these five is the keywords, indicating that this representation works well with current retrieval models.

4.1.2 Retrieval settings

We use the Indri IR system[5]. For ranking we use two versions of the language model: with Dirichlet (Dir) and Jelinek-Mercer (JM) smoothing [4]. We tune their parameters using the tuning range in Zhai and Lafferty (2002) [24]:

- DIR's $\mu \in \{100, 500, 800, 1000, 2000, 3000, 4000, 5000, 8000, 10000\}$

- JM's $\lambda \in \{0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$

We report the best retrieval performance separately for Mean Average Precision (MAP), Normalised Discounted Cumulated Gain (NDCG)[6], and Binary Preference (BPREF) in the top 1000 retrieved documents, and also for Precision at 10 (P@10), NDCG at 10 (NDCG@10) and Mean Reciprocal Rank (MRR). These measures contribute different aspects to the evaluation: Both MAP and BPREF measure the average precision of a ranked list, but BPREF differs from MAP because it does not treat non-assessed documents as explicitly non-relevant (whereas MAP does) [2]. NDCG measures the gain of a document based on its position in the result list. The gain is accumulated from the top of the ranked list to the bottom, with the gain of each document discounted at lower ranks. This gain is relative to the ideal based on a known recall base of relevance assessments [11]. P@10 and NDCG@10 focus on the early precision of the top 10 retrieved documents. MRR [23] corresponds to the multiplicative inverse of the rank of the first relevant document retrieved, i.e. it focuses on the retrieval quality of the very top of the ranked list.

---

[4]The recommendation operator can become equivalent to Shafer's discounting function [19] as explained in Lioma et al. (2009) [14], section 3.2.

[5]http://www.lemurproject.org/indri/

[6]with the following gain values: very relevant = 3, fairly relevant = 2, marginally relevant = 1, non-relevant = 0.

The user query (Q) and two representations (A and B) of its context. Each representation contains positive and negative evidence about the strength of the combination $A \oplus B$ or $A \otimes B$ with respect to the query.
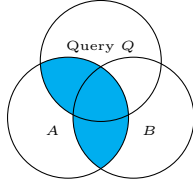

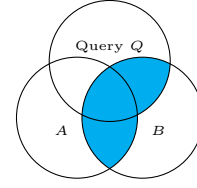**Figure 1: Positive evidence of A (in $A \oplus B$).**

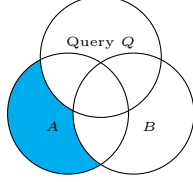**Figure 2: Positive evidence of B (in $A \oplus B$).**

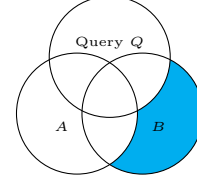**Figure 3: Negative evidence of A (in $A \oplus B$).**
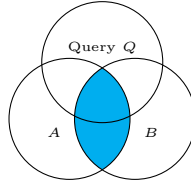

Figure 4: Negative evidence of B (in $A \oplus B$).

Figure 5: Positive evidence of A about B and also of B about A (in $A \otimes B$).

4.2 Experimental Findings

4.2.1 Mapping Opinions to Evidence

According to the model of Lioma et al. (2010) summarised in section 3, each representation of query context corresponds to an opinion about the original query. The belief, disbelief and uncertainty of these opinions must be computed from features of these query context representations (completely independently of the retrieval process). These features are referred to as evidence, and can be either positive or negative with respect to the original query, depending on whether they support it or not. Subjective logic maps this type of evidence to opinions as follows [12]. Let r denote positive evidence, and s denote negative evidence. Then, the correspondence between this evidence and the belief, disbelief, and uncertainty b, d, u is:

$$b = \frac{r}{r+s+2} \qquad d = \frac{s}{r+s+2} \qquad u = \frac{2}{r+s+2} \quad (6)$$

What constitutes positive and negative evidence can be defined in various ways. In this work, we use the terms of the query context representations in a simple bag of words approach (regardless of their frequency, co-occurrence, grammatical or any other feature). We consider each query context representation and the original query as sets of terms. Then, positive evidence is the number of terms that occur in overlaps of these sets, and negative evidence is the number of terms that occur in complements of these sets, as described below.

For the consensus of two query context representations A

and B ($A \oplus B$) with respect to a user query Q, A's positive evidence is the number of terms occurring in $A \cap B$ and also in $A \cap Q$ (Figure 1). Similarly, B's positive evidence is the number of terms occurring in $A \cap B$ and also in $B \cap Q$ (Figure 2). A's negative evidence is the number of terms occurring in the complement of $A - B$ and $A - Q$ (Figure 3). B's negative evidence is the number of terms occurring in the complement of $B - A$ and $B - Q$ (Figure 4).

For the recommendation of A to B and of B to A ($A \otimes B$), both A's and B's positive evidence is the number of terms occurring in $A \cap B$ (Figure 5). The negative evidence of A and B is the same as in the consensus above.

We apply these separate pre-processing options when counting terms:

    (I) no-preprocessing at all

    (II) lower-case, punctuation removal

    (III) II + stopword removal

    (IV) III + stemming

We use the SMART stopword list, as appears in the appendices of Lewis et al. (2004) [13], and the Porter stemming algorithm [17].

Any type and amount of contextual information that can potentially be useful may be used as evidence, for instance statistical, linguistic, algorithmic or other features of the queries (e.g. see [14] for pragmatic query features). Additional evidence can be modelled by introducing more opinions.

Table 2: Pairwise polyrepresentation of query context representations through consensus ($\oplus$) and recommendation ($\otimes$) operations. Consensus and recommendation reflect the strength of each combination as a probability: the higher the probability, the more likely that the combination will benefit retrieval performance. Recommendation is non-commutative, so $A \otimes B$ and $B \otimes A$ denote the combination order.

POLYREPRESENTATION PROBABILITIES			
(I) No Pre-Processing			
Query Context Representations	Consensus $A \oplus B$	Recommendation $A \otimes B$	$B \otimes A$
user background - ideal answer	0.1850	0.4126	0.4192
user background - work task	0.1965	0.3527	0.4518
information need - user background	0.1862	0.3759	0.4594
information need - ideal answer	**0.2359**	0.3687	0.4450
information need - work task	0.2064	0.3096	**0.4764**
work task - ideal answer	0.1769	**0.4258**	0.3859

POLYREPRESENTATION PROBABILITIES			
(II) Lower-Case, No Punctuation			
Query Context Representations	Consensus $A \oplus B$	Recommendation $A \otimes B$	$B \otimes A$
user background - ideal answer	0.2285	0.3967	0.4153
user background - work task	0.2347	0.3319	0.4540
information need - user background	0.2426	0.3470	0.4702
information need - ideal answer	**0.2939**	0.3379	0.4584
information need - work task	0.2534	0.2680	**0.4895**
work task - ideal answer	0.2123	**0.4137**	0.3811

POLYREPRESENTATION PROBABILITIES			
(III) Lower-Case, No Punctuation, No Stopwords			
Query Context Representations	Consensus $A \oplus B$	Recommendation $A \otimes B$	$B \otimes A$
user background - ideal answer	0.1972	0.4365	0.4499
user background - work task	0.1826	0.4030	0.4641
information need - user background	0.2457	0.3928	0.4787
information need - ideal answer	**0.3154**	0.3493	0.4734
information need - work task	0.2715	0.2991	**0.4907**
work task - ideal answer	0.1902	**0.4436**	0.4274

POLYREPRESENTATION PROBABILITIES			
(IV) Lower-Case, No Punctuation, No Stopwords, Stemming			
Query Context Representations	Consensus $A \oplus B$	Recommendation $A \otimes B$	$B \otimes A$
user background - ideal answer	0.2278	0.4211	0.4424
user background - work task	0.2203	0.3771	0.4616
information need - user background	0.2885	0.3607	0.4837
information need - ideal answer	**0.3674**	0.3138	0.4868
information need - work task	0.3106	0.2615	**0.5027**
work task - ideal answer	0.2172	**0.4307**	0.4170

Having collected this evidence from the term statistics of the query context representations, we compute their combinations by consensus (using Equations 2 - 3) and recommendation (using Equations 4 - 5). The output of these equations is fused beliefs, disbeliefs, and uncertainties for each combination, which we feed into Equation 1 to compute the final strength of each combination - we refer to this as *polyrepresentation probability*. In Equation 1, we set the prior $\alpha = 0.5$ assuming a binary frame of discernment (i.e. having two states and dividing α uniformly across these).

4.2.2 Polyrepresentation Findings

Table 2 shows the probabilities of each combination of query context representations computed as described above. The three query context representations that give higher polyrepresentation probabilities consistently for all preprocessing options are the user's information need, work task and ideal answer. Among these, work task and information need give the highest polyrepresentation probability at all times, using the recommendation operation (in this order). The assumption behind this combination is that (a) work task and information need are dependent, and that (b) the work task is a better contextual representation of the information need than of the original query. This can be seen in the very low probability fetched by the recommendation information need ⊗ work task (which is the lowest probability for all recommendation combinations). Simply speaking, what we see here is two query context representations that are potentially valuable to retrieval effectiveness but only if their effect is channelled in a specific way so that their effect is traced according to their dependence. Their dependence in this case is that the work task context supports the information need, and the information need context supports the query. This is in line with the cognitive interpretation of user's work task in interactive IR laid down in Ingwersen (1996) [9].

Table 2 also shows that recommendation results in higher overall polyrepresentation probabilities than consensus, i.e. combinations that are likely to be more reliable. To assess the predictions of the query context combinations in Table 2, we compare them to the actual retrieval performance of each combination, described next. No part of the retrieval process has been informed by the polyrepresentation computation described above, and vice versa.

4.2.3 Retrieval Findings

Table 3 shows the retrieval performance of the combinations of query context representations we saw in Table 2. For these runs, we use the text of each query context representation as query text, and the same four types of preprocessing reported in Section 4.2.1. We do not weight separately any of the query fields; we simply concatenate all text into one query. Table 3 also displays the retrieval performance when using the original query without any context, for reference (there is no polyrepresentation of context in this run). We see that the combination with the highest polyrepresentation probability in Table 2, information need and work task, gives also the best retrieval performance among all combinations. This is consistent for all six evaluation measures, and for both retrieval models. Comparing these results to the performance of the original query without context, we see that the best polyrepresentation run improves retrieval performance over the original query for MAP, P@10 and

NDCG@10, and is comparable to the original query for the remaining evaluation measures. Overall, these findings indicate that the polyrepresentation probabilities tend to agree with observed retrieval performance. This agreement is also visually displayed in Figure 6. We see that there is overall consistency among the two, apart from the case of background - work task.

The scores shown in Table 3 are averaged over all queries, meaning that they can be affected by outliers. Figures 7 - 30 present a detailed per-query overview of the retrieval performance of each query (measured in MAP only for DIR) against the belief and uncertainty of the query context representations combined with consensus and recommendation[7]. This belief and uncertainty are the ingredients used to compute the polyrepresentation probabilities with Equation 1. Hence, these figures serve to explain how the polyrepresentatin probabilities can be decomposed and how their components correlate to retrieval performance. The correlation between MAP and the respective belief and uncertainty values is reported using Spearman's rank correlation coefficient. We can report similar trends for runs with JM and for the other evaluation measures that average retrieval precision in the top 1000 documents.

Figures 7 - 30 show that the combination of information need and work task (which gave the highest polyrepresentation probability and also the highest retrieval performance) does not have the highest correlation with MAP on a per query basis, neither for belief nor for uncertainty. The highest correlation with respect to these is given by the combination of information need with user background. However, the combination of information need and work task has overall the highest belief values among all combinations. This shows that this combination has somewhat stronger evidence than the others, which may not always correlate strongly with the MAP score flunctuations for each query, but which nevertheless is an overall better representation of the query context.

5. DISCUSSION

The experimental findings presented above can be summarised in three points.

The experiments indicate that the polyrepresentation predictions regarding the optimal combination of query context representations were in agreement with observed retrieval performance. This validates the model of Lioma et al. (2010) for this dataset and retrieval scenario. Further experiments with other settings and using more than pairwise combinations of representations are needed to ground the model on firmer grounds - we consider this study as a first step in that direction.

We showed how the components of each polyrepresentation combination were induced from naive bag of words term statistics and converted to belief, disbelief and uncertainty. This was a straight-forward application of the mapping proposed in Jøsang (2001) [12] and of features common to IR experimentation. Further features could be used, potentially challenging their mapping to binary (positive or negative) evidence, resulting in new definitions and mappings of graded features as contextual evidence for IR.

---

[7]For the recommendation, we present one order of combination (out of the two); we can report that the other combination order follows very similar trends.

Table 3: Retrieval performance using the language model with Dirichlet (DIR) and Jelinek-Mercer (JM) smoothing. The best combination of query context representations is shown in bold. † marks the best overall score per evaluation measure.

Query Context Representations	MAP		NDCG		BPREF	
	DIR	JM	DIR	JM	DIR	JM
user background - ideal answer	0.0588	0.0703	0.1928	0.2217	0.2042	0.2347
user background - work task	0.0933	0.1095	0.2589	0.3062	0.2737	0.3059
information need - user background	0.1073	0.1229	0.2919	0.3341	0.2844	0.3213
information need - ideal answer	0.0945	0.1077	0.2670	0.2983	0.2508	0.2834
information need - work task	**0.1175**	**0.1310†**	**0.3117**	**0.3563**	**0.2974**	**0.3270**
work task - ideal answer	0.0849	0.0925	0.2489	0.2895	0.2480	0.2960
original query (no context)	0.1156	0.1268	0.3339	0.3572†	0.3201	0.3340†

Query Context Representations	P@10		NDCG@10		MRR	
	DIR	JM	DIR	JM	DIR	JM
user background - ideal answer	0.1446	0.1477	0.1298	0.1314	0.3069	0.3310
user background - work task	0.2077	0.2277	0.1911	0.2061	0.3884	0.3907
information need - user background	0.2154	0.2138	0.2004	0.2075	0.4222	0.4600
information need - ideal answer	0.2185	0.2154	0.1976	0.1919	0.4394	0.4167
information need - work task	**0.2523**	**0.2554†**	**0.2246**	**0.2352†**	**0.4582**	**0.4880**
work task - ideal answer	0.2000	0.1892	0.1883	0.1805	0.4134	0.3997
original query (no context)	0.2492	0.2431	0.2287	0.2175	0.5267†	0.4958
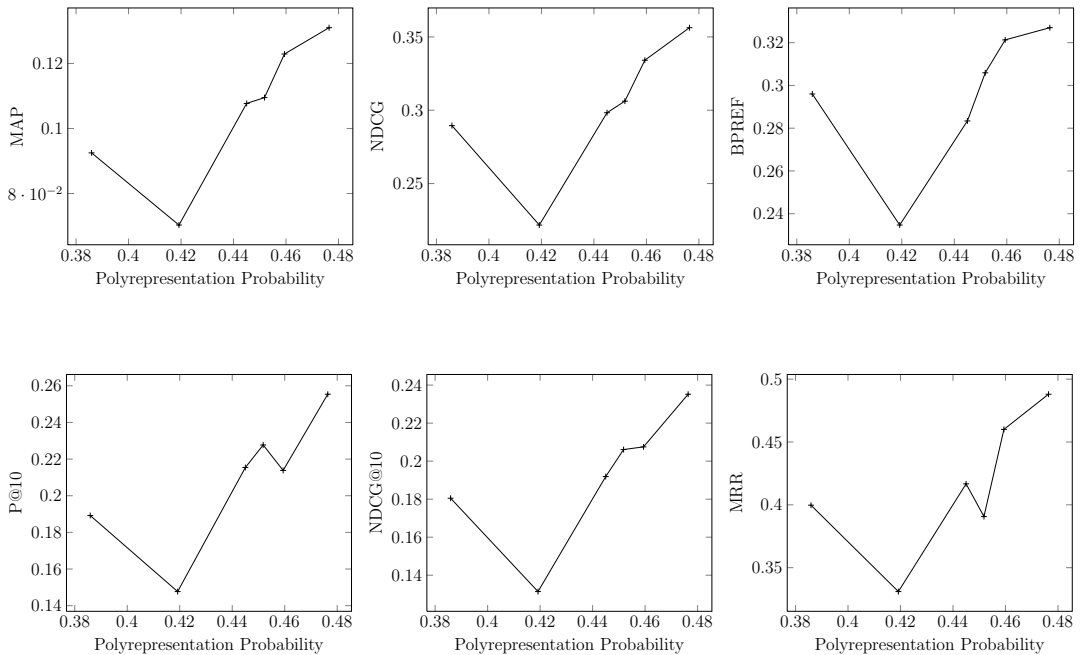
Figure 6: Probability of each polyrepresentation combination (x axis) vs. retrieval performance with JM (y axis) for all query context representation combinations (the points correspond to the scores in Tables 2 & 3).

The belief and uncertainty (y axis) of the representations used for the consensus combination (this page) and the recommendation operator (next page) against MAP (x axis) per query, and their rank correlation coefficients (Spearman's ρ). The order of the combination for recommendation is shown in the title of each figure.
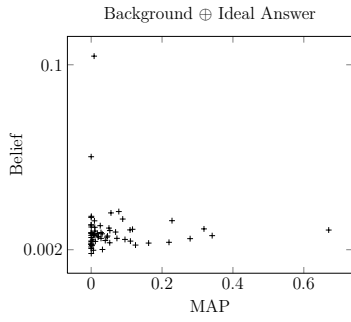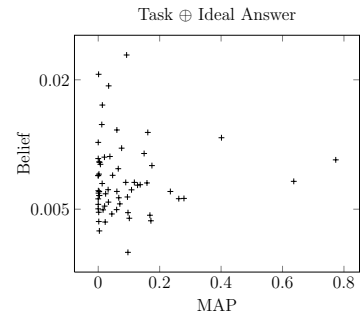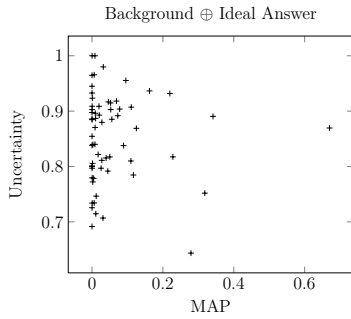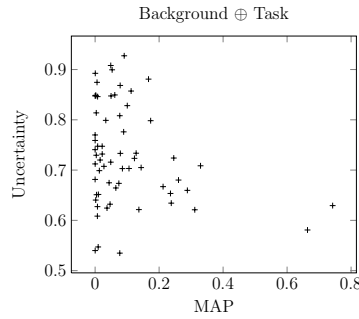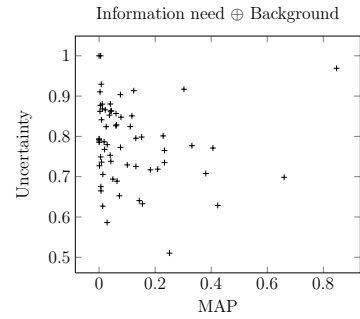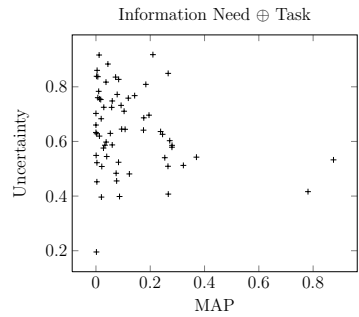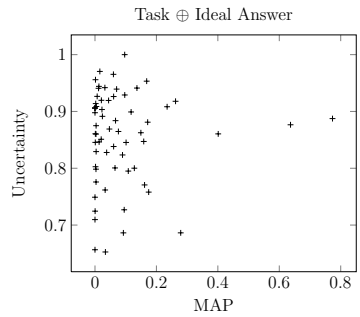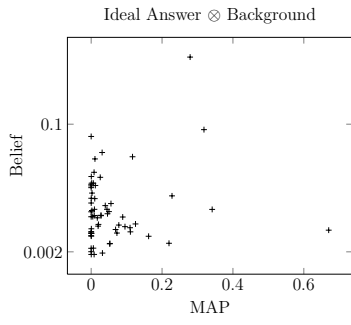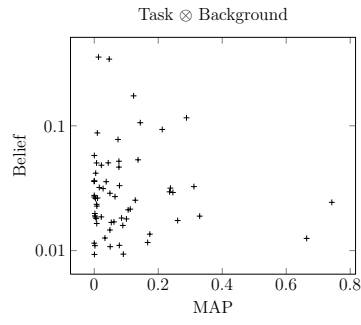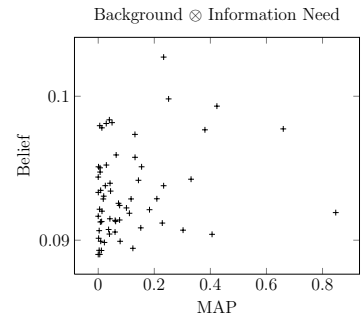

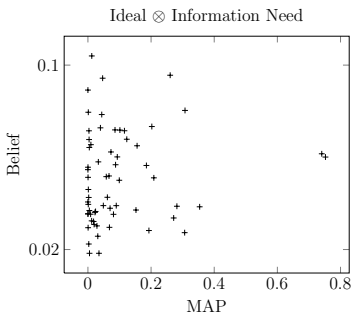Figure 7: $\rho = 0.123$
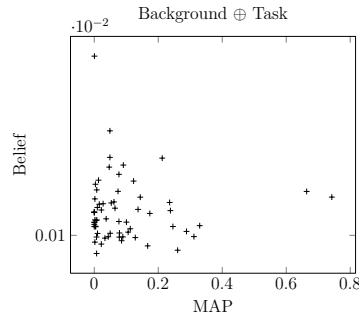
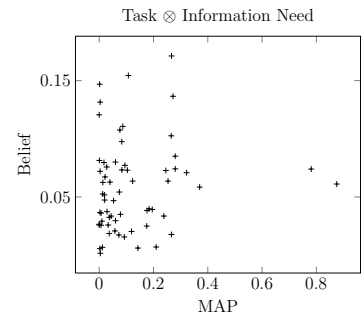

Figure 8: $\rho = -0.014$
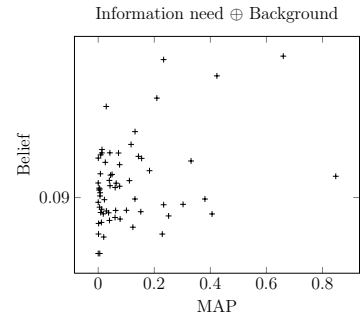
Figure 9: $\rho = 0.483$
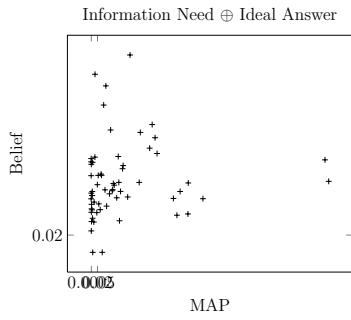
Figure 10: $\rho = 0.411$
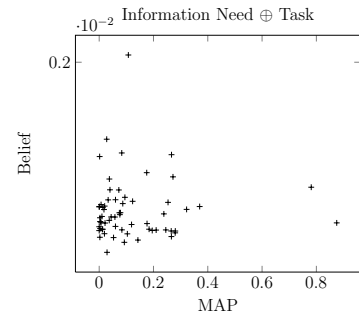
Figure 11: $\rho = 0.359$

Figure 12: $\rho = 0.049$

Figure 13: $\rho = -0.009$
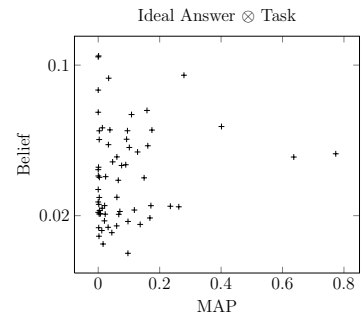

Figure 14: $\rho = -0.198$
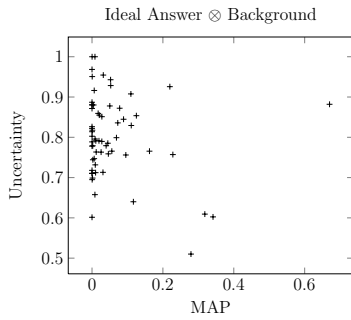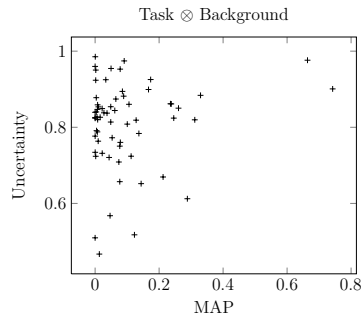

Figure 15: $\rho = -0.209$
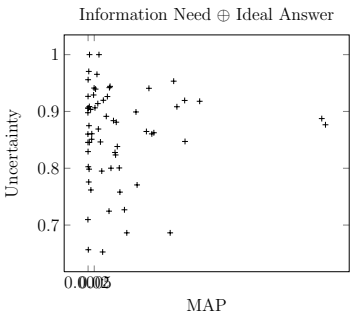
Figure 16: $\rho = -0.146$

Figure 17: $\rho = -0.065$

Figure 18: $\rho = -0 - 007$

Figure 19: $\rho = 0.062$

Figure 20: $\rho = 0.044$

Figure 21: $\rho = 0.238$

Figure 22: $\rho = 0.086$

Figure 23: $\rho = 0.117$

Figure 24: $\rho = 0.105$

Figure 25: $\rho = -0.080$
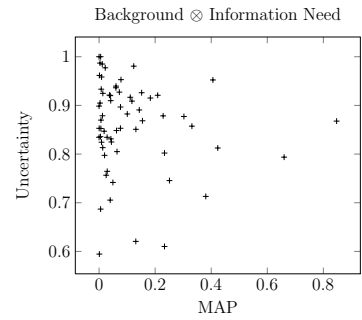
Figure 26: $\rho = 0.019$
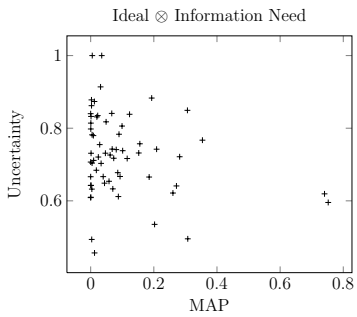
Figure 27: $\rho = -0.136$
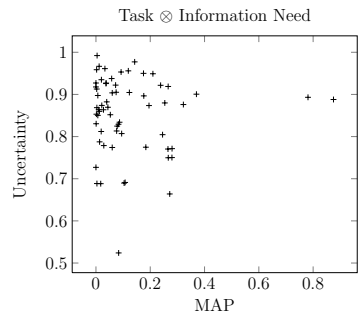
Figure 28: $\rho = -0.064$
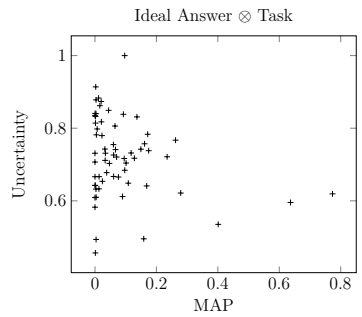
Figure 29: $\rho = -0.061$

Figure 30: $\rho = -0.136$

Finally, we showed how the effect of one representation upon the other can be traced and modelled in a combination. This is an attractive feature that could be used to model and practically adjust the user retrieval experience, through for instance interface functionalities, such as timely interventions or suggestions.

6. CONCLUSIONS

This work provided a practical application and analysis of the principle of polyrepresentation formalised using subjective logic, as initially proposed in Lioma et al. (2010). We showed how to map the abstract notions of belief and uncertainty used in that model to real-life evidence drawn from a retrieval test collection, and how to estimate two different types of combinations for polyrepresentation assuming either (a) independence or (b) dependence between the information objects that are combined. Experimental evidence on the polyrepresentation of different types of context relating to user information needs (i.e. work task, user background knowledge, ideal answer) using two state of the art retrieval models, six standand evaluation measures and 65 queries showed that the model of Lioma et al. (2010) can predict their optimal combination prior and independently to the retrieval process.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] N. J. Belkin, C. Cool, W. B. Croft, and J. P. Callan. Effect of multiple query representations on information retrieval system performance. In R. Korfhage, E. M. Rasmussen, and P. Willett, editors, *SIGIR*, pages 339–346. ACM, 1993.

[2] C. Buckley and E. M. Voorhees. Retrieval evaluation with incomplete information. In M. Sanderson, K. Järvelin, J. Allan, and P. Bruza, editors, *SIGIR*, pages 25–32. ACM, 2004.

[3] W. B. Croft. Combining approaches to ir. In *DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries*, 2000.

[4] W. B. Croft and J. Lafferty. *Language Modeling for Information Retrieval*. Kluwer Academic Publishers, Norwell, MA, USA, 2003.

[5] A. P. Dempster. A Generalization of Bayesian Inference. *Journal of the Royal Statistical Society*, B(30):205–247, 1968.

[6] A. Diriye, A. Blandford, and A. Tombros. A Polyrepresentational Approach to Interactive Query Expansion. In *JCDL*, pages 217–220, 2009.

[7] M. Efron and M. A. Winget. Query polyrepresentation for ranking retrieval systems without relevance judgments. *JASIST*, 61(6):1081–1091, 2010.

[8] I. Frommholz, B. Larsen, B. Piwowarski, M. Lalmas, P. Ingwersen, and K. van Rijsbergen. Supporting polyrepresentation in a quantum-inspired geometrical retrieval framework. In *Proceedings of the third symposium on Information interaction in context*, IIiX '10, pages 115–124, New York, NY, USA, 2010. ACM.

[9] P. Ingwersen. Cognitive Perspectives of Information Retrieval Interaction - Elements of a Cognitive IR Theory. *Journal of Documentation*, 52(1):3–50, 1996.

[10] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[11] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, 20(4):422–446, 2002.

[12] A. Jøsang. A logic for uncertain probabilities. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 9(3):279–311, June 2001.

[13] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li. RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5:361–397, 2004.

[14] C. Lioma, R. Blanco, R. M. Palau, and M.-F. Moens. A Belief Model of Query Difficulty that Uses Subjective Logic. In *ICTIR*, pages 92–103, 2009.

[15] C. Lioma, B. Larsen, H. Schuetze, and P. Ingwersen. A subjective logic formalisation of the principle of polyrepresentation for information needs. In *Proceedings of the third symposium on Information interaction in context*, IIiX '10, pages 125–134, New York, NY, USA, 2010. ACM.

[16] M. Lykke, B. Larsen, H. Lund, and P. Ingwersen. Developing a test collection for the evaluation of integrated search. In C. Gurrin, Y. He, G. Kazai, U. Kruschwitz, S. Little, T. Roelleke, S. M. Rüger, and K. van Rijsbergen, editors, *ECIR*, volume 5993 of *Lecture Notes in Computer Science*, pages 627–630. Springer, 2010.

[17] M. F. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[18] T. Saracevic and P. Kantor. A study of information seeking and retrieving. iii. searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39(3):197–216, 1988.

[19] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[20] P. Smets. *What is Dempster-Shafer's model?* Wiley, 1994.

[21] D. R. Sorensen, T. Bogers, and B. Larsen. An exploration of retrieval-enhancing methods for integrated search in a digital library. In B. Larsen, C. Lioma, and A. P. de Vries, editors, *TBAS2012*, pages 4–8. http://ceur-ws.org, 2012.

[22] H. R. Turtle and W. B. Croft. Evaluation of an inference network-based retrieval model. *ACM Trans. Inf. Syst.*, 9(3):187–222, 1991.

[23] E. M. Voorhees and D. M. Tice. The trec-8 question answering track evaluation. In *TREC*, 1999.

[24] C. Zhai and J. D. Lafferty. Two-stage language models for information retrieval. In *SIGIR*, pages 49–56. ACM, 2002.