

Can user and task characteristics be used as predictors of success in health information retrieval sessions?

Melinda Oroszlányová, Carla Teixeira Lopes, Sérgio Nunes and Cristina Ribeiro

Introduction. The concept and study of relevance has been a central subject in information science. Although research in information retrieval has been focused on topical relevance, other kinds of relevance are also important and justify further study. Motivational relevance is typically inferred by criteria such as user satisfaction and success.

Method. Using an existing dataset composed by an annotated set of health Web documents assessed for relevance and comprehension by a group of users, we build a multivariate prediction model for the motivational relevance of search sessions.

Analysis. The analysis was based on lasso variable selection, followed by model selection using multiple logistic regression.

Results. We have built two regression models; the full model, which considers all variables of the dataset, has a lower estimated prediction error than the reduced model, which contains the statistically-significant variables from the full model. The higher values of evaluation metrics, including accuracy, specificity and sensitivity in the full model support this finding. The full model has an accuracy of 91.94%, and is better at predicting motivational relevance.

Conclusions. Our findings suggest features that can be considered by search engines to estimate motivational relevance, to be used in addition to topical relevance. Among these features, a high level of success in Web search and in health information search on social networks and chats are some of the most influencing user features. This shows that users with higher computer literacy might feel more satisfied and successful after completing the search tasks. In terms of task features, the results suggest that users with clearer goals feel more successful. Moreover, results show that users would benefit from the help of the system in clarifying the retrieved documents.

Introduction

Information on health topics started to appear on the Web during the 1990s ([Pallen, 1995](#)). With the increased availability of health information, health search on the Web started to have an impact on people's health care routines ([Fox and Rainie, 2002](#); [Fox, 2006](#); [Fox, 2011](#); [Espanha and Villanueva, 2008](#)). Regarding health information, it has been observed that the Internet is a popular source of information ([Kim, 2009](#); [Savolainen, 2008](#)). A survey reported that, in 2012, 72% of all adults looked online for health information ([Fox and Duggan, 2013](#)). Users' habits concerning their Web and health search have been studied in several surveys ([Espanha and Villanueva, 2008](#); [Fox and Rainie, 2002](#); [Fox, 2006](#); [Fox, 2011](#); [Fox and Duggan, 2013](#)).

Relevance represents a key concept for evaluating the effectiveness of information retrieval. Although there are several types of relevance, most of the attention has been given to topical relevance, that is, the 'relation between the topic expressed in a query, and topic covered by retrieved texts' ([Saracevic, 1996](#), p. 12). Fewer studies have focused on motivational relevance, defined by

Saracevic as the '*relation between the intents, goals, and motivations of a user, and texts retrieved by a system or in the file of a system, or even in existence; satisfaction, success, accomplishment, and the like being criteria for inferring motivational relevance*' (Saracevic, 1996, p. 12).

The importance of motivational relevance is visible in the definition above. Since the aim of any retrieval system is to fulfil the user's goals, we think this kind of relevance should not be disregarded. The objective of the present study is to analyse which characteristics influence the satisfaction of users during health information-seeking, with the help of a set of annotated Web pages that were assessed by users in a user study (Lopes, 2013) during which they were assigned a set of search tasks. We want to address the following major research question: How can motivational relevance, that is whether a user feels successful after the search tasks or not, be predicted? More specifically, we want to determine whether (1) it can be predicted through user characteristics such as their experience with Web search, how frequently they conduct health searches or their habits regarding the terminology used during health query formulation; (2) it can be predicted through task features such as the level of clarity, simplicity and familiarity of the tasks for the users or by previous searches conducted by the user about the given tasks. Our goal is to find good descriptors and potential predictors of motivational relevance.

Literature review

The notion of relevance has been studied for decades, and the retrieval of relevant information has become the main concern of any information retrieval system (Manning, et al., 2009). Relevance is estimated by search engines and is considered '*a measure of the effectiveness of an interaction between a source and a destination*' (Saracevic, 1975, p. 321).

The exhaustive literature review by Mizzaro (1997) introduces and discusses the concepts, types and history of relevance in information retrieval, summarising the research studies from 1959 to 1996. He gives references to the origins of the problem of finding relevant information before 1958. Regarding motivational relevance, there are references to user satisfaction and the success of the search as judged by the user, as considered measures for the evaluation of an information retrieval system (Su, 1991; 1992; 1994). O'Connor (1968) discusses the types of relevance in terms of satisfying users' information needs. Sandore's (1990) discovery of a correlation between user satisfaction and the precision of their search is also pointed out in Mizzaro's study (1997), as well as other features that affect user satisfaction such as interaction with the intermediary and the library's location (Tessier et al., 1977). Schamber (1994) also studies user satisfaction in her literature review from 1983 to 1994. Reviews the literature on the themes of behaviour, measurement and terminology, she proposes three different views of relevance (system, information, and situation views) on the basis of a classical IR interaction model; discusses recall, precision, utility, and satisfaction; describes the factors affecting relevance judgments; reports recent results of the criteria identified by the users; sketches the interdisciplinary models and the theoretical approaches to relevance, and discusses some methodological problems.

Saracevic's first literature review of relevance from 1975 (Saracevic, 1975), was followed by two more of his research works focusing on the nature and manifestations of relevance (Saracevic, 2007) and its behaviour and effects (Saracevic, 2007). These works distinguish between user and system relevance.

System (or algorithmic) relevance is defined as the *'relation between a query and information objects in a given system, where the aim is to retrieve a set of information objects that the system inferred as being relevant to a query'* (Saracevic, 1996, p. 12). User relevance is subjective and dependent on the user and context. It is divided into four major categories: topical, cognitive, situational and motivational (Saracevic, 1996). Topical (or subject) relevance is defined as the *'relation between the topic expressed in a query, and topic covered by retrieved texts'* (Saracevic, 1996, p. 12). Cognitive relevance (or pertinence) is the *'relation between the state of knowledge and cognitive information need of a user, and texts retrieved'* (Saracevic, 1996, p. 12). Situational relevance (or utility) is the *'relation between the task and texts retrieved by a system'* (Saracevic, 1996, p. 12). Finally, motivational (or affective) relevance is defined as the *relation between the intents, goals, and motivations of a user, and texts retrieved by a system or in the file of a system, or even in existence; satisfaction, success, accomplishment, and the like being criteria for inferring motivational relevance* (Saracevic, 1996, p. 12).

Dumais (2012) summarises various studies on the evaluation of interactive information retrieval, referring to Cole *et al.* (2009) with respect to motivational relevance. The authors, focusing on the evaluation of interactive information retrieval systems, propose *usefulness* as a measure of system performance. Belkin (2010) also suggests that usefulness could be an appropriate criterion for evaluating interactive information retrieval systems. In a previous study, we showed that user and task characteristics are good descriptors and possible predictors of motivational relevance (Oroszlányová, 2015). Relevance assessment was shown to be influenced by query, document characteristics, user and task (e.g., age, health search experience, task clarity) (Lopes, 2010).

Method

This work studies both how and to what extent task and user characteristics are useful in predicting motivational relevance. With the help of a multivariate prediction model for the motivational relevance of search sessions, we analyse which user and task characteristics influence the motivational relevance of health Web documents. A detailed description of the user and task characteristics is provided in the following subsection.

Description of the dataset

The present study is based on an existing dataset composed by a sample of 4533 annotated health Web documents. This set of documents was initially collected for a user study (Lopes, 2013), where forty participants performed eight tasks, associated with different health information-seeking situations, based on questions submitted to the health category of the Yahoo! Answers service. From the list of open questions of this category, starting with the most popular, eight questions about treatments for a symptom/disease were selected. For each question, four different search queries were defined, two in English and two in the participant's native language. In each language, the two queries were formulated by using lay and medico-scientific terminology, respectively. Queries were built by concatenating the eight symptoms or diseases (painful urination/dysuria; head itching/head pruritus; high uric acid/hyperuricaemia; mouth inflammation/stomatitis; bone infection/osteomyelitis; heartburn/pyrosis; hair loss/alopecia; joint

pain/arthritis) with the word *treatment*, with different medical terminology (lay/medico-scientific). To reduce the risk of Google learning from the previously-submitted queries, it was ensured that returned links were never clicked. Further, to prevent changes in the search engine, all queries were submitted within a very short time span. For each query, the top-thirty results were collected. The documents were assessed by a set of Information Science students in terms of relevance and comprehension using a three-valued scale.

For these documents, a metadata scheme was defined and used for a latter annotation with manual and automatic approaches (Sousa, 2011). For instance, the specificity of the vocabulary (related to technical and scientific terms of the health area) was one of the document characteristics, and it was evaluated by annotating a value according to the defined scale: 1: barely perceptible; 2: perceptible; 3: completely perceptible. To evaluate the quality of the annotation, 10% of the documents were also assessed by an external health professional (Sousa, 2011). The agreement rate between both assessments was very good (93%), thus, the way the characteristics were evaluated/annotated was, in general, well defined.

Our study includes only user and task variables plus six aggregated variables containing the number of documents in each session assessed as non-relevant (nrel0), partially relevant (nrel1), totally relevant (nrel2), not understood (ncomp0), partially understood (ncomp1) and completely understood (ncomp2). In the present work, motivational relevance is assessed by a question posed at the end of the search session. In this question, users were asked to evaluate their feeling of success with the task in a five-level scale (1-extremely unsuccessful; 5-extremely successful). The metadata scheme that was used to annotate the dataset contains specific characteristics of tasks and users, listed in Table 1. Task-related characteristics include users' feedback on the tasks' clarity, simplicity and familiarity. User characteristics describe the user in terms of their health literacy, Web search and health search experience.

Characteristics/Variables	Scale
<i>Tasks</i>	
Correct answers in the task	0-No 1-Yes
The user had an exact idea about the information in the tasks	1 (Disagree) to 5 (Agree)
Level of clarity, simplicity and familiarity of the tasks for the users	1 (Unclear/Easy/Unfamiliar) to 5 (Clear/Complex/Familiar)
Language of the query	Nominal
Medical terms in the query	0-No 1-Yes
Previous search by the user about the given tasks	0-No 1-Yes
Whether the users knew the technical terms	0-No 1-Yes
Number of totally relevant documents	Continuous
Number of partially relevant documents	Continuous
Number of non-relevant documents	Continuous
Number of completely understood documents	Continuous
Number of partially understood documents	Continuous
Number of not understood	Continuous

documents	
<i>Users</i>	
English proficiency of the users	Continuous
Health literacy of the users	Continuous
Number of medical concepts included in the query, that the user knows	Continuous
Age of the users	Continuous
Gender of the users	Nominal
Health status of the users	1 (Not healthy) to 5 (Very healthy)
Experience of the users with Web search and with health search	Continuous
Frequency of the users' Web search and health search	1 – Once a year
	2 – Once a month
	3 – Once a week
	4 – Once a day
	5 – More often
	NULL – No response provided
Success of the users with Web search and health search	1 (Never) to 5 (Always)
Health search in Portuguese, English and other language	1 (Never) to 5 (Frequently)
Use of medico-scientific terminology during Web searches about health subjects	1 (Never) to 5 (Always)
Level of satisfaction of the users' health information need on Web pages, blogs, forums, social networks, chats, newsletter and RSS feeds	1 (Never) to 5 (Frequently)

Table 1: Task and user characteristics used for predicting motivational relevance.

Statistical analysis

In the subsequent section Multivariate analysis of motivational relevance, we analyse multiple variables from our data collection in relation to motivational relevance. First, we selected the variables to produce a model that best fits our data, using the least absolute shrinkage and selection operator (lasso). The model selects the best subset of predictors by shrinking the regression coefficients towards zero, and estimates their coefficients based on logistic regression (James, et al., 2013). The logistic regression models the probability of motivational relevance given the task and user characteristics/variables. We can write it as probability (relevance = yes|characteristics), where the probability values $p(\text{characteristics})$ range between 0 and 1.

Originally, our model had a multinomial distribution with five relevance levels (0, 1, 2, 3, 4 and 5). Here we merged motivational relevance levels 1, 2, 3 and 4, 5, inducing a binomial distribution of the model.

After the lasso variable selection, we included the chosen characteristics in the multiple logistic regression model and used leave-one-out cross-validation (LOOCV) to estimate the accuracy of the model (James, et al., 2013). The LOOCV error rate is estimated by averaging the misclassified observations over the total number of observations.

The LOOCV approach splits the set of observations into a single observation

used for the validation set, and into the remaining observations which form the training set, where the prediction is made for the former observation. We also built a second model, which we call reduced model. It contains only the variables which were significant in the full model. Finally, we compared the LOOCV estimates of prediction (or test) errors for the two models.

Results

Multivariate analysis of motivational relevance

This section describes how we built the prediction models, the models themselves, and their evaluation, with the aim of predicting whether or not a user feels successful after the search tasks, based on user and task characteristics. Following the statistical strategy defined in the previous section (Statistical analysis), after classifying tasks as successful (fourth and fifth level of the scale) or not (first, second and third levels), we built and evaluated two logistic regression models for motivational relevance. We built the reduced model to analyse whether we could reach similar results using a lower number of features, which would be advantageous for practical reasons.

Full model

The first model considers all variables. The analysis began by fitting a lasso model on the training set. In the next step we chose the best tuning parameter using cross-validation and used it to fit the lasso model on the full dataset (section Model definition process). Finally, with the variables selected by the lasso model, we fitted a multiple logistic regression model (section Logistic regression model) and evaluated the results (section Evaluation).

Model definition process

Applying the lasso to our dataset and using the potential predictor variables discussed in the section Method, we built a model predicting the motivational relevance. The lasso, with the minimal tuning parameter chosen by cross-validation, yielded a prediction model containing candidate variables to be analysed with the multiple logistic regression model.

Logistic regression model

The lasso helped with variable selection, and we continued the analysis with model selection using logistic regression. The resulting variables from the lasso model were added to the multiple logistic regression model which is summarized in Table 2. The first column contains the category of the feature, where T and U refer to task and user, respectively. The variables are listed in the second column, where the numbers in the parentheses indicate the levels of the variables (according to the scales defined in Table 1). The variables' corresponding estimated coefficients are in the third column. The fourth column lists the standard error when assessing the accuracy of the coefficient estimates. In the fifth column, the z-statistics are listed, and their large (absolute) value indicates evidence against the null hypothesis of the coefficients being equal to zero. The last column contains the corresponding p-values. Those marked in bold relate to the variables that are statistically significant at $\alpha = 0.05$ in the full regression model.

Cat.	Variable	Estimate	Std. Error	Z-score	Pr(> z)
T	Is the user familiar with the task? (5)	19.534	393.867	0.050	0.960
T	Is the task clear? (5)	18.549	206.990	0.090	0.929
T	Is the task easy? (5)	13.390	230.042	0.058	0.954
T	Is the task easy? (4)	3.572	0.101	35.366	5.60E-274
T	Does the user have an idea about the information? (5)	3.567	0.378	9.441	3.69E-21
T	Is the task clear? (4)	2.018	0.090	22.487	5.55E-112
T	Does the user have an idea about the information? (4)	0.383	0.079	4.846	1.26E-06
T	Did the user answer the task correctly? (1)	0.308	0.064	4.828	1.38E-06
T	Did the user answer the task correctly? (2)	0.276	0.110	2.512	0.012
T	Did the user find the document relevant? (2)	0.036	0.003	13.181	1.13E-39
T	Is the task clear? (2)	-0.438	0.114	-3.833	1.26E-04
T	Is the task easy? (2)	-1.328	0.094	-14.094	4.12E-45
T	Does the user have an idea about the information? (2)	-2.405	0.121	-19.928	2.31E-88
U	Is the user healthy? (4)	20.195	697.783	0.029	0.977
U	Is the user healthy? (5)	17.961	697.783	0.026	0.979
U	Is the user healthy? (3)	17.438	697.783	0.025	0.980
U	Does the user health search in chats? (4)	5.705	0.211	27.037	5.48E-161
U	Does the user health search in chats? (3)	4.479	0.263	17.050	3.52E-65
U	Does the user health search in social networks? (5)	4.400	0.294	14.943	1.74E-50
U	Is the user successful in web search? (5)	3.406	0.218	15.636	4.17E-55
U	Does the user health search frequently? (2)	1.007	0.102	9.842	7.42E-23
U	Does the user health search in newsletters? (2)	0.292	0.096	3.030	2.44E-03
U	Does the user health search in	-0.557	0.105	-5.279	1.30E-07

	blogs? (2)				
U	Experience of the user with web search (in years)	-0.663	0.021	-31.868	7.34E-223
U	Does the user health search in social networks? (4)	-0.764	0.235	-3.255	1.14E-03
U	Does the user health search on webpages? (4)	-0.821	0.112	-7.328	2.34E-13
U	Does the user health search on webpages? (5)	-1.141	0.138	-8.249	1.60E-16
U	Does the user health search in social networks? (2)	-1.339	0.121	-11.030	2.74E-28
U	Does the user health search on webpages? (2)	-2.072	0.281	-7.370	1.71E-13
U	Does the user know health search terminology? (2)	-2.275	0.157	-14.463	2.07E-47
U	Does the user know health search terminology? (5)	-2.376	0.177	-13.457	2.79E-41
U	Does the user health search frequently? (NULL)	-2.718	0.149	-18.186	6.64E-74
U	Does the user health search in RSS feeds? (3)	-3.252	0.154	-21.177	1.54E-99
U	Does the user health search in newsletters? (4)	-3.584	0.182	-19.685	2.89E-86
	LOOCV estimate of prediction error	0.067			

Table 2: Summary of the coefficient estimates in the full model. Bold values show variables statistically significant at 0.05. .

Evaluation

Our regression model was verified by leave-one-out cross-validation, and its results are reported in the last row of Table 2. The p-values associated with the variables, marked with bold in Table 2, are statistically significant at $\alpha = 0.05$. The negative coefficients indicate that users with the corresponding features/variables are less likely to feel successful after the search task than the users without these characteristics, for fixed values of the remaining variables. Variables with large coefficient estimates highlight the importance of such variables (e.g. high level of success in Web search, and in health information search on social networks and chats, or clearly given/defined search tasks) for motivational relevance. To assess the accuracy of the model, we have fitted the model using half of the data (a training dataset), and then examined how well it predicts the held-out data (a test dataset) as explained by James *et al.* (2013). Using the test dataset, we then computed the probabilities of the document being relevant, allowing us to compute the accuracy, sensitivity and specificity of the model. Given these predictions, we determined how many observations

were correctly or incorrectly classified. Our logistic regression has an accuracy of 91.94%, a specificity (true negative rate) of 89.71% and sensitivity (true positive rate) of 93.17%. The LOOCV estimate of prediction error from Table 2 is very low (0.067), meaning that the regression model is of high accuracy.

Reduced model

We built a second model, including only the statistically significant variables from the full model. In this second model, all variables remained significant except the one pertaining to the second level of health search in social networks. Table 3 shows the coefficient estimates for the reduced logistic regression model, i.e., for a logistic regression model that uses the selected 34 variables to predict the probability of a user feeling successful or not after completing the search task. The p-values marked with bold indicate that the corresponding variables are associated with motivational relevance and are statistically significant at $\alpha = 0.05$ in the reduced regression model. The model's accuracy was assessed using leave-one-out cross-validation, with estimated prediction error of 0.0913. The logistic regression has an accuracy of 89.03%, a specificity of 86.27% and sensitivity of 90.53%. The LOOCV estimate of prediction error for this model is higher than the error estimate for the full regression model in Table 2.

Cat.	Variable	Estimate	Std. Error	Z-score	Pr(> z)
T	Does the user have an idea about the information? (5)	5.304	0.335	15.828	1.98E-56
T	Is the task easy? (4)	3.767	0.087	43.062	0.000
T	Does the user have an idea about the information? (4)	2.118	0.064	32.892	2.81E-237
T	Is the task clear? (4)	1.002	0.076	13.271	3.42E-40
T	Did the user answer the task correctly? (1)	0.148	0.050	2.945	3.23E-03
T	Did the user find the document relevant? (2)	0.047	0.002	19.668	4.05E-86
T	Is the task clear? (2)	-0.720	0.090	-7.962	1.69E-15
T	Is the task easy? (2)	-0.887	0.071	-12.406	2.44E-35
T	Does the user have an idea about the information? (2)	-1.786	0.095	-18.780	1.11E-78
U	Does the user health search in chats? (4)	4.863	0.132	36.961	4.89E-299
U	Does the user health search in social networks? (5)	4.285	0.230	18.654	1.17E-77
U	Is the user successful in web search? (5)	3.679	0.160	22.941	1.83E-116
U	Does the user health search in chats? (3)	2.068	0.190	10.856	1.88E-27
U	Does the user health search frequently? (2)	0.367	0.073	5.044	4.55E-07
U	Does the user health search in newsletters? (2)	0.348	0.088	3.966	7.32E-05
U	Does the user health search in blogs? (2)	-0.168	0.071	-2.354	0.019

U	Does the user health search in social networks? (2)	-0.185	0.095	-1.954	0.051
U	Experience of the user with web search (in years)	-0.703	0.019	-36.368	1.35E-289
U	Does the user health search on webpages? (5)	-1.160	0.120	-9.655	4.67E-22
U	Does the user know health search terminology? (5)	-1.259	0.160	-7.852	4.09E-15
U	Does the user health search on webpages? (4)	-1.488	0.100	-14.806	1.33E-49
U	Does the user know health search terminology? (2)	-2.008	0.150	-13.378	8.09E-41
U	Does the user health search in social networks? (4)	-2.487	0.149	-16.650	3.02E-62
U	Does the user health search in RSS feeds? (3)	-2.777	0.121	-22.992	5.61E-117
U	Does the user health search in newsletters? (4)	-3.160	0.146	-21.641	7.32E-104
U	Does the user health search on webpages? (2)	-3.176	0.208	-15.259	1.43E-52
U	Does the user health search frequently? (NULL)	-3.604	0.121	-29.698	8.23E-194
	LOOCV estimate of prediction error	0.091			

Table 3: Summary of the coefficient estimates in the reduced model.

Discussion

As expected, the best model to predict motivational relevance is the one containing all variables suggested by lasso, although the reduced model was very close in terms of error rates and has the advantage of not requiring as much information. In Table 4 we summarize the evaluation metrics of both models. The first row contains the number of variables included in each model. Looking at the second row we can observe that the full model has the lowest prediction error estimate (LOOCV error). The higher values of accuracy, specificity and sensitivity in the full model also support this finding.

	Full model	Reduced model
Number of variables	39	27
LOOCV error	6.70%	9.13%
Accuracy	91.94%	89.03%
Specificity	89.71%	86.27%
Sensitivity	93.17%	90.53%

Table 4: Comparison of the full and reduced logistic regression models in terms of number of variables and evaluation rates. Figures in bold show the best performance values in each row.

The suggested models let us answer our research questions and were useful in understanding which characteristics are more relevant when estimating whether a user feels successful after the search tasks. The characteristics that

significantly contribute to the prediction of motivational relevance, either positively or negatively, might be important for this purpose. For example, search engines might use this information to improve their performance.

The production and analysis of these models allowed us to identify important user and task features to estimate motivational relevance. We found that users who frequently conduct health search on chats or social networks feel more successful after completing the search tasks. Users who consider themselves healthy succeeded better in their search task. We ponder whether healthier people are less demanding than less-healthy people on health information seeking. The more successful a user is in Web search, the more successful s/he feels in completing the search tasks.

As expected, users feel more successful when they find more totally-relevant documents. This is reasonable, because when the user encounters useful information, s/he will feel more satisfied with the completion of his search task. It did not surprise us to discover that having an initial idea about what to search for also contributes in increasing the users' feeling of success with the search task. Users who gave appropriate or somewhat appropriate answers to the information need after the search session are considered more successful in completing the health search tasks. This shows that users are somehow aware of the accuracy of the obtained knowledge.

On the other hand, the analyses show that users who generally conduct less frequent health search in RSS feeds, newsletters, blogs, and Web pages tend to feel less successful in completing their search tasks. We discovered that the more frequently users use medico-scientific terminology during their Web searches on health subjects, the less they will feel satisfied with their search task. This result is aligned with the findings of Saracevic (2007, p. 2136) which state that *'lesser subject expertise seems to lead to more lenient and relatively higher relevance ratings'*. Considering the number of years users have been searching the Web, we also hypothesize that older users are less successful with their health search tasks. This is aligned with our suspicion that people with more health problems might be more demanding.

The findings of the present study revealed that, besides the above user characteristics, some of the task features such as the clarity and simplicity of the search task are also useful for estimating motivational relevance.

Clear and easy tasks positively contribute to the success of users within the search session. These results support our previous findings, where we observed significant positive association between the clarity and simplicity of the search tasks, user familiarity with them, their information idea, success in Web search, respectively, and motivational relevance (Oroszlányová, 2015).

There are many variables with multiple levels, which might suggest that regardless of the high values of estimates, some of the features included in the model might be less useful in real situations. Therefore, in practice, it might be useful to consider only one level at once. For instance, we might prefer the fourth level to the third one for the variable health search on chats, because its estimate is higher or because we want to make predictions for users who use chats more frequently, or we might prefer to exclude the variables with no available response (Does the user frequently conduct health search? (NULL))

Conclusions

We conducted a multivariate analysis focused on how and to what extent user and task characteristics are useful in predicting motivational relevance. For this purpose, we built two regression models. Our best model, the full model, had a LOOCV estimate of prediction error of 6.70%, a sensitivity of 93.17%, an accuracy of 91.94%, and a specificity of 89.71%. Several of the features used to estimate motivational relevance are related to users' search habits, and to characteristics of the task such as their clarity and simplicity.

Among the variables which were identified to predict motivational relevance, a high level of success in Web search, and in health information search on social networks and chats are some of the most influencing user variables. This might show that users with higher abilities in both Web and health search might feel more satisfied and successful after completing the search tasks. In terms of tasks features, our findings suggest that users need help with performing the search tasks and with clarifying their goals and retrieved information to feel more satisfied with the search tasks.

The present study resulted in several potential predictors of motivational relevance. These features can be useful in improving the estimation of motivational relevance, particularly in the domain of health Web search. The application of our models to other datasets might be also interesting for the generalisation of our results. As future work, we aim to use the findings of this study to improve information retrieval in the health domain, for example making the search tasks clearer and easier. Since some of the features included in these models can be assessed automatically, they can be useful to improve the estimation of motivational relevance by search engines, particularly on health Web documents. Having found that the more successful a user is in Web search, the more successful he feels in completing the search tasks, we also would like to study the relationship between personality, experience and motivational relevance in more detail. For instance, will positive and confident people naturally feel more successful in Web search or in completing a health search task? We will also work on the development of methods to automatically detect these features.

Acknowledgments

This work was supported by Project “NORTE-01-0145- FEDER-000016” (NanoSTIMA), financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

About the authors

Melinda Oroszlányová - Department of Informatics Engineering (DEI), Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal - melinda@fe.up.pt

Carla Teixeira Lopes - Department of Informatics Engineering (DEI), Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal; INESC TEC, Porto, Portugal - ctl@fe.up.pt

Sérgio Nunes - Department of Informatics Engineering (DEI), Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal; INESC TEC, Porto, Portugal - ssn@fe.up.pt

Cristina Ribeiro - Department of Informatics Engineering (DEI), Faculty of Engineering of the University of Porto (FEUP), Porto, Portugal; INESC TEC, Porto, Portugal - mcr@fe.up.pt

References

- Belkin, N. (2010). On the evaluation of interactive information retrieval systems. In B.Larsen, J.W. Schneider & F. Aström (Eds.) *The Janus faced scholar. A festschrift in honour of Peter Ingwersen*, s.n. (pp. 13-22) Copenhagen: Royal School of Library and Information Science.
- Cole, M., Liu, J., Belkin, N.J., Bierig, R., Gwizdka, J., Liu, C., Zhang, J. & Zhang, X. (2009). Usefulness as the criterion for evaluation of interactive information retrieval. In, *Proceedings of the third Workshop on Human-Computer Interaction and Information Retrieval* (pp. 1-4). Washington, DC: Catholic University of America.
- Dumais, S. (2012). [Whole-session evaluation of interactive information retrieval systems](#), *NII Shonan Workshop*, Oct 8-12. Retrieved from <https://pdfs.semanticscholar.org/dac0/626c7f2191f505d1164da62916493f6b9eee.pdf> (Archived by WebCite® at <http://www.webcitation.org/72ESpd8gW>)
- Espanha, R. & Lupiáñez-Villanueva, F., (2008). [Health and the Internet: autonomy of the user](#). Lisbon: Lisbon Internet and Networks. Retrieved from http://www.lini-research.org/np4/?newsId=11&fileName=RESPANHA_FLUPPIANEZ_VILLANUEVA_WP6.pdf (Archived by WebCite® at <http://www.webcitation.org/72ESzUPvN>)
- Fox, S. (2006). [Online health search](#). Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf (Archived by WebCite® at <http://www.webcitation.org/72ETC6V7N>)
- Fox, S. (2011). [The social life of health information](#). Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Social_Life_of_Health_Info.pdf (Archived by WebCite® at <http://www.webcitation.org/72ETN6nDf>)
- Fox, S. & Duggan, M. (2013). [Health online 2013](#). Washinton, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/PIP_HealthOnline.pdf (Archived by WebCite® at <http://www.webcitation.org/72ETWM7ZA>)
- Fox, S. & Rainie, L. (2002). [Vital decisions: a Pew Internet health report](#). Washington, DC: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/~//media//Files/Reports/2002/PIP_Vital_Decisions_May2002.pdf.pdf (Archived by WebCite® at <http://www.webcitation.org/72ETh1dBf>)
- James, G. Witten, D. Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. New York, NY: Springer.
- Kim, J. (2009). Describing and predicting information-seeking behavior on the Web. *Journal of the American Society for Information Science and Technology*, 60(4), 679-693.
- Lopes, C. T. & Ribeiro, C., (2010). Context effect on query formulation and subjective relevance in health searches. In, *IiiX'10 Proceedings of the third symposium on Information interaction in context*, (pp. 205-214) New York, NY: ACM.
- Lopes, C. T. & Ribeiro, C., (2013). Measuring the value of health query translation: An analysis by user language proficiency. *Journal of the American Society for Information Science and Technology*, 64(5), 951-963.
- Manning, C. Raghavan, P. & Schütze, H. (2009). *Introduction to information retrieval*. Cambridge: Cambridge University Press.
- Mizzaro, S. (1997). Relevance: the whole history. *Journal of the American Society for Information Science and Technology*, 48(9), 810-832.
- O'Connor, J. (1968). Some questions concerning "information need". *American Documentation*, 19(2), 200-203.

- Oroszlányová, M., Lopes, C. T., Nunes, S. & Ribeiro, C., (2015). The influence of documents, users and tasks on the relevance and comprehension of health web documents. *Procedia Computer Science*, 64, 771-778.
- Pallen, M. (1995). Guide to the Internet. The world wide web. *British Medical Journal*, 311(7019), 1552-1556.
- Sandore, B. (1990). Online searching: what measures satisfaction? *Library and Information Science Research*, 12(1), 33-54.
- Saracevic, T. (1975). Relevance: a review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26(6), 321-343.
- Saracevic, T. (1996). Relevance reconsidered. In, *Proceedings of the 2nd International Conference on Conceptions of Library and Information Science: Integration in Perspective (CoLIS2)*, (pp. 201-218). Copenhagen: Royal School of Librarianship.
- Saracevic, T. (2007). Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 1915-1933.
- Saracevic, T. (2007). Relevance: a review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 2126-2144.
- Savolainen, R. (2008). Source preferences in the context of seeking problem-specific information. *Information Processing and Management*, 44(1), 274-293.
- Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology*, 29, 3-48.
- Sousa, M. I., (2011). *Characterization of health web documents*. Unpublished master's dissertation, University of Porto, Porto, Portugal.
- Su, L.T. (1991). *An investigation to find appropriate measures for evaluating interactive information retrieval*, Unpublished doctoral dissertation, Rutgers, the State University of New Jersey, New Brunswick, N.J., USA
- Su, L.T. (1992). Evaluation measures for interactive information retrieval. *Information Processing & Management*, 28(4), 503-516.
- Su, L.T. (1994). The relevance of recall and precision in user evaluation. *Journal of the American Society for Information Science*, 45(3), 207-217.
- Tessier, J.A., Crouch, W.W. & Atherton, P. (1977). New measures of user satisfaction with computer-based literature searches. *Special Libraries*, 68(11), 383-389.

How to cite this paper

Oroszlányová, M., Teixeira Lopes, C., Nunes, S., Ribeiro, C. (2018). Can user and task characteristics be used as predictors of success in health information retrieval sessions? *Information Research*, 23(3), paper 801. Retrieved from <http://InformationR.net/ir/23-3/paper801.html> (Archived by WebCite® at <http://www.webcitation.org/72Nq9kfP0>)

Find other papers on this subject

Check for citations, [using Google Scholar](#)

© the author, 2018.

72 Last updated: 12 September, 2018

[Contents](#) | [Author index](#) | [Subject index](#) | [Search](#) |
[Home](#)
