Analyzing Query Reformulation Data using Multi-level Modeling

Kun Lu[1], Soohyung Joo[2], Taehun Lee[1], Rong Hu[3]
[1] University of Oklahoma
[2] University of Kentucky
[3] Jiang Xi University of Finance and Economics

Abstract
This study explores the adoption of multi-level modeling to analyze query reformulation data. Thus far, the dependency among query reformulations within the same search session has not been adequately treated in the experimental design. This has limited the analysis of users' query behavior. This study introduces multi-level modeling to query reformulation data analysis. Multi-level modeling is capable of handling the correlations among query reformulations and provides an avenue to analyzing the nested data structure. A demonstration of fitting query reformulation data to two types of multi-level models is provided. The method introduced in this study provides a potential solution to the analysis of query reformulations.
Keywords: Query reformulation; multi-level modeling; data analysis; interactive information retrieval
doi: 10.9776/16473

Contact: kunlu@ou.edu

1 Introduction

Understanding users' query behavior has been one of the important topics in the field of information retrieval. Early studies examined the characteristics of user queries based on transaction logs (Jansen *et al*., 2000; Spink *et al*., 2001). Also, investigators looked at how users reformulate their queries (Rieh & Xie, 2006; Jansen *et al.*, 2007; Jansen *et al*., 2009), and more recent studies have explored the effects of contextual factors (e.g., search task types, cognitive status, domain knowledge, and search skills) on users' query reformulation behavior (Liu *et al.*, 2010; Hu *et al.*, 2013). When analyzing query reformulations, one hurdle that impedes the adoption of traditional statistical analysis methods has been that query reformulations are nested within a search session, and thus they are prone to be correlated with each other. Traditional statistical analysis methods, such as regression or analysis of variance (ANOVA), require the observations to be independent with each other. This has limited the use of inferential statistics in query reformulation studies. For example, Joo and Lee (2011) analyzed only the first occurring reformulation types to avoid the dependency problem among query reformulations when using ANOVA. Liu *et al.* (2010) viewed each individual reformulation as a unit of analysis, and disregarded the dependency among reformulations in the analysis. Similarly, Hu *et al.* (2013) analyzed query reformulation without considering the session effect by aggregating query reformulations across sessions. As is shown in these examples, researchers have had difficulty in controlling for the inter-dependency among observations in the analysis of query reformulations.

Multi-level modeling can be a compelling solution to analyzing this type of nested data structure (Gelman & Hill, 2007). In the context of multi-level modeling, query reformulations can be considered as level one variable (individual level) that is nested within search sessions (level two variable, or group level). It is capable of handling the correlations among query reformulations within each search session. The purpose of this study is to introduce multi-level modeling to query reformulation data analysis. The study will demonstrate how different types of multi-level models can be applied to analyze query reformulation data.

2 Multi-level Modeling

The history of multi-level modeling goes back to the work of Robinson (1950), which discussed the fallacy of ecological correlation, the papers by Blau (1960) and Davis et al (1961) on the contextual effects, and the paper by Eisenhart (1947) on the distinction between fixed and random effects. After their seminal papers, there has been a bloom of methodological articles aiming at showing that serious inferential errors may result from applying fixed parameter regression models to the analysis of hierarchically structured data (Laird & Ware, 1982; Blalock, 1984; Aitkin & Longford, 1986; Bryk & Raudenbugh, 1987; Kenny & Judd, 1986; Hoffman, 1997). Behavioral and social sciences data commonly have hierarchically structured systems. For example, workers are nested within firms in organizational/management research, and students are nested within schools in educational research. In the analysis of such data, multi-level models provide a set of integrated methods for modeling dependence due to the hierarchically nested structures.

3 Data Analysis

3.1 Experimental Data

The data was collected from a user study in which forty five subjects were recruited from a state university in the United States to test two different search interfaces. One search interface (SimpleMed) is similar to Google with a simple search box and users can click a search result to pop up the full document. The other search interface (MeshMed) provides additional components, the MeSH tree browser and the MeSH term browser, to allow users to interact with MeSH thesaurus while searching (Figure 1). The tree browser displays the hierarchical structure of MeSH that allows a user to navigate the thesaurus. The term browser allows a user to search related MeSH terms and their definitions with natural language queries. The data corpus used in the experiment was Ohsumed test collection (http://ir.ohsu.edu/ohsumed/). Six search topics were randomly selected from the 106 topics in Ohsumed. Each participant searched six topics on two search systems: three on SimpleMed and three on MeshMed. The sequences of the search topics and search systems were randomized to balance the learning effects.

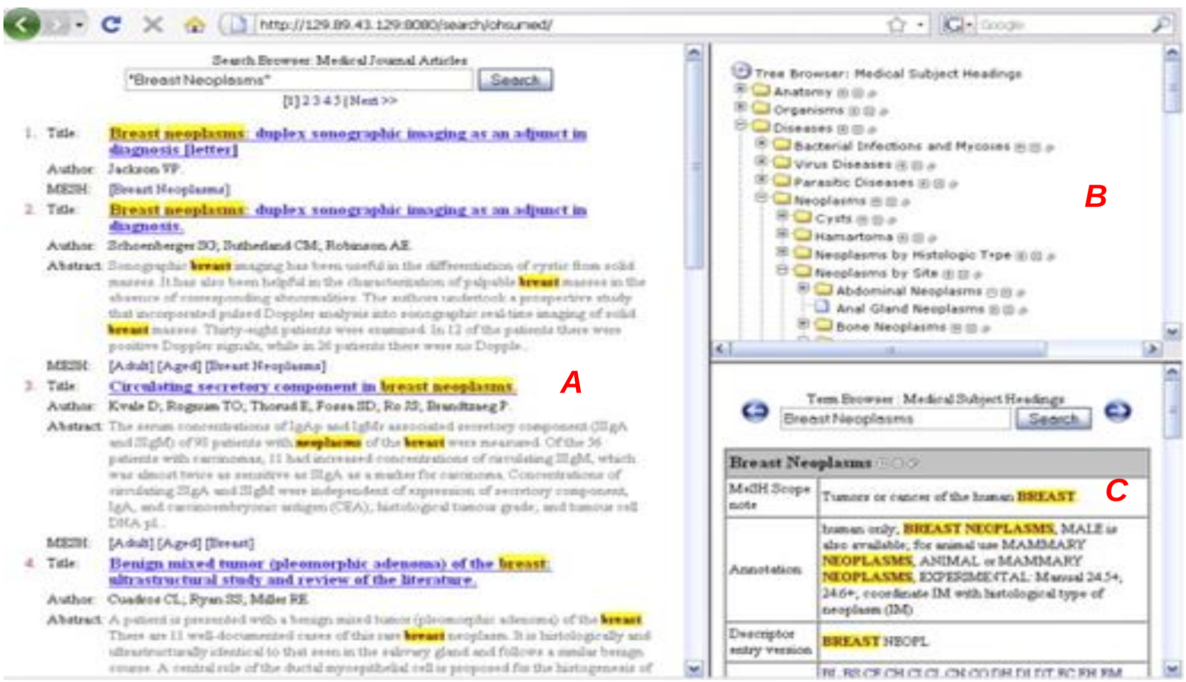

Figure 1. MeshMed search interface (A is the search browser, B is the tree browser, and C is the term browser).

3.2 Coding

The study collected a total number of 270 search sessions, with 135 from SimpleMed and 135 from MeshMed. The total number of sessions with at least one reformulation is 199, with 112 from SimpleMed and 87 from MeshMed. The query reformulation data was then independently coded by two of the authors using the coding scheme in Table 1. The inter-coder reliability turned out to be 0.851 according to Cohen's kappa, which suggests a high level of reliability.

Facets	Sub-facets	Example
Content changed	Specification(*SPE*)	lymphoma→ lymphoma definition
	Generalization(*GEN*)	Diabetic gastroparesis → gastroparesis

	Parallel movement(*PAR*)	lymphoma definition→ lymphoma small bowel
Content unchanged	Synonym(*SYN*)	Menopausal→ menopause
Format	Format(*FOR*)	definition:menopausal→ definition menopausal
Error	Error(*ERR*)	COPT→COPD

Table 1. Coding scheme

Reformulation type	Frequency	Percentage
ERR	34	5.57%
FOR	80	13.11%
GEN	106	17.38%
PAR	194	31.80%
SPE	193	31.64%
SYN	3	0.50%

Table 2. Frequency of reformulation types

3.3 Multi-level Models for Query Reformulation

In this study, a simple multi-level modeling with one level 2 variable (search session) and two level 1 variables (query reformulation and search performance) is employed to demonstrate the use of multi-level modeling for query reformulation data. As is shown in Figure 2, the model hypothesizes that query reformulations affect search performance with search session as the level 2 variable (query reformulations are nested in search sessions). As for the search performance, p@10 (precision value at the top 10th retrieved documents) is used. The model used the search session as the level 2 indicator. However, due to the limited space, the model does not include any level 2 predictors in this poster. It should be noted that level 2 predictors, such as search system types or search topic familiarity, can be easily incorporated in multi-level modeling to improve the inference.

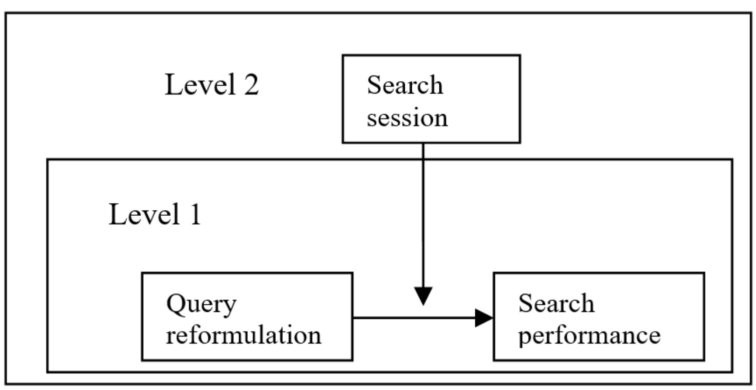

Figure 2. Diagram of a multi-level model for query reformulation data.

There are three approaches that multi-level modeling models the random effects: random intercepts (differences in the overall level of level 2 units), random slopes (differences in the effects of predictors across level 2 units), or both random intercepts and slopes. In this study, we demonstrate the random intercepts and both random intercepts and slopes. In random intercepts model, only the coefficient that varies across groups is the intercept. The model can be written as:

Level 2: $$\alpha_j = \mu_\alpha + \eta_j$$

$$\text{Level 1:} \quad y_i = \alpha_{j[i]} + \beta x_i + \varepsilon_i$$

where y_i is the p@10 (precision at top 10th retrieved documents) of the *i*th observation, x_i is *i*th query reformulation, β is the regression coefficient, and α_j is the group level intercept that can vary across different groups (*j* refers to the *j*th session).
The random intercepts and slopes model can have varying intercepts and slopes, which can be written as:

$$\text{Level 2:} \quad \alpha_j = \mu_\alpha + \eta_j^\alpha \qquad \beta_j = \mu_\beta + \eta_j^\beta$$
$$\text{Level 1:} \quad y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \varepsilon_i$$

where a_j is the varying intercept, β_j is the varying slope, other variables are the same as those in random intercepts model.

4 Initial Results

Table 3 lists the average search performance after different types of query reformulations as is measured by p@10. As Table 3 shows, *SPE* showed the highest p@10 among all, followed by *ERR*, *SYN*, *PAR*, *FOR*, and *GEN*. Then, the two types of multi-level models were applied. The *SYN* reformulations were excluded from the multi-level modeling since there were only three observations. *SPE* was selected as the reference level as it had the highest p@10.

With the random intercepts model, the p@10 after *FOR* was significantly lower than that after *SPE* (t=-3.879; $d.f.$=600; p<0.05), the p@10 after *GEN* was significantly lower than that after *SPE* (t=-5.475; $d.f.$=600; p<0.05), the p@10 after *PAR* was significantly lower than that after *SPE* (t=-3.829; $d.f.$=600; p<0.05). There was no significant difference between the p@10 after *ERR* and p@10 after *SPE* (t=-0.399; $d.f.$=600; p>0.05). The data was then fitted with the random intercepts and slopes model where both the intercept and slope can vary across search sessions. According to the statistical test results from random intercepts and slopes model, the p@10 after *FOR* was significantly lower than that after *SPE* (t=-3.554; $d.f.$=586; p<0.05), the p@10 after *GEN* was significantly lower than that after *SPE* (t=-5.619; $d.f.$=586; p<0.05), the p@10 after *PAR* was significantly lower than that after *SPE* (t=-2.418; $d.f.$=586; p<0.05), while the p@10 after *ERR* was not significantly different from that after *SPE* (t=-0.100; $d.f.$=586; p>0.05).

Reformulation type	P@10
ERR	0.109
FOR	0.044
GEN	0.032
PAR	0.095
SPE	0.144
SYN	0.100

Table 3. Search performance after different reformulation types

5 Conclusion

This study introduces multi-level modeling to query reformulation data analysis. Query reformulations are generally nested within search sessions, which limits the adoption of traditional statistical analysis such as regression or ANOVA. Multi-level modeling offers a compelling method to analyze such nested data in query reformulation studies. With multi-level modeling, predictors at different levels (session level, move level) can be integrated into a single research model. It also avoids the violation of independent assumption. Future study will further explore more complex multi-level models for query reformulation.

6 References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society*, 149(1), 1-43.

Blalock, H. M. (1984). Contextual-effects models: Theoretical and methodological issues. *Annual Review of Sociology*, 10, 353-372.

Blau, P. M. (1960). Structural effects. *American sociological review*, 25(2), 178-193.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, 101(1), 147.

Davis, J. A., Spaeth, J. L., & Huson, C. (1961). A technique for analyzing the effects of group composition. *American Sociological Review*, 26(2), 215-225.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, 3(1), 1-21.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of management*, 23(6), 723-744.

Hu, R., Lu, K., & Joo, S. (2013). Effects of topic familiarity and search skills on query reformulation behavior. In *Proceedings of the Association for Information Science and Technology 2013 (ASIST 2013)*. Montreal, Canada.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management*, 36(2), 207-227.

Jansen, B. J., Spink, A., & Narayan, B. (2007). Query modifications patterns during Web searching. In *Proceedings of the International Conference on Information Technology 2007 (ITNG'07)*, pp. 439-444.

Jansen, B. J., Booth, D. L., & Spink, A. (2009). Patterns of query reformulation during Web searching. *Journal of the American Society for Information Science and Technology*, 60(7), 1358-1371.

Joo, S., & Lee, J. (2011). Assessing effectiveness of query reformulations: Analysis of user-generated information retrieval diaries. In *Proceedings of the Association for Information Science and Technology 2011 (ASIS&T 2011)*, New Orleans, LA, USA.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, 99(3), 422.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4), 963-974.

Liu, C., Gwizdka, J., & Belkin, N. J. (2010). Analysis of query reformulation types on different search tasks. In *Proceedings of the iConference 2010*, Urbana Champaign, IL.

Metzler, D., & Croft, W. B. (2004). Combining the language model and inference network approaches to retrieval. *Information Processing and Management*, 40(5), 735-750.

Mu, X., Lu, K., & Ryu, H. (2014). Explicitly integrating MeSH thesaurus help into health information retrieval systems: An empirical user study. *Information Processing and Management*, 50(1), 24-40.

Rieh, S. Y., & Xie, H. (2006). Analysis of multiple query reformulations on the Web: The interactive information retrieval context. *Information Processing and Management*, 42(3), 751-768.

Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociology Review*, 15(3), 351-357.

Spink, A., Wolfram, D., Jansen, B.J., & Saracevic, T. (2001). Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3), 226-234.