Interactions Between Decision Goals Applied to the Calculation of Context Dependent Re-rankings of Results of Internet Search Engines

Rudolf Felix

F/L/S Fuzzy Logik Systeme GmbH, Joseph-von-Fraunhofer-Str. 20, 44227
Dortmund, Germany
felix@fuzzy.de

Abstract. Search keywords can be considered as decision criteria (goals), whereas the documents contained in the original result list of a search engine can be considered as decision alternatives in the sense of adequate search results. The paper shows how a decision making model based on interactions between decision goals can be used for re-ranking of the search results obtained by classical search engines. The decision making model was previously applied to real world problems in production optimization and business process management. In contrast to other approaches, the interactive structure of decision goals for each decision situation is calculated explicitly based on fuzzy types of interaction. In this paper, after a brief description of the model an application to the calculation of context depending re-rankings of search results of internet search engines is presented. It is shown that keywords of search queries can be understood as decision goals and that the decision making model can be successfully applied for context dependent re-ranking of the search results. Using the model, popularity based rankings can be re-ranked making use of context dependent information derived from the query keywords.

Keywords: Decision making, interaction of decision goals, context depending rankings, internet search, re-ranking of search results.

1 Introduction

Humans efficiently decide whether or not a listed document is interesting from the point of view of the search query they previously entered to the search engine. They rather perform a cross check of the documents instead of reading them in detail and rank them with respect to the initial intention of the search query expressed by the keywords of the query. The process of re-ranking the search results can be considered as a decision making process with the keywords used in the query being the decision goals. The decision alternatives are the selected documents themselves which are to be re-ranked with respect to the keywords. It is shown how the concept of interactions between decision goals (here: keywords) can be applied for ranking the documents (here: decision alternatives). In the subsequent sections first basic definitions used to formalize the concept of interactions between decision goals are presented. It is also discussed in which sense interactions between the goals are used to model the decision making. Then it is shown how the model is applied to the

O. Castillo et al. (Eds.): Theor. Adv. and Appl. of Fuzzy Logic, ASC 42, pp. 81–88, 2007.
springerlink.com

problem of context dependent ranking of documents being search results of a classical internet search engine. Subsequently the results obtained from a system prototype are discussed and a summarizing conclusions are made.

2 Basic Definitions

Before we define interactions between goals as fuzzy relations, we introduce the notion of the positive impact set and the negative impact set of a goal. A more detailed discussion can be found in [6] and [7].

Def. 1)
Let A be a non-empty and finite set of potential alternatives, G a non-empty and finite set of goals, $A \cap G = \emptyset, a \in A, g \in G, \delta \in (0,1]$. For each goal g we define the two fuzzy sets S_g and D_g each from A into [0, 1] by:

1. Positive impact function of the goal g: $S_g(a) := \delta$, if a affects g positively with degree δ then $S_g(a) = \delta$, $S_g(a) := 0$ *else.*
2. Negative impact function of the goal g: $D_g(a) := \delta$, if a affects g negatively with degree δ then $D_g(a) = \delta$, $D_g(a) := 0$ *else.*

Def. 2)
Let S_g and D_g be defined as in Def. 1). S_g is called the *positive impact set of* g and D_g the *negative impact set of* g.
The set S_g contains alternatives with a positive impact on the goal g and δ is the degree of the positive impact. The set D_g contains alternatives with a negative impact on the goal g and δ is the degree of the negative impact.

Def. 3)
Let A be a finite non-empty set of alternatives. Let $\mathcal{P}(A)$ be the set of all fuzzy subsets of A. Let $X, Y \in \mathcal{P}(A)$, x and y the membership functions of X and Y respectively.

The *fuzzy inclusion* $\mathrm{I}: \mathcal{P}(A) \times \mathcal{P}(A) \rightarrow [0,1]$ is defined as follows:
$\mathrm{I}(X,Y) := \Sigma_{a \in A} (\min(x(a), y(a))) / \Sigma_{a \in A}$, for $X \neq \emptyset$ and $\mathrm{I}(X,Y) := 1$ for $X = \emptyset$, with $x(a) \in X$ and $y(a) \in Y$.
The *fuzzy non-inclusion* $\mathrm{N}: \mathcal{P}(A) \times \mathcal{P}(A) \rightarrow [0,1]$ is defined as:

$$\mathrm{N}(X,Y) := 1 - \mathrm{I}(X,Y)$$

The inclusions and non-inclusions indicate the existence of interaction between two goals. The higher the degree of inclusion between the positive impact sets of two goals, the more cooperative the interaction between them. The higher the degree of inclusion between the positive impact set of one goal and the negative impact set of the second, the more competitive the interaction. The non-inclusions are evaluated in a similar way. The higher the degree of non-inclusion between the positive impact sets of two goals, the less cooperative the interaction between them. The higher the degree of non-inclusion between the positive impact set of one goal and the negative impact set of the second, the less competitive the relationship.

The pair (S_g, D_g) represents the whole known impact of alternatives on the goal g. Then S_g is the fuzzy set of alternatives which satisfy the goal g. D_g is the fuzzy set of alternatives which are rather not recommendable from the point of view of satisfying the goal g.

3 Interactions Between Goals

Based on the inclusion and non-inclusion defined above, 8 basic fuzzy types of interaction between goals are defined. The different types of interaction describe the spectrum from a high confluence between goals (analogy) to a strict competition (trade-off).

Def. 4)
Let S_{g_1}, D_{g_1}, S_{g_2} and D_{g_2} be fuzzy sets given by the corresponding membership functions as defined in Def. 2). For simplicity we write S_1 instead of S_{g_1} etc.. Let $g_1, g_2 \in G$ where G is a set of goals.

The fuzzy types of interaction between two goals are defined as relations which are fuzzy subsets of $G \times G$ as follows:

1. g_1 is independent of g_2: <=> $\min(\mathrm{N}(S_1, S_2), \mathrm{N}(S_1, D_2), \mathrm{N}(S_2, D_1), \mathrm{N}(D_1, D_2))$
2. g_1 assists g_2: <=> $\min(\mathrm{I}(S_1 S_2), \mathrm{N}(S_1, D_2))$
3. g_1 cooperates with g_2: <=> $\min(\mathrm{I}(S_1, S_2), \mathrm{N}(S_1, D_2), \mathrm{N}(S_2, D_1))$
4. g_1 is analogous to g_2: <=> $\min(\mathrm{I}(S_1, S_2), \mathrm{N}(S_1, D_2), \mathrm{N}(S_2, D_1), \mathrm{I}(D_1, D_2))$
5. g_1 hinders g_2: <=> $\min(\mathrm{N}(S_1, S_2), \mathrm{I}(S_1, D_2))$
6. g_1 competes with g_2: <=> $\min(\mathrm{N}(S_1, S_2), \mathrm{I}(S_1, D_2), \mathrm{I}(S_2, D_1))$
7. g_1 is in trade-off to g_2: <=> $\min(\mathrm{N}(S_1, S_2), \mathrm{I}(S_1, D_2), \mathrm{I}(S_2, D_1), \mathrm{N}(D_1, D_2))$
8. g_1 is unspecified dependent from g_2: <=>
   $\min(\mathrm{I}(S_1, S_2), \mathrm{I}(S_1, D_2), \mathrm{I}(S_2, D_1), \mathrm{I}(D_1, D_2))$

The interactions between goals are crucial for an adequate orientation during the decision making process because they reflect the way the goals depend on each other and describe the pros and cons of the decision alternatives with respect to the goals. For example, for cooperative goals a conjunctive aggregation is appropriate. If the goals are rather competitive, then an aggregation based on an exclusive disjunction is appropriate. The interactions between the decision goals have been applied in a problem independent decision making model that has already been applied in many relevant real world solutions [9], [11]. For a more detailed discussion see for instance [7].

4 Approaches Related to the Decision Making Model Based on Interactions Between Goals

Since fuzzy set theory has been suggested as a suitable conceptual framework of decision making [2] several different approaches have been developed [3], [4], [5], [13], [14], [15], [17], [19],[20], [21]. The decision making approach based on interaction between goals significantly differs from these approaches. For a more detailed discussion see for instance [6], [8], [10], [12]. It also significantly differs from fuzzy rule based approaches applied for instance to the selection of web service compositions [1] as these approaches do not explicitly use negative selection information.

5 Application to Context Dependent Ranking of Internet Search Results

In the subsequent sections the application of the decision making model to the context dependent ranking of documents that are search results of a classical internet search engine is described. First the ranking problem is defined. Then it is indicated how the positive and negative impacts sets are used in order to translate the context of a search query to the decision making model. Subsequently examples of search queries are shown and the results of the rankings of the documents are discussed. The discussion of the resulting rankings is made by comparing the results with the rankings obtained by a well-known internet search engine.

5.1 Description of the Ranking Problem

Classical internet search engines generate rankings which are not based on content dependent information. The search keywords which represent the context of the query are used as index information. The rankings of the documents are then rather popularity based and concentrate on information like the number of input and output links. On the other hand a particular user evaluates the search result from his own point of view, that means from the point of view of the context of his own query. In consequence, the ranking generated by a classical search engine may not fit the intention of the user's search query. In such cases a document may be part of the set of documents found but may be ranked with a lower value and placed towards the end of the ranking generated by the search engine. The aim of the approach presented below is to re-rank the result of a classical search engine in a way that better fits the context of the query.

5.2 Positive and Negative Context Information

The search keywords can be considered as decision criteria (as goals in the sense of Def. 1) whereas the documents contained in the original result list of the search engine can be considered as decision alternatives. Using the decision making model presented in the previous sections the ranking problem can be solved in the following way.

For each search query the keywords used are considered as the set G of decision goals. The result list to this query generated by the conventional search engine is considered as the set of decision alternatives A. For every goal (keyword) the positive impact function's value for a particular document a $\in A$ is defined by counting the number of occurrences of the keyword (goal) normalized by the total number of words contained in the document (alternative). The more occurrences of the search keyword in the document the higher the value of the positive impact function of the keyword (goal) for the document (alternative) and the lower the value of the negative impact function of the keyword (goal) for the document (alternative). In an analog way we model: The less occurrences of the positive keyword (goal) in the document the lower the value of the positive impact function and the higher the value of the negative impact function.

The user also has the possibility to define the so-called negative keywords which are key-words that should not appear in the document. Using the concept of impact functions for each negative keyword again the documents are evaluated in an analog way as the positive keywords but with opposite increasing and decreasing values.

In the sense of the search query defined by the user the positive keywords describe what the user is tentatively looking for, the so-called positive context of the search query. The negative keywords describe what the user tentatively is not looking for, the so-called negative context of the search query.

Every search query is evaluated in the way that every document which is contained in the ranking generated by the conventional search engine is considered as a decision alternative $a \in A$. For every keyword the values of both the positive and the negative impact functions are calculated as described above. If for every keyword both the positive and the negative impact functions are calculated and the impact sets are passed to the decision model based on interactions between decision goals (here search keywords). The decision model calculates a new ranking of the documents which better correspond then to the context of the search query defined by the user.

5.3 Examples of Search Queries and Resulting Rankings

Let us explain some experimental results of the search concept obtained by using some common sense examples. The examples presented subsequently are based on the internet search results of Google and Google's index information. As already explained each search query consists of two sets of keywords. The first one is the positive keyword set (PKWS). The PKWS consists of keywords which describe what the user is searching for, that means the positive context of the search query. The second set of keywords is the negative keyword set (NKWS). The NKWS consists of keywords that describe the context the user is rather not interested in, that means the negative context of the search query. A search query is always a pair (PKWS: ={positive keywords},NKWS:={negative keywords}). The search is organized as follows: First the search engine Google is used to generate a list of documents based on a search query with the PKWS as keyword list and the result is the initial search result list (ISRL). The order of the links in the ISRL is the initial ranking (IR) of the links or documents. The ISRL is used as input for the re-ranking procedure based on the interactions between the keywords (goals) as presented in the previous sections. The result of the re-ranking is the final ranking (FIR). For simplicity, in the subsequent examples the ISRL consisted of at most 50 links. Therefore both IR and FIR have at maximum 50 ranking positions.

Let us consider a first search query Q1=(PKWS:={Klose},NKWS=:{}). The positive keyword set consists of the keyword "Klose", the negative keyword set is empty. The IR of the query consists of 50 positions of which 24 links refer to documents about the German soccer player Miroslav Klose. If we modify the search query by adding the negative keyword "Fussball" which is the German word for "soccer" we obtain the search query Q2=(PKWS:={Klose}, NKWS:={Fussball}). The first 10 ranking positions of FIR of Q2 compared to the first 10 ranking positions FIR of Q1 and the links are given in Fig. 1. It can be seen that links with relation to the soccer player Miroslav Klose dominate the FIR of Q1 and that the dominance in the result of Q2 is reduced to 0 because in the FIR of Q2 the positions 1 to 10 do not contain any link related to the soccer player Miroslav Klose.

FIR of Q1				FIR of Q2		
Links	Resulting Ranking	Google Ranking		Links	Resulting Ranking	Google Ranking
Miroslave Klose – die offfizielle Webseite *http://www.miroslavklose.de*	1	1		Friedrich Klose – Wikipedia http://de.wikipedia.org/wiki/F...	1	3
Miroslav Klose – Wikipedia http://de.wikipedia.org/wiki/M...	2	2		Die Klose Kollektion GmbH http://www.klosekollektion.de/	2	4
Friedrich Klose – Wikipedia http://de.wikipedia.org/wiki/F...	3	3		Staudengaertner-Klose, Paeonien, Staudenversand, Garten *http://www.staudengaertner-klo...*	3	6
Die Klose Kollektion GmbH http://www.klosekollektion.de/	4	4		Galerie Klose http://www.galerie-klose.de/	4	11
Staudengaertner-Klose, Paeonien, Staudenversand, Garten *http://www.staudengaertner-klo...*	5	6		BBQ Pits by Klose – Houston, TX *http://www.bbqpits.com/*	5	12
2006 FIFA World Cup Germany – Player Profile Page – KLOSE Miroslav – Germany *http://fifaworldcup.yahoo.com/...*	6	7		Rainer Klose Werkzeuge für's Handwerk, Münster *http://www.klose-ms.de/*	6	13
2006 FIFA World Cup Germany – Player Profile Page – KLOSE Miroslav – Germany *http://fifaworldcup.yahoo.com/...*	7	8		Fehlermeldung http://www.bundestag.de/mdb15/...	7	14
ZEIT online - Fussball --- WM 2006: Es traf Klose *http://www.zeit.de/2006/24/Fus...*	8	9		Klose, Hans-Ulrich http://www.bundestag.de/mdb/bi...	8	15
ZEIT online – Fussball --- Länderspiel: Überragender Klose *http://www.zeit.de/online/2006...*	9	10		BW – Hilchenbach.de http://www.klose-antriebstechn...	9	18
Galerie Klose http://www.galerie-klose.de/	10	11		Familie Klose aus Heisingen *http://www.klose-family.de/*	10	19

Fig. 1. Comparison of the results of the search queries Q1 and Q2

Let us now consider an another search query Q3=(PKWS:={Ronaldo}, NKWS=:{}). Again for simplicity the IR was reduced to at most 50 links. Without going into details it was observed that the ranking of Q3 contained in total 34 links which refer to the Brazilian soccer player Ronaldo and 7 links referring to the Portuguese soccer player Christiano Ronaldo. Obviously there is an ambiguity because of the nickname of the Brazilian player and the last name of the Portuguese player. However, among the first 20 links in the ranking there were 17 links to Ronaldo and only 2 links to Christiano Ronaldo. This overrepresentation of the Brazilian Ronaldo was due to the fact that he is more famous and therefore there were more links to internet sites referring to him. Let us compare the results of Q3 with the results obtained for Q4=(PKWS:={Ronaldo}, NKWS=:{Brasil}). We used the radical "Brasil" of the German word Brasilien for Brazil and defined it as element of the NKWS of the search query Q4. With this query we tried to lower the ranking positions of the Brazilian Ronaldo. And in fact we observed that this was the case: Among the first twenty positions of the ranking for Q4 there were 15 links to the Brazilian Ronaldo and 5 links to the Portuguese Christiano Ronaldo. Please, note that we did not obtain the same result by using the Google search with the excluding keyword “Brasil”. The Google search result with “Ronaldo –Brasil” provided a ranking which still contained 17 links to the Brazilian among the first 20 positions in the ranking and 2 to the Portuguese. We considered the first 10 ranking positions of both Q3 and Q4 observed that in the ranking of Q3 we had 8 links to the Brazilian Ronaldo and only 2 to Christiano Ronaldo. Compared to this Q4 provided a ranking with 5 links to the Brazilian Ronaldo and also 5 to the Portuguese Cristiano Ronaldo.

6 Usability of Positive and Negative Keywords

When using search engines, users normally make their queries with two or three words. One might assume that it will be difficult motivating the user to provide negative keywords. However, if the search queries are not trivial, the number of keywords may increase. It can be expected that if the classical popularity ranking doesn’t match the user’s interest (like in the Ronaldo-Example) the documents on the top of the classical ranking will be negative examples of search results that do not fit what the user was looking for. If so, it will be easy for the user to identify negative keywords for instance by observing the titles and /or some keywords of these documents. Already after a few clicks there will be enough information available for writing down two or three negative keywords and starting a new (extended) search query again. This kind of user behavior may also be part of the user model and the user’s interaction may be evaluated in the sense of an automated acquisition of both positive and negative query information [16].

Please note that using negative keywords as considered in the paper is different to the notion of negative selectivity as discussed in [18], where the selectivity is understood as the cardinality of the result of the query and is called negative if the cardinality is equal to 0.

7 Conclusions

The approach presented in this paper shows that the decision making model based on interactions between decision goals may be successfully used to re-rank search results of internet search engines with classical popularity based rankings. The process of cross-checking documents can be modeled as a decision making process within which the keywords are the decision goals and the decision alternatives are the documents to be re-ranked. Using both positive and negative keywords helps to better describe the search context. The interaction between the keywords considered as interaction between goals during the execution of the search queries is in opinion of the author a promising idea of extending popularity based search by keyword oriented context depending re-ranking of search results.

References

1. Agarwal, S. and Lamparter, S., "User Preference Based Automated Selection of Web Service Compositions", IBM Research Report, Proceedings of the International Workshop on Dynamic Web Processes (DWP 2005), Amsterdam, Netherlands, 2005
2. Bellman, R. E. and Zadeh, L. A., "Decision Making in a Fuzzy Environment", Management Sciences, 17, 141-164, 1970.
3. Biswal, M. P., "Fuzzy programming technique to solve multi-objective geometric programming problems", Fuzzy Sets and Systems, Vol.51, 67-71, 1992.
4. Dubois, D. and Prade, H., "Criteria Aggregation and Ranking of Alternatives in the Framework of Fuzzy Set Theory", Studies in Management Sciences, Volume 20, 209-240, 1984.
5. Dubois, D. and Prade, H., "Fuzzy Sets and Possibility Theory: Some Applications to Inference and Decision Processes", Fuzzy Logic - Theorie und Praxis, ed B Reusch: Springer-Verlag, 66-83, 1992.
6. Felix, R., "Entscheidungen bei qualitativen Zielen", PhD Dissertation, University of Dortmund, Department of Computer Sciences, 1991.
7. Felix, R., "Relationships between goals in multiple attribute decision making", Fuzzy Sets and Systems, Vol.67, 47-52, 1994.
8. Felix, R., "Fuzzy decision making based on relationships between goals compared with the analytic hierarchy process", Proceedings of the Sixth International Fuzzy Systems Association World Congress, Vol.II, Sao Paulo, Brazil, 253-256, 1995.
9. Felix, R., Kühlen, J. and Albersmann, R., "The Optimization of Capacity Utilization with a Fuzzy Decision Support Model", Proceedings of the fifth IEEE International Conference on Fuzzy Systems, New Orleans, USA, 997-1001, 1996.
10. Felix, R., "On fuzzy relationships between goals in decision analysis", Proceedings of the 3rd International Workshop on Preferences and Decisions, Trento, Italy, 37-41, 2000.
11. Felix, R., "Fuzzy decision making with interacting goals applied to cross-selling decisions in the field of private customer banking", Proceedings of the 10th IEEE International Conference on Fuzzy Systems, Melbourne, Australia, 2001.
12. Felix, R., "Efficient Decision Making with Interactions Between Goals", Proceedings of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, Honolulu, Hawaii, USA, 2007.
13. Freeling, A. N. S., "Possibilities Versus Fuzzy Probabilities - Two Alternative Decision Aids", Studies in Management Sciences, Volume 20, Fuzzy Sets and Decision Analysis, 67-82, 1984.
14. Grabisch, M., “Fuzzy integral in multicriteria decision making”, Fuzzy Sets and Systems, 69, 279-298, 1995.
15. Modave, F., Dubois, D., Grabisch, M. and Prade, H., “A Choquet integral presentation in multicriteria decision making”, AAAI Fall Symposium, Boston, Ma, November 1997.
16. Parcus de Koch, N., "Software Engineering for Adaptive Hypermedia Systems", PhD Dissertation, Ludwig-Maximilians-Universität, Munich, Department of Mathematics and Computer Sciences, 2000.
17. Saaty, T. L., "The Analytic Hierarchy Process", Mc Graw-Hill, 1980.
18. Spiegel, J., Polyzotis, N., "Graph-Based Synopses for Relational Selectivity Estimation", SIGMOD 2006, June 27-29, Chicago, Ill., 2006.
19. Torra, V., "Weighted OWA operators for synthesis of information“, Proceedings of the fifth IEEE International Conference on Fuzzy Systems, New Orleans, USA, 966-971, 1996.
20. Yager, R.R., “Families of OWA operators”, Fuzzy Sets and Systems, 59, 125-148, 1993.
21. Zimmermann, H. J., "Fuzzy Set Theory and its Applications", Kluver/Nijhoff, 1991.