SmartPick: Augmenting Advertisement Service with Domain Knowledge and Qualitative Reasoning

Aarno LEHTOLA, Teppo VEIJONEN, Jussi AHOLA, Paula SILVONEN, Mika TIMONEN, Markus TALLGREN
VTT Information Technology, PB 1201, Espoo, 02044 VTT, Finland
Tel: +358 20 722111*, Fax: +358 20 722 6027, Email: firstname.surname@vtt.fi

Abstract: This paper describes how advertisement services can be augmented to better serve their customers by accepting also qualitative selection criteria. Until now, the services have been confined by the actual contents of the advertisements and the connection to related domain knowledge has been very elementary or totally missing. Such knowledge can however provide significant added value by enabling selection by qualitative criteria. We have developed a prototype system, SmartPick, to demonstrate the benefits of qualitative reasoning in an advertisement service for used cars. Similar benefits can be expected for other types of products as well. We discuss also possible future developments.

1. Introduction

According to many surveys a major cause for the poor usability of B2C eCommerce services is that the customers cannot find the requested products. "No product found" is an insufficient answer, if the selection includes comparable goods or if the user just happens to use words that are missing from the product database. Services should find the offerings that best match the user requirements also when an exact match is not available. In order to do so the systems would need qualitative reasoning capabilities and access to knowledge resources related to the domain.

With certain goods the wideness of supply makes the choosing very difficult for consumers. For instance, the selling of used cars in Finland is estimated to rise to 550,000 units in 2005. The three top advertising portals for used cars, at any given moment, contain altogether advertisements of almost 70,000 cars. Individual cars can be queried by stating crisp conditions for car makes and models, geographic regions of the seller, prices, year models, driven kilometres or other explicit contents of the advertisement. Criteria that go beyond the advertisement contents are not supported.

Similar situations prevail in other categories of goods mediated by eCommerce services. The use is easy, if one knows in exact terms what one is looking for and if the selection includes the desired item with the matching information explicit in the advertisement. However, these services are not suitable for situations, when more general qualitative criteria are used. One will not find answer for a query of “a table with a fade proof and durable surface” or of “a pastel coloured and shrink-resistant blouse”. Moreover, the current services do not assist the user when information requests do not result in answers.

This paper describes a prototype system, SmartPick, which solves the above-mentioned problems by using qualitative reasoning techniques and value-adding knowledge resources, which are related to the artefacts on sale. The system has been made for sales of used cars, but there is no reason why it cannot be adapted to other domains of advertising such as furniture, clothing, apartments etc. First we discuss about the research background and
objectives of our work. After that we describe the methodology and techniques used. Next we evaluate the current solutions. Finally we outline what would be the business benefits.

2. Objectives behind SmartPick

Ontologies have been widely recognised as a central solution for sharing conceptions of goods and services among parties in eCommerce [1]. In Mkbeem eCommerce mediation system we applied crisp ontologies to engineer multilinguality [7,8,9]. Domain ontologies included product models that specified components and properties of the mediated products. All ontologies handled crisp concepts. Limited qualitative reasoning about materials and colours was included using related generic ontologies and crisp inference rules. Main objective of SmartPick is to demonstrate more extensive solutions for qualitative product selection and innovative ways of combining knowledge from multiple sources of qualitative data. Used car selecting was chosen for the test case because of its difficulty, its semantic richness combining both crisp and imprecise information, and its evident business potential.

![Figure 1: Product finding using qualitative criteria and multiple knowledge resources.](image)

Figure 1 illustrates the multiple objective optimisation problem a car seeker is solving when trying to find the most favourable match from the available used car offerings in an online advertisement system. The listed requirements are qualitative in nature and numerous cars could potentially match with the criteria. Checking systematically with respect to a large selection of cars would be extremely tedious task without automation. External knowledge resources related to solving these constraints include:

- Car model data
- Safety assessment information sources (EuroNCAP, US NCAP, IIHS etc.)
- Fault statistics sources (ADAC, car inspection offices, insurance companies etc.)
- Used car reference price statistics sources (several car magazines and vendors collect information about realised purchase prices)
- Catalogue of company contact information
- Geographical information service

All of the listed information is nowadays accessible through the Internet, although Web Services or corresponding interfaces for easy automated access are for the most part non-existent. Furthermore, the solving would need ontologies, like a product model for cars and...
a colour model. Our objective has been to develop a system to demonstrate how qualitative
criteria and knowledge from multiple additional resources can be used to greatly facilitate
the product selection from a large number of advertisements. In other words, we try to solve
the product matching problem described in Figure 1 as well as possible. The main customer
benefit is that the selection of relevant items gets greatly enlarged.

3. Methodology: Quality Dimensions and Fuzzy Logic

People do their qualitative product comparisons using conceptual models of those products.
Usually product related information contains both numeric and symbolic data. There may
be data sets that as a whole form meaningful concepts, e.g., through some sort of aggregate
functions. Our system needs to solve the mapping from product data to such natural
concepts that participate in the related human decision making processes. Our approach is
influenced by the theory of Conceptual Spaces by Gärdenfors [3]. The theory sees (natural)
concepts as (convex) regions in conceptual spaces defined by quality dimensions. The
mapping from measurable quantities to such concepts retains the possibility to use classical
AI symbol manipulation approaches at the higher level. In SmartPick quality dimensions
may be based on different metrics depending on their characteristics.

To deal with imprecise and imperfect information we decided to use neurofuzzy
methodologies. Concepts in a product domain may have imprecise definitions. Moreover,
user queries typically contain both crisp criteria and imprecise criteria. For example, in
Figure 1 “air conditioning” is a crisp condition while “good city drivability” is an imprecise
one. Fuzzy data models have been found suitable for comparison of complex objects and
finding relevant partial matches, e.g. [10]. For symbolic reasoning we decided to use fuzzy
logic [13,14]. Values from different dimensions are translated into membership degrees in
fuzzy sets. User input is formalised as conditional expressions in fuzzy logic. They are used
to infer whether an item is acceptable or not.

4. Technical Solutions in the Car Sales Use Case

In the domain of car sales, one can find quality dimensions, such as fuel consumption,
colour, safety, drivability, robustness, age etc. Customer requirements refer to these
dimensions. They are translated into fuzzy logic rules of the form:

\[
\text{if } \begin{array}{l}
\text{fuel consumption is economical and age mileage is newish and}
\text{city drivability is very good and colour is greenish and}
(\text{highway drivability is average or good or very good})
\end{array}
\text{then suitability is very good for MrSmith.}
\]

The technical solution, which is used to handle inexactness of information, varies
depending on the quality dimension. In the following subsections we give examples
concerning three quality dimensions.

4.1 Quality Dimension for Fuel Consumption – Fuzzy Sets

Fuel consumption has been defined using fuzzy sets. Definitions for qualifiers are
economical, average, and hungry (Figure 2). The degree of membership of a car in each set
can be computed straightforward after querying its consumption figures from the car model
information database.
4.2 Quality Dimension for Colours – Colour Ontology Using CIELab Model

The system must know more about colours than just their names that are present in the advertisements. It must associate verbally given colour names with colour hierarchy and distance metrics. We have developed a colour ontology for this purpose. It combines a hierarchy of discrete colours with location in colour space. For the colour space model we chose CIELab [4]. CIELab (also called CIE L*a*b*) is common in applications where closeness of colours must be quantified, such as colorimetry, gemstone evaluation, or dye matching. Colour appearance has a categorical nature [5], e.g. in English there are eleven categories (i.e. black, white, red, green, yellow, blue, brown, purple, pink, orange and grey). Our colour ontology model contains these eleven colours as basic colour categories. Each category is defined by its names and its region in the CIELab colour space. Specific colours are defined by their names and for the moment by their crisp colour space co-ordinates. Further on, they are included in a colour hierarchy.

Distance can be calculated between any two colours to find how similar they are, or when fuzzification is concerned, what is the membership of a colour in another. Two heuristically set thresholds are in use as limits when closeness of colours is scaled between the interval $[0,1]$. These are tolerance for accepting as fully equal and limit for judging as fully unequal.

Advertisements can contain descriptive colour names such as almond, emerald, orange, ruby, sand etc. After the customer has chosen preferred colours, the system can evaluate the membership of, e.g., an ochre car in the set of selected colours (red 0.85, yellow 0.66 etc.). When colours belong to several categories, the fuzzy membership functions have particular importance in computing similarities.

4.3 Quality Dimensions for Use Purposes – Self-Organising Maps

Suitability of a car into particular purpose of use can be addressed by the well-known strict categories like family car, city car, sports car etc. However, a family car can also be, to some extent, a city car or a sports car. Thus, a more flexible classification of cars is needed to better categorise different cars based on their use purposes. We decided to apply self-organising maps (SOM) for analysing what quantities in car type specific data best distinguish between different use purposes [6,11].

The first task of the analysis was determining a set of representative features to be used in the categorisation. The data that was used in the feature selection consisted of 42 facts of over 2500 car models. The features that best represent the desired car purposes of use were selected based on human expertise together with visual and statistical interpretation of the data. In particular, the strong visualisation capabilities of the SOM were utilised in finding significant patterns and dependencies in the data. The features that were chosen to be used in the further analysis were number of doors, maximal speed, acceleration, motor power, motor moment, axle spacing, front track width, total mass, tyre diameter, height and length.

For the flexible categorisation of the cars based on their use purposes, a SOM was trained with the data set resulting from the feature selection. This included first determining
the use purpose categories and then defining the membership of each car to different categories. In the first phase, a crisp classification using hierarchical clustering approach was used for category determination. The categories were then analysed and labelled accordingly to different use purposes. The membership of a car into a use purpose category was defined subsequently as a distance to the category centre. Both distance on the discrete map neighbourhood and Euclidean distance on the original feature space were considered. Other possible approaches that could be used in the categorisation are fuzzy or even mixture model clustering, which would enable more probability-like membership definition. Figure 3 illustrates the car purpose categorisation using the SOM. In the future this kind of visualisation could also be utilised in the user interface to help users in setting their criteria. Users could select areas from the map and get lists of corresponding models in response.

4.4 Semantic Considerations in Specifying Concepts

Some concepts involve more complex semantics. For instance, “newish” cannot be confined solely to the age, as a few years old car may already be well-worn, if it has been driven much, e.g., as a taxi. Moreover, small city cars tend to last shorter mileage than big cars. We estimate the worn condition of cars with a function of their age, mileage and size. However, chance for misjudgement remains as a car may be in far better condition than its reference group, if it has been well taken care of by a devoted owner.

5. The SmartPick System

The current SmartPick system addresses most of the objectives outlined in Section 2. When it is solving which cars are acceptable, it automatically complements the contents of car advertisements with additional knowledge about car model data (45 facts of over 2500 models collected by Tekniikan Maailma, a Finnish magazine for popularised technology), car safety (EuroNCAP scoring), fault statistics (ADAC statistics and Swedish car inspection statistics), and colours. Customer requirements can contain fuzziness. For instance, the customer requirement specification "quite fuel economical, relatively robust and greenish car with good city drivability, fair highway drivability, good safety, drawhook, central locking, power steering, ABS brakes" can be accepted. There is no need to mention any particular car make or model in the query. The query can be purely based on qualitative criteria. Depending on the nature of a quantity, a criterion can be crisp or imprecise. It is straightforward to check from car model data crisp technical conditions, such as the ABS brakes. The system uses fuzzy logic to evaluate imprecise conditions, like newish, greenish, fair highway drivability etc.
The user interface of SmartPick is shown in Figure 4. One goal in its development has been to find out what kind of controls and scales are the best for each criterion. The interface is divided into five panes: requirement specification pane (top left), query invocation pane (bottom left), pane for answer list (bottom middle), alert invocation pane (bottom right), and pane for showing individual car details (top right). The requirement specification pane starts from the top with model and feature specification. The user interface is kept consistent with respect to the selected features. For instance, when the user starts with selecting city car category, he gets immediately a set of cars from that class into the included models list and the feature prioritisation adjusters are set to corresponding positions. If he uses them to set further requirements, for instance to highway drivability, those ones that no longer fulfil the new adjustment will be removed from the model list.

The user can also select particular car model and check the box "include alike models". In that case the system uses the model as a prototype and finds models that are equal enough by their properties. The lower part of the requirement specification pane contains conditions for the advertised car instances.
Although the computing model uses qualitative values from best to worst, for many quantities only positive values are relevant to ask in the user interface. The user would never choose the negative values. For instance, nobody wants to have a car with bad safety. In these cases, the adjustments are made by setting priorities to positive values. In case of age and mileage the user may deliberately want to choose also the worst case, namely "repairable", if he is a good car mechanic or wants to buy a car in order to disassemble it for spare parts. Some accessories, like a drawhook, are obtainable afterwards. If a used requires them they are not used to prune out cars but merely to prioritise them.

Main parts of the software are in Java. We decided to equip the system with a Java Servlet interface. This enables easy integration to commercial services and mobile communications. The user interface is an applet, which transmits the search criteria to the Servlet for processing. There is also a possibility to run the system as a back-office tool, if the calculation times increase intolerably as the system is taken into commercial use. In that case the queries would be stored and alerts to the users of the system would be sent, when products that fulfil their needs are found from the database. To increase scalability, the advertisements can be pre-processed in the back-office to better conform to the selection logic.

Currently we use a Prolog based fuzzy logic engine for inferencing. There is also a chance of using an active database system incorporating fuzzy reasoning [12]. At this phase, we store the car information gathered from different sources in local databases. Later on agents can be added to the system for the retrieval and updating of information.

6. Business Benefits

SmartPick demonstrates how domain knowledge and qualitative reasoning can be used to obtain significant added value to advertisement and eCommerce portals and their customers. The main stream of current services is limited to simple search facilities by categories, crisp properties and string search. Compared to them SmartPick brings a real improvement. SmartPick is based on the qualitative analysis of the mediated goods.

The main benefit from SmartPick for a customer is that the selection of relevant items gets greatly enlarged. The system does on behalf of the customer the laborious task of systematically checking all qualitative conditions for each available item.

In the domain of used cars, the system can recognise car models from several brands based on the user requirements. The user does not need to know specific properties of various car models. In fact, he can totally skip the choice of models. If in traditional car portals the choosing starts from specifying the car models, in SmartPick one gets the list of suitable models as a response to setting his qualitative criteria.

While the user makes adjustments, the system updates the list of included models already before query launching. This helps to avoid null queries as the user quickly sees how adjustments affect to the available selection.

The user can select an item as a prototype and let the system find comparable other items. As for the car sales use case, this means selecting a model to obtain comparable cars. This is another way of enlarging the selection.

Our approach makes possible the implementation of natural concepts as aggregates from several product properties, e.g., city car. The concepts of the system can be made very close to the ones that the user is using in his personal evaluation process. This makes the service more usable and facilitates the decision making process.

For advertisers, qualitative analysis of user queries would provide a chance for better focused marketing. For instance, if safety is in the focus of the user, paid advertisements of closely matching, especially safe cars, could be shown. The approach enables collecting statistics about the qualitative criteria required by the users and selling these criteria, e.g., to distributors. They can use it when they plan future selection or marketing.
7. Conclusions

This paper described an innovative way to combine domain knowledge and qualitative reasoning in an intelligent advertisement service. The developed system, SmartPick, uses fuzzy logic as the formalisation for evaluating the suitability of advertised products to the qualitative criteria given by the user. In order to do so, the system does not limit itself to the contents of advertisements but uses also other knowledge resources of the domain. This far the tests have shown the approach viable and practical.

The underlying metrics of SmartPick could be extended also for dynamic generation of qualitative comparison tables for selected products. These comparisons could be made personalised along the criteria set by the user. Until now we have tested the approach in the domain of used car sales. This far the quality related concepts have been modelled manually or by using self-organising maps. In order to make the technology easily transferable to other advertising domains, like electronics, housing and apartments, we plan to develop semiautomatic modelling tools.

Acknowledgements

The authors would like to thank professor Seppo Linnainmaa for his valuable support for the work, Martti Merilinna from Tekniikan Maailma for letting us use their car data resources, and Russell Sladek for doing the language check.

References