Design and Implementation of a Data Warehouse for Quality Management, System Evaluation and Knowledge Discovery in the Medical Domain

Martin Atzmueller[1], Stephanie Beer[2], Alexander Hörnlein[1], Ralf Melcher[2], Hardi Lührs[2], and Frank Puppe[1]

[1] University of Würzburg, Department of Computer Science, Am Hubland, 97074 Würzburg, Germany
{atzmueller, hoernlein, puppe}@informatik.uni-wuerzburg.de
[2] University-Hospital of Würzburg, Department of Gastroontologics, Josef-Schneider-Str. 2, 97080 Würzburg, Germany
{beer_s, melcher_r luehrs_h}@klinik.uni-wuerzburg.de

Abstract. System evaluation and quality management of intelligent systems are essential tasks. In this paper, we describe a data mining approach for handling these: The techniques are based on a data warehouse that contains data from the intelligent system to be evaluated and from external sources. We discuss the integration of different heterogenous knowledge sources into the data warehouse and the application of the collected knowledge for the tasks outlined above. Additionally, we describe how the data can be applied for knowledge discovery. The context of our work is given by an intelligent documentation and consultation system in the medical domain of sonography. We provide several real-world examples demonstrating the applicability and benefit of the presented approach.

1 Introduction

Intelligent documentation systems are in wide-spread use, for example, in the service support domain for technical applications, or in the diagnostic domain for medical applications. For the latter, systems that cover both documentation and consultation features are often utilized. For evaluating the input – output behavior, and for quality management, manual evaluations by domain specialists are then usually applied. For these, the documented output of the system is compared with a (manually acquired) gold-standard. Then, the performance of the system can be concisely evaluated, even in complicated cases.

However, another important quality parameter that is orthogonal to such evaluations is given by the performance of the users of such a system corresponding to the *input quality*. In order to tackle this issue we need to consider the persons that are entering the input data before the system provides an output based on this data. If the entered input findings are not consistent, for example, then this may lead to incorrect conclusions that were not accounted for by the designers of the system: The specific situation may not be captured in the applied knowledge base which can significantly decrease the quality of the system. In a way, both these evaluations options can complement each other, since effects detected for one option can often lead to causes by the respective other option.

Since a purely manual evaluation for such purposes is usually time-consuming and costly, semi-automatic techniques are often a promising option. In this paper, we describe a data mining approach that is based on the implementation of a data warehouse [1] containing the data from the system to be evaluated and other external sources. The collected data can then be used for various automatic analysis options, and both the evaluation and the quality management objectives can easily implemented. As another benefit, the data warehouse can then also be directly used for knowledge discovery. This enables a multifunctional application of the data warehouse which is quite attractive for many domains and also rather cost-efficient. We describe the application and implementation of the proposed approach in the medical domain of sonography (ultrasound). We present examples from the real-world system SONOCONSULT [2,3], which is a multifunctional knowledge-based system for sonography, which has been in routine use since 2002 documenting more than 12000 patients in two clinics. An evaluation [3] of the diagnostic accuracy, acceptance, and clinical indicated very good results.

The rest of the paper is organized as follows: We first present the individual data mining methods of the proposed approach in Section 2, introduce the medical application context of the SONOCONSULT [3] system, describe the concrete implementation of the techniques, and finally provide first examplary results of their application. After that, we discuss the presented approach in Section 3. Finally, we conclude the paper with a summary in Section 4, and point out interesting directions for future work.

2 Method: Design, Implementation, First Results

In this section, we provide an overview on the individual methods of the proposed approach, and describe their implementation exemplified by case studies in the context of the SONOCONSULT system. As outlined above, we distinguish three main objectives for the presented approach:

1. System evaluation: Comparing and evaluating the input – output behavior of the system using external data for assessing the system solutions.
2. Quality management: Assessing the input *quality*: In the medical domain this corresponds to the *documentation quality* of the users (examiners).
3. Knowledge discovery: Correlating various parameters automatically in order to detect interesting patterns, for example, which diagnoses affect which laboratory parameters of blood samples.

In the following sections, we discuss these issues in detail. Furthermore, we exemplify the techniques by clinically interesting findings that were discovered using the proposed techniques. All the methods, specifically the quality management and knowledge discovery steps were implemented using the VIKAMINE system [4] – a versatile environment for knowledge-intensive data mining. The system provides both interactive and automatic methods, using subgroup discovery [5,6] as the core mining technique. The powerful visual methods were regarded as especially helpful during the implementation of the presented approach.

2.1 System Evaluation - Solution Profiling

The input and output relations for a given set of cases can be transparently evaluated by domain specialists that provide a gold standard for the solutions of the system. Then, the acquired cases are compared by a before-after strategy, that needs to be done manually by the domain specialists. Such a procedure usually works relatively well,e.g., [3]. However, providing the gold-standard takes a lot of time and is thus very cost-intensive. Therefore, other options that do not need to rely on a domain specialist are promising. Additionally, a continous monitoring of all the possible test cases would potentially require an unlimited set of expert-rated solutions, which is rather unfeasible.

We provide a data mining method that relies on external data from other data sources (examinations). If all the available data has been integrated into a data warehouse, then the evaluation of the input – output relations is straight-forward using a gold-standard given by laboratory and/or other examinations, for example, magnetic resonance imaging or CT tomography. The accumulated gold-standard data is then automatically compared to the solutions of the system in order to identify potentially incorrect solutions that were documented using the intelligent system.

However, these solutions may be incorrect due to two different reasons. First, the solution of the system may be wrong, and secondly, the provided input findings may be wrong and/or inconsistent with respect to the true input description. In the first case, we need to refine the knowledge system, whereas in the second case we need to make sure, that the input findings are entered in the correct way, e.g., by tutoring sessions or by special training of the users. In order to clarify such situations, we also need to consider the performance of the users that provide the input to the system, as discussed in the next section. Using the solution approach we can only detect situation of the first kind, therefore both analysis options complement each other.

In the following we discuss preliminary results obtained from the SONOCONSULT data. By integrating different data sources into the warehouse it is possible to measure the conformity of sonographic results with other methods or inputs.

Total Case Number	SONO CONSULT Diagnoses	SAP Diagnoses	% Conformity with SONO CONSULT	CT/MR Diagnoses	% Conformity with SONO CONSULT	Discharge Letter Diagnoses	% Conformity with SONO CONSULT
Liver cirrhosis							
16	12	6	20	1	33	9	50
Liver metastasis							
28	16	11	65	15	87	17	94

Table 1. Conformity of various sources of diagnosis input. Correlation of the different sources with SONOCONSULT diagnoses.

Table 1 shows the correlation of SONOCONSULT based diagnosis with CT/MR, diagnoses listed in the discharge letter and diagnoses contained in the hospital information system for a first number of cases. It was quite interesting that the conformity between SONOCONSULT based diagnoses with the diagnoses contained in the hospital information system was quite low. Evaluating this issue it was obvious that various diagnosis were not listed in the hospital information system because they were not revenue enhancing. Therefore, we looked at the accordance with the discharge letters which were found to be highly concordant at least for the diagnosis of liver metastasis.

Liver cirrhosis is more awkward to be diagnosed with ultrasound and has to be in a more advanced stage. Therefore, some of the discharge diagnoses "liver cirrhosis" were only detected using histology or other methods. In one case liver cirrhosis was listed in the hospital information system but was neither found with ultrasound nor in the discharge letter. It came out that the input was performed by another department (neurology), for which documenting the disease was not really relevant.

These first results of the correlations of diagnoses of various input sources indicate that there is a promising high conformity between SonoConsult and the discharge letters. However, for further quality improvement the correlation with other imaging techniques is very important. With a larger number of cases it should be possible to measure the sensitivity of different techniques for various diagnoses in more detail.

2.2 Quality Management

For the task of quality management or quality control we consider all the users individually that can enter findings for generating cases with a specific solution. In the context of the medical documentation system SONOCONSULT, the users are sonographic examiners that document the case by entering the specific findings they identify on the sonographic images. Since the sonographic examination is highly subjective and dependent on the experience of the examiner, the quality management is essential in order to identify examiners that deviate from the norm, if we assume a "similar" share of patients. Ultrasound is a method which is strongly dependent on the examiner's degree of knowledge. Therefore, it is interesting to know how well the results of the examination agree between different examiners and with the results of other methods.

For the quality management step we build user profiles statistically and test for all solutions, for which examiners have a significant different distribution: Essentially, we compare the frequency of documented diagnoses depending on the examiner. This method is easily implemented using subgroup discovery by regarding the examiner as the target variable, and considering all diagnoses and/or findings as independent variables. Table 2 shows the results of this analysis for various diagnoses.

The results show that there are some major differences in the frequency of diagnosing a specific disease for the different examiners. For example, several examiners do not document certain diagnoses at all ("aorta sclerosis, not calcified" or "portal hypertension"). The principal cause for this discrepancy may lie in the incident that the different examiner work in different departments and therefore the patients stock of each sonographer has a different distribution of diseases. Since the data warehouse is continously being extended, we plan to investigate these issues in more detail given a larger number of cases in more detail.

Diagnoses	All		Examiner 1		Examiner 2		Examiner 3		Examiner 4	
	F	%	F	%	F	%	F	%	F	%
All	2498	100	757	34	104	4,7	392	17,6	359	16,1
fatty liver	683	27,3	212	33,3	20	3,1	136	21,4	117	18,4
liver cirrhosis	42	1,7	22	52,4	0	0	15	35,7	0	0
aortic sclerosis, non calcified	29	1,2	3	10,7	0	0	0	0	18	64,3
aortic sclerosis, calcified	510	20,4	96	19,7	27	5,5	126	25,8	82	16,8
ascites	160	6,4	60	41,1	4	2,7	27	18,5	16	11
cholezystolithiasis	345	13,8	107	35,7	13	4,3	41	13,7	47	15,7
chron. deg. kidny disease	219	8,8	66	30,3	9	4,1	50	22,9	45	20,6
gut disease	35	1,4	10	28,6	2	5,7	18	51,4	1	2,9
mass/liver	119	4,8	39	33,9	3	2,6	21	18,3	25	21,7
obstructive cholestasis	15	0,6	3	23,1	0	0	6	46,2	3	23,1
lymphnode intraabdominal	33	1,3	16	59,3	2	7,4	3	11,1	1	3,7
pleural effusion	128	5,1	31	27	3	2,6	31	27	24	20,9
portal hypertension	35	1,4	16	45,7	0	0	11	31,4	1	2,9
prostata disease	334	13,4	55	18,1	4	1,3	50	16,4	79	26
liver size = very enlarged	178	7,1	25	15,8	3	1,9	26	16,5	35	22,2

Table 2. Diagnostic profiles for several examiners. The rows denote different diagnoses, for which the column *F* specifies the absolute and % the relative frequency.

2.3 Knowledge Discovery

A data warehouse [1] typically contains data from different heterogenous sources [7] that need to be accumulated, standardized, and finally imported into the data warehouse [8]. As we have outlined above, we have integrated several heterogenous data sources into the clinical data warehouse ranging from structured data records containing the examination data from the sonographic records, the various laboratory parameters, the final diagnoses for billing, but also unstructured data given by the textual discharge letters. From these, several additional data about further examinations is extracted.

As always in the data mining process (c.f., [8]), the design and implementation of the data warehouse took a lot of effort, since all the data neglecting the SonoConsult data needed to be extracted from legacy database systems, and a lot of time needed to be invested into the data cleaning approach. The data warehouse was completed after an initial design and several incremental refinement cycles for which the data sources and the selected data needed to be adapted and tuned.

Then, besides the already described methods, the application for knowledge discovery purposes is also very attractive: It enables a transparent and easy access to several independent knowledge sources and their combined analysis and mining. It is then not only possible to check and improve the quality of ultrasound examination but also to use it for knowledge discovery in order to find out, for example, which laboratory parameters of blood samples are altered with a specific diagnoses. Table 3 shows an exemplary selection of simple but clinically interesting relations between the diagnosis liver cirrhosis and several laboratory parameters.

Subgroup Size	True Positive	False Positive	Laboratory Parameter (Subgroup)	Population (Defined Population)	Relative Gain	Significance Level (P)
761	63	698	total-Bilirubin=HH	4304	3,11	<0,0000001
769	64	705	Cholinesterase=L	3542	2,53	<0,0000001
415	39	376	Quick=L	4199	3,56	<0,0000001
966	53	913	GOT (ASAT)=HH	4287	1,71	<0,0000001
18	6	12	Cholesterol=LL	3866	15,03	<0,0000001
887	44	843	Alk. Phosphatase=H	4359	1,46	0,000000000006
481	30	451	Albumin=L	4315	2,07	0,000000000008
40	7	33	HbA1c=LL	2138	8,51	0,000000000012
335	21	314	Albumin=LL	4315	2,08	0,000000017016

Table 3. Exemplary clinically interesting subgroups for liver cirrhosis and various laboratory parameters.

The results shown in Table 3 indicate, for example, a reduced synthetic efficacy of the liver, e.g. decreased cholinesterase, reduces blood clotting (Quick/INR); also albumin are all significantly reduced whereas the parameters indicating cell loss, e.g. GOT and alk. phosphatase are increased. The discovered relations are highly significant. Since the absolute frequencies are quite low in some cases, we plan to repeat this analysis using a larger number of cases in order to increase its statistical power further.

3 Discussion

First results of the application of the presented approach demonstrate its effectiveness and applicability for the sketched scenario. The validation of solutions for system evaluation already shows a high share of correct diagnoses. However, due to the highly subjective documentation procedure of the input diagnoses, we also suspect a dependency on the experience of the examiner. If the results of the examiner profiling study are only attributable to this specific issue, or if they depend on other confounding factors, for example, different patient distributions due to different departments needs to be clarified using a larger number of cases.

The main advantage of the data mining approach using a data warehouse compared to the manual validation approach is its cost-effectiveness, ease of use, and potentially continous application throughout the life-cycle of the intelligent system. Ultimately, it can be automated and can provide important feedback to certain types of user, for example, inexperienced examiners. Then, the quality of the documentation and of the input findings can be significantly increased.

As we have seen, a further attractive application is given by (semi-automatic) knowledge discovery, after the necessary data has been acquired and has been integrated into the data warehouse. Thus, the data warehouse serves as a multifunctional tool for an integrated approach from knowledge acquisition, quality control, monitoring, and further knowledge discovery.

4 Conclusion

In this paper, we have presented an approach for the evaluation, quality management, and knowledge discovery of intelligent documentation and consultation systems using a data warehouse. We have discussed the design and implementation of the data warehouse exemplified by a real-world system in the medical domain. We have shown, how the different objectives are implemented and have discussed their (clinical) impact in several examplary case studies that are nevertheless clinically relevant. The results indicate, that the presented approach is well suited for the evaluation and quality management. Furthermore, the knowledge collected within the data warehouse also provides for an ideal basis for further knowledge discovery by relating the different data sources.

For future work, we aim to integrate the system into the standard medical procedure. Furthermore, an extended integration of textual content contained in the different discharge letter should also provide for an increased benefit and utility of the presented approach. Additionally, we plan to extend the analysis to further knowledge sources. Of course, we also need to apply the techniques to a growing number of cases throughout the continous increase in the size of the data warehouse.

Acknowledgements

This work has been partially supported by the German Research Council (DFG) under grant Pu 129/8-2.

References

1. Kimball, R., Ross, M.: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. John Wiley & Sons, Inc., New York, NY, USA (2002)
2. Huettig, M., Buscher, G., Menzel, T., Scheppach, W., Puppe, F., Buscher, H.P.: A Diagnostic Expert System for Structured Reports, Quality Assessment, and Training of Residents in Sonography. Medizinische Klinik **99**(3) (2004) 117–122
3. Puppe, F., Atzmueller, M., Buscher, G., Huettig, M., Luehrs, H., Buscher, H.P.: Application and evaluation of a medical knowledge-system in sonography (sonoconsult). In: Proc. 18th European Conference on Artificial Intelligence (ECAI-08), Prestigious Applications of Intelligent Systems (PAIS-08). (2008) 683–687
4. Atzmueller, M., Puppe, F.: Semi-Automatic Visual Subgroup Mining using VIKAMINE. Journal of Universal Computer Science **11**(11) (2005) 1752–1765
5. Klösgen, W.: 16.3: Subgroup Discovery. In: Handbook of Data Mining and Knowledge Discovery. Oxford University Press, New York (2002)
6. Atzmueller, M., Puppe, F., Buscher, H.P.: Exploiting Background Knowledge for Knowledge-Intensive Subgroup Discovery. In: Proc. 19th Intl. Joint Conference on Artificial Intelligence (IJCAI-05), Edinburgh, Scotland (2005) 647–652
7. Kerkri, E.M., Quantin, C., Allaert, F.A., Cottin, Y., Charve, P., Jouanot, F., Yetongnon, K.: An Approach for Integrating Heterogeneous Information Sources in a Medical Data Warehouse. Journal of Medical Systems **25**(3) (2001) 167–176
8. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publisher (2000)