USE OF AN E-R DIAGRAM IN THE DESIGN OF A FEATURE DICTIONARY FOR A MULTI-DOMAIN MEDICAL KNOWLEDGE BASE

F. Naeymi-Rad[1], T. Koschmann[2], R Rosenthal[1], D. Trace[1], S. Naeyini-Rad[1], J. Swanson[1], C Lee[1], R Carlson[3], and M.H. Weil[1], M. Evens[3].

[1]University of Health Sciences/The Chicago Medical School, North Chicago, IL 60004
[2] Xerox Al Systems, Des Plaines, IL 60013
[3] Illinois Institute of Technology, Department of Computer Science, Chicago, IL 60616

ABSTR ACT

Past research has shown that the performance of a Bayesian system is crucially dependent on a consistent set of features. Because different terminology is often used by physicians of difficult specialties or locations to refer to the same feature, it is necessary that the knowledge base provide a means to compensate for discrepancies in terminology or definition. For this purpose, a new dale base called the Feature Dictionary has been designed with an E-R model. This dictionary will provide a means for transcending the medical "language barrier" by recognizing the entry of a feature into the knowledge base no matter which of the common terms the particular expert has used. For instances in which the expert comes across an unfamiliar term in the system, he or she will be able to call up a definition of up to four lines in length which will describe exactly what the feature is to which the term refers. The feature dictionary would enable us to merge existing MEDAS databases, evaluate and compare knowledge bases, or translate between MEDAS knowledge bases and medical knowledge bases of different types. II could be used in educating health care professionals and in collecting and analyzing dale for epidemiological research independent of the MEDAS system. The Feature Dictionary will be used in conjunction with the MEDAS (Medical Emergency Decision Assistance System) knowledge engineering tool TOOLBOX. The MEDAS TOOLBOX has been used in the design of diagnostic consultants for a variety of medical domains.

THE ENTITY-RELATIONSHIP MODEL

An entity-relationship model has been used to design and develop a relational database for MEDAS (the Medical Emergency Decision Assistance System) on the DEC System 20. This database has served as the foundation for a set of knowledge engineering tools called the TOOLBOX. The TOOLBOX has made it possible for a medical expert to add new disorders and new features to the database easily and a number of physicians and medical students are now taking advantage of [his opportunity. The microcomputer version now being developed in cooperation with NASA is likely to lead to further proliferation of MEDAS databases. We are now designing a feature dictionary that will contain definitions of symptoms, normal/abnormal ranges for standard medical tests, feature interactions, units of measurement, hierarchical categories, synonyms and antonyms, and lost factors.

MEDAS is a multimembership Bayesian system developed at the University of Southern California by Ben-Bassat a al. [1] with the support of the Institute of Critical Care Medicine under the leadership of Max Harry Weil. Four years ago the project moved to the University of Health Sciences / The Chicago Medical School. Since then MEDAS has been rewritten to run on the DEC System 20 and to use the relational database system ACCENT-R.

The Entity-Relationship (ER) model, developed by Peter Chen .f21 has become the fundamental standard for database design. In using Chen's approach the designer begins by isolating the important entities, the central concepts, in the database. For MEDAS these are disorders and features - a feature is any item of information about the patient that can be used in making a diagnosis, beginning with age sex, and other items that appear in the physical examination and medical history, the chief complaint and other current symptoms, and test results. Disorders are grouped in feature categories. Figure I shows an ER diagram for MEDAS. Entities are represented by rectangles, the attributes associated with an entity by ovals, the relationships by diamonds. The most important relationship in MEDAS is the one that describes a disorder in terms of its diagnostic features. Since MEDAS uses a pattern recognition strategy, it is essential for the database to record the conditional probabilities that a given symptom will be present in a patient with the disorder.

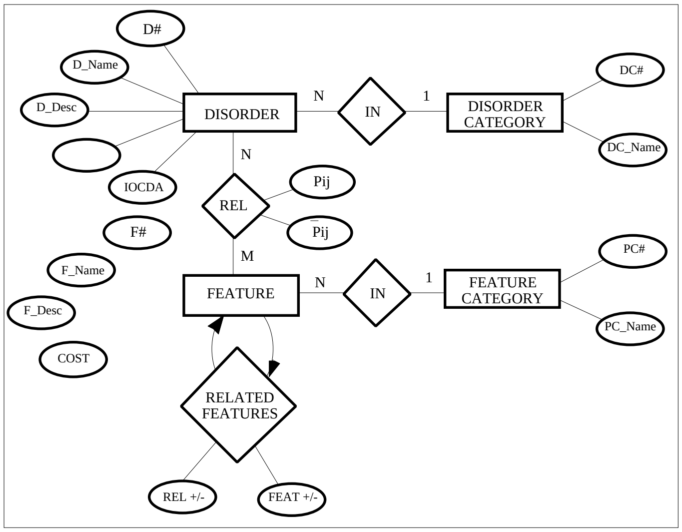

Figure 1: ER-Diagram for MEDAS

THE TOOL BOX AND THE FEATURE DICTIONARY

Using a relational database, a set of knowledge engineering tools called the TOOLBOX was implemented to make it easier for experts to enter and modify data for the MEDAS knowledge data base [3, 4, 5, 6]. The TOOLBOX has eliminated the need for the author of a medical knowledge base to understand the data organization or the relationship of entities within the Bayesian system. Indeed, the medical expert need only concentrate on entering and updating information concerning the two entities, the disorder file and the feature file.

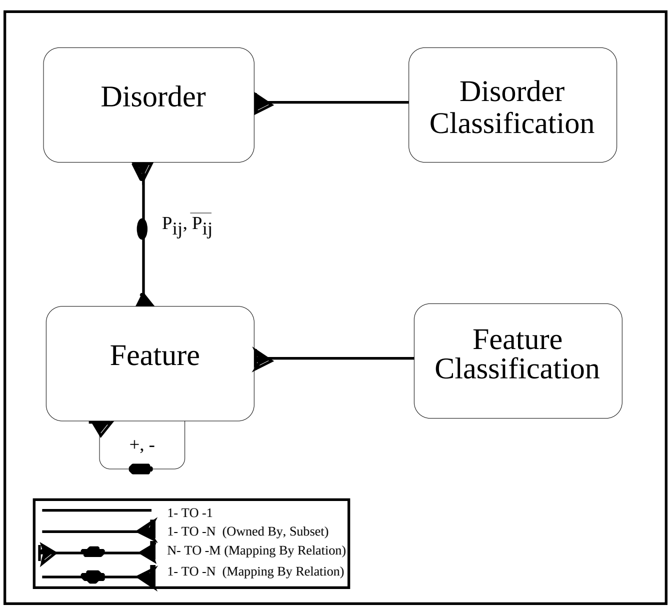

Figure: 2
ER Diagram of the MEDAS Data Base
Using New Graphic Constructs

The TOOLBOX has proven to be an invaluable aid in the production and entry of disorder patterns for the MEDAS system. In the span of six months one expert (Rosenthal) entered over 4500 features and developed 170 disorder patterns. Additionally, our team has refined the disorder and feature categories info a more consistent and usable framework. Another benefit of the TOOLBOX is that it has made the MEDAS system completely domain independent, so that it can now maintain a number of knowledge bases for any of the medical specialties.

When experts create knowledge for the medical knowledge base for MEDAS, it is essential that rhe system recognizes a consistent set of features not only within a single domain but also when features interact among a variety of domains. Dombal's research [7] has shown that the Performance of a Bayesian system is crucially dependent on the consistent labeling of features. The terms experts use to refer to particular features vary, depending on an expert's specialty, preference, or regional vernacular. These differences create the need for a feature dictionary which can translate medical terms into those meaningful to the system, and provides definitions for terminology foreign to the user. Figure 3 shows an ER diagram of the feature dictionary supporting multi-domain knowledge,

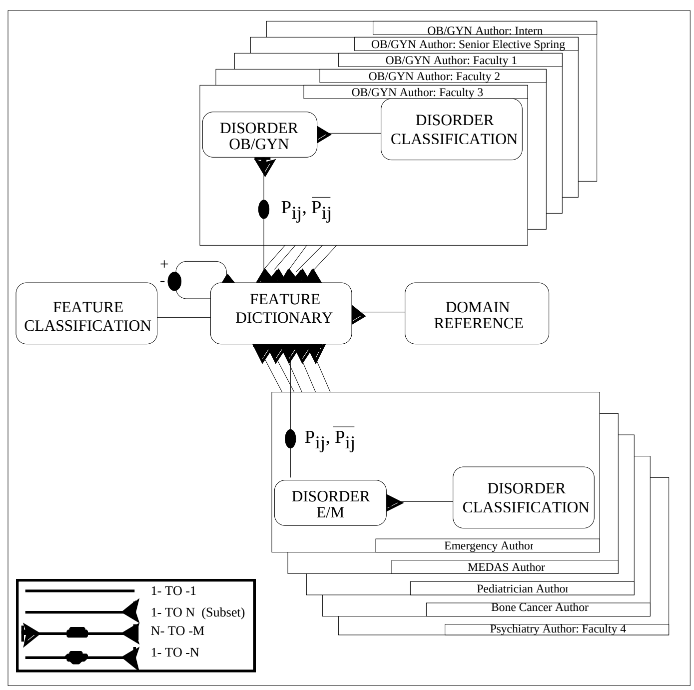

Figure 3
The Feature Dictionary Supporting Multi-Domain Knowledge

ADVANTAGES OF THE FEATURE DICTIONARY

The feature dictionary can not only serve individual physicians using MEDAS, but it has important long-term implications. The feature dictionary is an essential tool in merging existing MEDAS data bases, evaluating and comparing existing knowledge bases, educating health care professionals, and providing a means of communicating between multidomain MEDAS knowledge bases. The feature dictionary could even permit a translation between a MEDAS knowledge base and a medical knowledge base of a completely different type, such as a rule- based EMYCIN system [8]. The feature dictionary would also be invaluable for the collection and analysis of data for epidemiological research independent of the MEDAS system

The feature dictionary would facilitate the evaluation of the performance of competing knowledge bases. Without the feature dictionary, patient data would have to be interpreted and re-entered for each knowledge base, since each has its own set of terminology. The feature dictionary would be used to perform this interpretation automatically.

In the same manner, the feature dictionary would allow us to compare knowledge bases for the same domain. One could always compare knowledge bases externally by processing patient data through them and comparing the disorder rankings they produce. One needs the feature dictionary to translate terms into common feature names for easy comparison.

We have used MEDAS in senior medical electives. When students are creating knowledge it is essential that they use feature sets defined by an expert and that they understand each feature and its definition. The feature dictionary makes this information readily available. Once the students have developed their own MEDAS knowledge bases, the attending physician supervising them can evaluate the students' knowledge bases, internally as well as externally.

The expert system has an important role to play in translation. The feature dictionary would allow us to translate already-entered patient information into formats demanded by different medical expert systems developed using M- I or EMYCIN.

Over the long run, we plan to update probabilities in MEDAS using data collected from clinical data bases. In order to take advantage of existing clinical data bases, the feature dictionary is needed to translate patient data into the standard format of MEDAS knowledge.

The feature dictionary can be used, independent of MEDAS, as a framework for the collection of clinical data for use in epidemiological research. Currently there is no registry of consistent clinical data available on a nationwide level.

CONCLUSION

Developers of expert systems have been slow to take advantage of database technology. We find this particularly distressing since for the MEDAS project, at least, the use of a DBMS has brought rich rewards. The relational database management system has made it possible for us to provide disorder and feature pattens at any point, to store competing knowledge bases, and to associate items of knowledge with their author. The use of the Entity-Relationship paradigm converted the often painful chore of database design into a very simple and natural process.

The design of the feature dictionary was inspired by the data dictionary utilities now provided by many database management systems. The feature dictionary will provide a standard and consistent set of features with prerecorded feature interactions and produce detailed definitions of features along with normal / abnormal values for test results and other information useful for interpreting the feature. These detailed definitions can be obtained whenever desired by striking a simple key. The feature dictionary will make it possible to evaluate MEDAS knowledge bases, to compare competing knowledge bases, and to update the MEDAS knowledge base from clinical databases. Above and beyond its value for the MEDAS system, the feature dictionary has the potential for providing for the first time a standardized set of features for collecting clinical data or normalizing existing clinical databases for epidemiological studies.

REFERENCES

1. Ben-Bassat, M., Carlson, R.W., Puri, U.K., Davenport, M.D. Schriver, J.A., Latif, M., Smith, R., Portigal, LR., Lipnick E.H., Weil M.H. "Pattern-Based Interactive Diagnosis of Multiple Disorders: The MEDAS System", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAMI-2, no. 2, March, 19X0.

2. Chen, P.P. "A Historical Perspective and Future Directions" Proceedings of the Conference on the

E-R Approach to Software Engineering, (1983), pp. 71-77.

3. Koschmann, T., Solomon, D., Naeymi-Rad, F., Evens, M., and Weil, M.H. "Relational Storage Techniques Applied to a Medical Expert System". Proceedings of the Conference on Intelligent Systems and Machines. Rochester, Ml. April 1984. pp. 208-212.

4. Koschmann, T., Evens, M., Naeymi-Rad, F., Lee C, Gudipati R., Weil, M.H., "Knowledge Engineering Tools for a Bayes;an Diagnostic Consultant", Proc. Symposium on Computer Applications in Medical Care, Baltimore, November, 1985, 274-2~.

5. Naeymi-Rad, F., Evens, M., Koschmann, T., Lee, C. Gudipati R., Kepic, T., Rackow, E., Weil, M.H., "Maintaining a Knowledge Base Using the MEDAS Knowledge Engineering Tools", Proc. Symposium on Computer Applications in Medical Care, Baltimore, November, 1985, 298303.

6. Naeymi-Rad, F., Koschmann, T., Trace, D., Kepic, T., Carlson C.R., Weil, M.H., Evens, M., "Expert Knowledge Base Designed Using ER-Modeling Technique," Proc. Symposium on Computer Applications in Medical Care, Washington, November, 1986, 51-55.

7. Dombal, F.T. de. "Computers in Medicine", Surgical Annual. Vol. 11, 1979.

8. Buchanan, B.G., Shortliffe, E.H., "Rule-based Expert Systems: the MYCIN Experiments of the Stanford Heuristic Programming; Project, Reading, Massachusetts, Addison-Wesley, 1984.