User Modeling for Tailored Genomic e-Health Information

Nancy Green[1], Tami Britt, Karen Jirak, David Waizenegger, and Xuegong Xin

Department of Mathematical Sciences, University of North Carolina at Greensboro, Greensboro, NC 27402-6170 USA
[1] nlgreen @ uncg.edu

Abstract. GenIE, a prototype intelligent system currently under development, is intended to assist genetic counselors by creating editable first drafts of genetic counseling information tailored for their clients. In this paper we describe GenIE's approach to constructing the normative component of a genomic user model, a model of the patient's genomic health. Also we describe future plans to include patient-tailored interactive graphics for conveying risk information in electronic documents drafted by the system.

1 Introduction

We are developing GenIE, a prototype intelligent system intended to assist genetic counselors by creating editable first drafts of genetic counseling patient letters. In the USA, genetic counselors meet with clients to explain, e.g., genetic testing, inheritance of genetic disorders, and risks of multifactorial diseases such as cancer. The client may be a genetics patient or parent/guardian of the patient. In current practice, the information and clinical services provided to the client/patient are summarized by the counselor in the patient letter, a one- to two-page tailored print document [1].

GenIE will create a first draft of this type of letter using natural language generation (NLG) techniques [2, 3]. The genetic counselor will provide information about a patient's case (such as symptoms, test results, and diagnosis) through GenIE's graphical user interface. (It may be possible to acquire some patient information from medical records as well.) The system will create a user model from this information and a generic domain model on genomic health. By *user model* we mean a model of the patient's genomic health, including genomic information about family members. A letter will be synthesized by applying linguistic knowledge on standard topics, organization and writing styles for this genre [4, 5]. A document editor will present the synthesized first draft for review and editing by the counselor.

One intended benefit of GenIE is to save the genetic counselor time by partially automating the letter writing process. In current practice, genetic counselors have limited time available for letter writing, and thus may use "cut-and-paste" methods to reuse pieces of previously written letters. Time constraints on members of this profession are not likely to lessen. As information about human genetics has increased at a rapid pace in recent years, so have genetic testing options; this is likely to increase demand for the

services of genetic counselors. In addition, GenIE might be useful as an educational tool to help physicians learn how to communicate with patients on genetics-related problems. (Note, however, GenIE's role is not to perform decision-support functions, but to help explain the medical experts' beliefs and reasoning to a lay audience.)

Documents drafted by GenIE could be presented (after being reviewed by the counselor) to their audience electronically, which would enable written text to be augmented with graphics for low-literacy and low-numeracy recipients. Although in current clinical practice visual aids (e.g., [6]) may be used at the client meeting, it is not feasible for today's writers to include tailored graphics in patient letters. Thus, in addition to the goal of generating written text, we are currently investigating generation of tailored interactive graphics to help explain probability-related information in patient documents presented electronically.

2 Genomic User Model

GenIE's user model will integrate two components, one which we term an *affective component* and the other a *normative component*. The affective component represents variables relating to the client's appraisal of issues represented in the normative component [7]. The focus of this paper, the normative component of the user model, integrates generic genomic health information with evidence and expert judgments about the client's case. In this section, we describe the creation and structure of the normative component of the user model.

2.1 Corpus Project

The *normative* part of the user model is constructed by updating a generic genomic health domain model with patient-specific information. The design of the domain model has been informed by an encoded corpus of twenty-one genetic counseling patient letters. The letters were encoded by members of our project using a coding scheme we developed for representing the biomedical content in genetics counseling letters [8]. The coding scheme provides a small set of tags (History, Genotype, Mutation-event, Biochemistry, Physiology, Symptom, Test Result, Probability, Karyotype, and Complication), the first eight of which were formally evaluated and shown to have very good intercoder reliability [8]. These tags are used to annotate phrases in a letter referring to causally or probabilistically related concepts. These concepts may be viewed as variables in a Bayesian network (BN) model [9, 10] of a letter's biomedical content. For example, a sentence from a letter published in [1] followed by its encoding is shown in Figure 1. (The term *proband* in genetics refers to the person of interest in a genetic study, in this case, a child patient.)

Philip was diagnosed as having a pseudoarthrosis of the left tibia and also … several café-au-lait spots both of which are features of neurofibromatosis.

Philip was diagnosed as having <*symptom-3.1*/*proband* a pseudoarthrosis> of the left tibia and also … <*symptom-3.2*/*proband* several café-au-lait spots> both of which are features of <*genotoype-3*/*proband* neurofibromatosis>.

Figure 1. Sentence from patient letter in original and in encoded form.

2.2 Genomic Domain Model

Generalizing from cases modeled in the encoded corpus, we constructed BN graphs representing several genetic disorders. (Since we are not experts in clinical genetics, reference books were sometimes consulted for clarification.) A corpus-based approach to design is being used so that the domain model will represent the biomedical concepts and reasoning of the genetic counselor at the level presented in the letters, i.e., at a level of detail designed for a lay audience. From the encoded letters, we determined relevant variables (of the types defined in the coding scheme), their domains, and network topology needed to represent a disorder. However, to implement the domain model as a functioning BN would require acquiring the full conditional probability tables for each BN variable. Since it is not our goal to provide decision-support capabilities, the domain model is not required to compute probabilities. Thus, to avoid the cost of acquiring numeric probabilities, we decided to implement the domain model as a qualitative probabilistic network (or QPN) [11, 12]. A QPN is an abstraction of a BN where qualitative constraints are specified instead of probability distributions.

Qualitative constraints for the domain model are specified in terms of the relations of qualitative influence, product synergy, and additive synergy as defined in [11, 12]. Types of qualitative constraints characterizing the domain of clinical genetics seen in our corpus include the following [13]:

- Arcs in a BN represent positive and negative qualitative influence. An example of positive qualitative influence is the relationship between genotype and symptom, i.e., a genotype with one or two abnormal alleles may cause a certain symptom.
- An example of negative product synergy is the relation of two genetic disorders to a symptom, i.e., where they are assumed to represent mutually exclusive possible causes of the symptom. Negative product synergy licenses a kind of intercausal reasoning (sometimes called *explaining away*). Another example of negative product synergy is inheritance of an autosomal dominant disorder. (Autosomal dominant disorders are caused by having one abnormal allele.) Since a child inherits two alleles, one from each parent, if a child has an autosomal dominant disorder then increasing belief that one parent is a carrier of the abnormal allele explains away (reduces) belief that the other parent is the carrier.

- Autosomal recessive inheritance illustrates zero product synergy. (Autosomal recessive disorders result when both alleles of a gene are abnormal.) Zero product synergy does not permit explaining away.
- The combined effect of a treatment and a genetic disorder on some symptom can be defined in terms of negative additive synergy, i.e., the treatment may diminish the effect of the disorder.

Figure 2 shows part of a domain model for Neurofibromatosis 1 (NF) constructed from the letter published in [1]. As shown in the model, the genotype for NF (G3) may cause symptoms of pseudoarthrosis (S3.1) and/or café-au-lait spots (S3.2). NF is an autosomal dominant disorder. Either parent may have transmitted the abnormal allele to the proband (G3). If the mother or father had inherited the mutation (G24.1 or G24.2, respectively), then that parent might present symptoms of NF (S24.2 or S24.3, respectively). However, another possibility is that the mother or father experienced a mutation event (M26.1 or M26.2, respectively) that affected the reproductive cell of that parent (G26.4 or G26.5, respectively) that contributed DNA to the child (G3). (In that case, the parent carrying the mutation would not present symptoms.) All arcs in the figure represent positive qualitative influence. The four possible causes of G3 are shown in a relation of negative product synergy (denoted by X-) to G3.

2.3 User Model Creation and Use

The normative component of a user model is created by updating the generic domain model with information about a specific patient's case, including relevant information on family members. GenIE's graphical user interface presents the genetic counselor with a pedigree tool for diagramming a patient's family tree, as shown in Figure 3. The graphical user interface also enables the counselor to set the values (with probability estimates) of history, symptom, test result, and genotype variables for any member of the family tree, as shown in Figure 4. For example, the generic domain model shown in Figure 2 could be updated from information entered through the graphical user interface to represent a patient who has pseudoarthrosis and café-au-lait spots (updating S3.1 and S3.2) and whose most likely diagnosis is NF (updating G3). In addition, the genetic counselor could enter the information that neither parent has symptoms of NF (updating S24.2 and S24.3), and (therefore) it is not likely that they have inherited NF (updating G24.1 and G24.2).

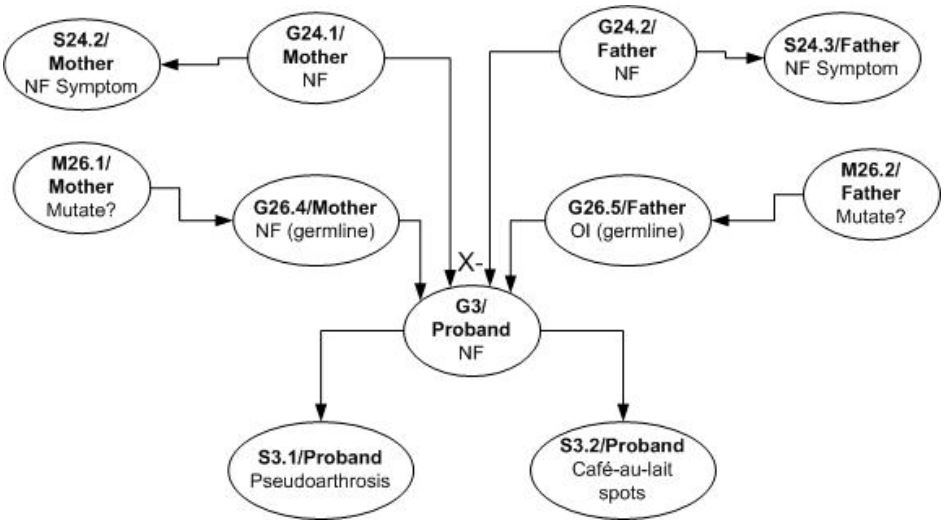

Figure 2. Domain model for NF.

Qualitative constraints may be applied to provide the natural language generator with derived information needed to draft a patient-tailored letter. For example, an explanation for why it is believed likely that one of the parents has germline NF can be derived from the beliefs that the proband has NF and that the proband's NF mutation must have been inherited, but that it is unlikely that either parent has inherited NF; thus, by elimination, one of the parents must have germline NF. Alternatively, if the generic domain model were updated to represent a case where the mother also has symptoms of NF, then the user model could help the natural language generator provide the following explanation: since she presents symptoms, it is likely that the mother has inherited NF, which she could have transmitted to her child. (Thus, belief in the other three possible sources of the child's NF mutation would be explained away.)

GenIE's NLG component has not been implemented yet. However, Figure 5 shows a prototype graphical user interface for the document editor that would enable the writer to review and edit the first drafts created by GenIE.

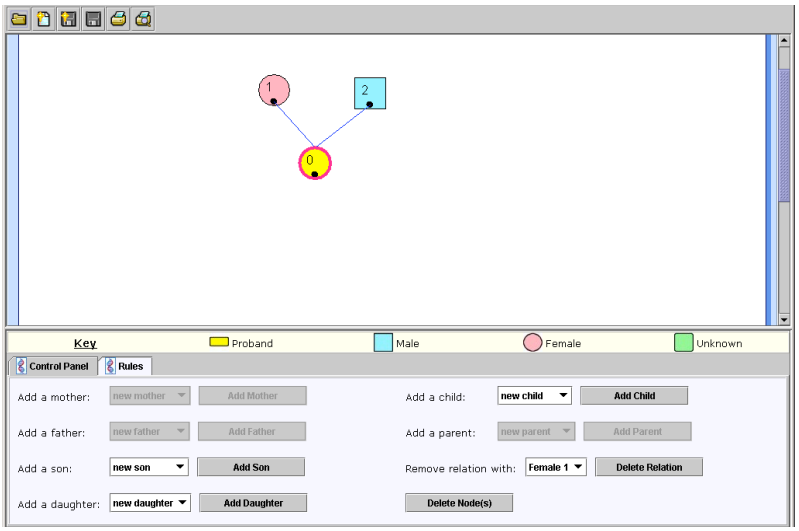

Figure 3. Pedigree drawing screen in GenIE prototype.

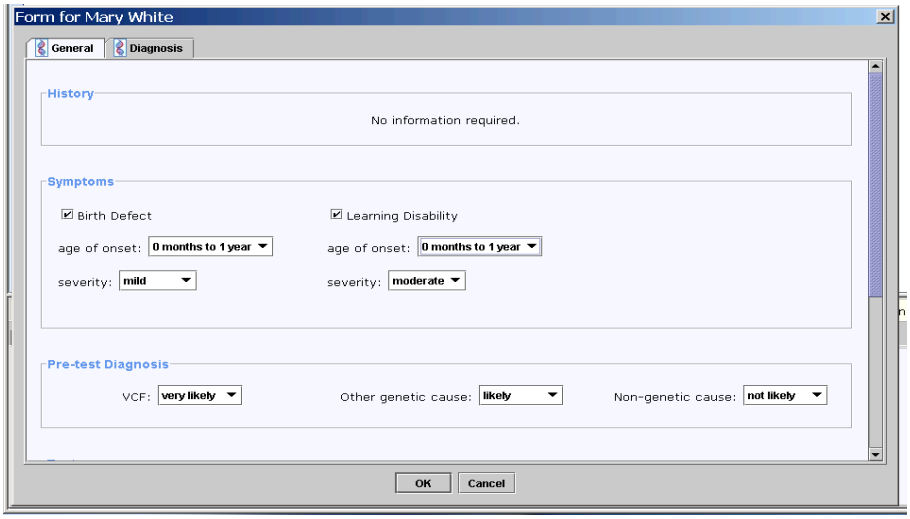

Figure 4. Screen for entering patient information in GenIE prototype.

3 Tailored Interactive Graphics

In genetic counseling, much of the information to be presented involves probability. Effective presentation of probability information to a lay audience faces numerous problems identified in the field of health risk communication (see, e.g., [14]). In addi-

tion to the goal of generating written documents for clinical genetics clients, we are interested in augmenting the text with computer graphics in order to address these problems.

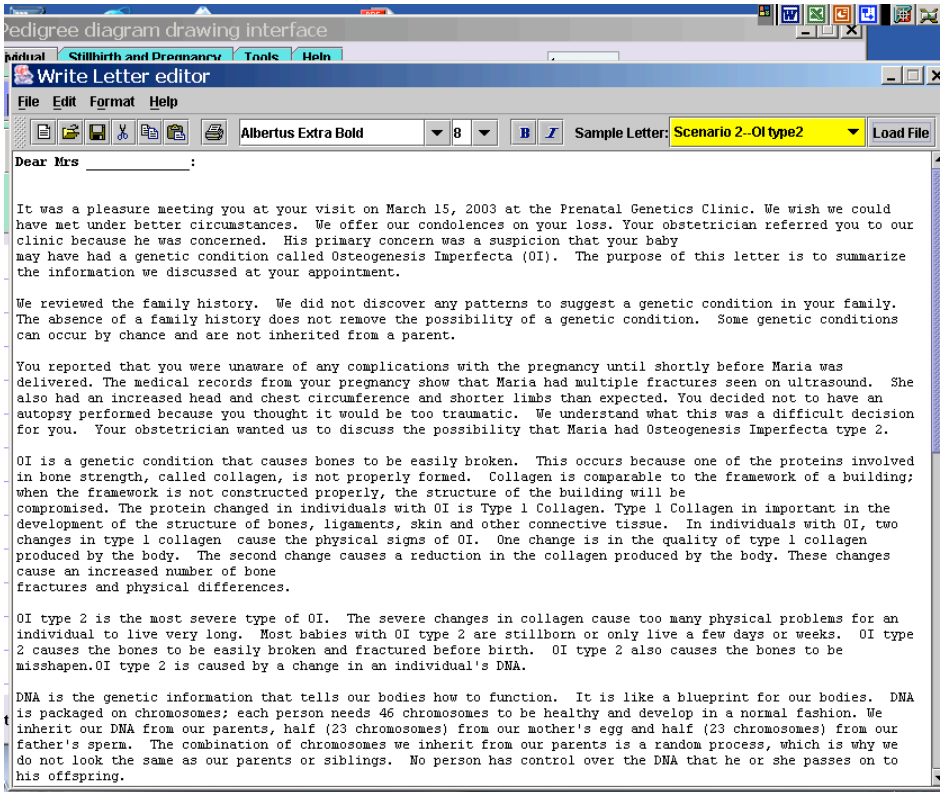

Figure 5. Editor for revising drafted letters in GenIE prototype under development. (Note: letter was drafted by a human author in order to demonstrate editing tools.)

As a starting point, we have designed a tailored interactive graphic that can be added to a generated document to illustrate simple autosomal inheritance. This is sometimes illustrated in genetics textbooks or counseling aids [6] with static pedigree diagrams representing the parents and four potential offspring, each illustrating a different possible combination of the two alleles from the parents. Such diagrams have several limitations. First, they are not tailored for a particular patient. Second, they do not convey pictorially the magnitude of the risk for and against a particular outcome. Third, they are subject to the fallacy of non-independent events. For example, a client who is told there is a 25% risk of having a child with a certain genetic disorder may mistakenly think that since she has had three children so far and none are affected then her next child will be affected.

We have designed and implemented an applet to address these problems. The applet presents two "wheels of fortune", one for each parent (Figure 6). The graphic is tailored by being initialized with the number of abnormal alleles (0, 1, or 2) in each parent's genotype according to the user model. (However, the user is able to reset the number of mutated alleles in a parent's genotype to explore other scenarios.) The magnitude of the probability that a parent will contribute a mutated allele to the child's genotype, and the complementary probability that the parent will *not* contribute a mutated allele, is depicted by dividing each wheel into two sections depicting the likelihood of the two possible outcomes.

The probabilistic nature of inheritance is conveyed through user-controlled animation. Each time the user presses a button (labeled *Spin* in Figure 6), the two wheels appear to spin and the applet randomly chooses the result of the spin (conveyed pictorially as the part of the wheel landing under a pointer). By running the animation several times, the user can see the non-deterministic nature of the outcome. In future work, we will evaluate the effectiveness of the applet.

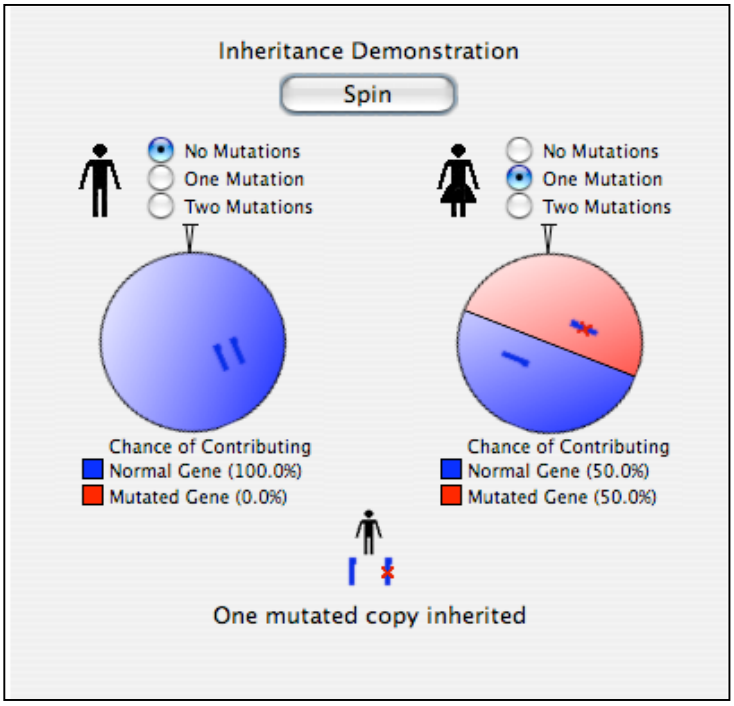

Figure 6. Screenshot of risk communication applet.

4 Related Work

MIGRAINE generated patient-tailored interactive explanations about migraine using information collected from the user (the patient) [15]. The HealthDoc project generated patient-tailored health-education documents using the medical record as user (reader) model [16]. The OPADE project generated user-adapted explanations of drug prescriptions for users with different information needs, e.g., the patient or the patient's healthcare provider [17]. Although these systems used patient data, they did not represent it in a way that supported reasoning about the relationships between medical problems and diagnosis or treatment, as in our user model.

5 Summary

We are currently developing GenIE, a prototype intelligent system intended to assist genetic counselors by creating editable first drafts of tailored genetic counseling patient letters. In this paper, we described the normative component of GenIE's user model, a model of the patient's genomic health. This part of the user model is created by updating a generic genomic domain model, implemented as a qualitative probabilistic network, with beliefs about the patient and the patient's family provided by the genetic counselor. This kind of user model supports derivation of qualitative inferences about the patient. In addition to its role in text generation, the user model will provide information for tailoring graphics for future versions of GenIE presenting documents in electronic media. In this paper we described a tailored interactive visual aid that we have developed to augment written material on inheritance risks.

Acknowledgments

This material is based upon work supported by the National Science Foundation under CAREER Award No. 0132821. Tami Britt also received support from DUE Award No. 9986757. The work of David Waizenegger was supported by an Undergraduate Research Assistantship in Fall 2004 from the University of North Carolina at Greensboro.

References

1. Baker, D.L., Eash T., Schuette, J.L., Uhlmann, W.R. Guidelines for writing letters to patients. *J. Genetic Counseling* 2002; 11(5):399-418
2. Cawsey, A.J., Webber, B.L., Jones, R.B. Natural language generation in health care. *J. Am. Med. Inform. Assoc.* 1997; 4(6):473-82
3. Hηske-Kraus D. Text Generation in Clinical Medicine – A Review. *Methods Inf. Med.* 2003; 42:51-60

4. Green, N. Analysis of linguistic features associated with point of view for generating stylistically appropriate text. *Working Notes of AAAI 2004 Spring Symposium on Exploring Attitude and Affect in Text*
5. Green, N. Argumentation in a Casual Probabilistic Humanistic Domain: Genetic Counseling. *International Journal of Intelligent Systems*. To appear
6. Stanislaw, C.L. et al. *Genetic counseling aids*. 4th ed. Greenwood, SC: Greenwood Genetic Center (2002)
7. Green, N. Affective Factors in Generation of Tailored Genomic Information. *Working Notes of User Modeling 2005 Workshop on Adapting the Interaction Style to Affective Factors*. To appear
8. Green, N. A Bayesian network coding scheme for annotating biomedical information presented to genetic counseling clients. *J. Biomedical Informatics* 2005; 38:130-144
9. Korb, K., Nicholson, A.E. *Bayesian artificial intelligence*. Chapman & Hall/CRC, Boca Raton, FL (2004)
10. Lucas, P.J.F., van der Gaag, L.C., Abu-Hanna, A. Bayesian networks in biomedicine and health-care. *AI in Medicine* 2004; 30(3):201-214
11. Wellman, M.P. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence* 1990; 44(3):257-303
12. Druzdzel, M.J., Henrion, M. Efficient reasoning in qualitative probabilistic networks. *Proc. 11th Nat. Conf. on AI* (1993) 548-553
13. Green, N., Britt, T., Jirak, K. Modeling the Communication of Uncertainty in Tailored Genetics Patient Information. Submitted
14. National Cancer Institute. *NCI Monographs* 25 (1999)
15. Mittal, V.O., Carenini, G., Moore, J.D. Generating Patient Specific Explanations in Migraine. *Proc. 18th Annual Symposium on Computer Applications in Medical Care*, McGraw-Hill Inc. (1994)
16. Hirst, G., DiMarco, C., Hovy, E., Parsons, K. Authoring and generating health-education documents that are tailored to the needs of the individual patient. In Jameson A, Paris C, and Tasso, C (Eds.), *User modeling: Proc. of 6th Int. Conf., UM97*. Vienna, New York: Spring Wien New York (1997)
17. de Rosis, F., Grasso, F., Berry, D.C. Refining instructional text generation after evaluation. *Artificial Intelligence in Medicine* 17 (1999), 1-36