13

Medical Concept Representation and Data Mining

Mila Kwiatkowska and Najib T. Ayas

Abstract. Medical data stored in clinical files and databases, such as patient histories and medical records, as well as research data collected for various clinical studies, are invaluable sources of medical knowledge. The computer-based data-mining techniques provide a tremendous opportunity for discovering patterns, relationships, trends, typical cases, and irregularities in these large volumes of data. The patterns discovered from data can be used to stimulate further research, as well as to create practical guidelines for diagnosis, prognosis, and treatment. Thus, a successful data-mining process may result in a significant improvement in the quality and efficiency of both medical research and health care services. Many studies have already demonstrated the practical values of data-mining techniques in various fields. However, in contrast with more traditional areas of data mining, such as mining of financial data or mining of purchasing records, medical data-mining presents greater challenges. These challenges arise not only from the complexity of the medical data, but more fundamentally from the difficulty of linking the medical data to medical concepts or rather medical concepts to medical data. Thus, although computerized medical equipment allows us to store increasingly large volumes of data, the problem lies in defining the meaning of the data and even more so in defining the medical concepts themselves.

This paper will address issues specific to medical data and medical data mining in the context of Dr. Kazem Sadegh-Zadeh's discussion of the typology of medical concepts. In his *Handbook of Analytic Philosophy of Medicine*, Dr. Sadegh-Zadeh outlines four main classes of medical concepts: individual, qualitative (classificatory), comparative, and quantitative. Moreover, he introduces a novel distinction between classical and non-classical concepts. We will explain how his typology can be utilized for conceptual modelling of medical data. Specifically we will illustrate how this typology can pertain to data used in the diagnosis and treatment of sleep disorders.

R. Seising and M.E. Tabacchi (Eds.): *Fuzziness and Medicine*, STUDFUZZ 302, pp. 191–203.
DOI: 10.1007/978-3-642-36527-0_13

13.1 Introduction

Medical data stored in clinical files and databases, such as patient histories and medical records, as well as research data collected for various clinical studies, are invaluable sources of medical knowledge. The rapidly increasing number of electronically stored patients' data provides a tremendous opportunity for data mining, the process of automatically discovering useful information (discovering knowledge) in large data repositories [4]. Data-mining techniques allow for discovering patterns, relationships, trends, typical cases, and irregularities in these large volumes of data. These newly discovered information and knowledge can be used to stimulate further research, as well as to create practical guidelines for diagnosis, prognosis, and treatment. Thus, a successful data-mining process may result in a significant improvement in the quality and efficiency of both medical research and health care services. Many studies have already demonstrated the practical values of data mining in various fields. However, in contrast with more traditional areas of data mining, such as mining of financial data or mining of purchasing records, medical data-mining presents greater challenges. These challenges arise not only from the complexity of the medical data, but more fundamentally from the difficulty of defining medical concepts. Medical concepts must be clearly defined in order to build appropriate data models in the data-mining process. Thus, although computers allow us to store and process increasingly large volumes of data, the problem lies in the creation of the suitable conceptual models for the data. These models should be unambiguously defined, and they should be explicitly connected with the related medical concepts. Evidently, the quality of the data-mining process depends on the quality of the conceptual data models and the quality of the data.

This paper addresses issues specific to conceptual modeling of medical data in the data-mining process. We will situate our discussion in the context of Dr. Kazem Sadegh-Zadeh's typology of medical concepts [17]. In his *Handbook of Analytic Philosophy of Medicine*, Dr. Sadegh-Zadeh outlines four main classes of medical concepts: individual, qualitative (classificatory), comparative, and quantitative. Furthermore, Dr. Sadegh-Zadeh divides medical concepts into classical concepts and non-classical concepts. We demonstrate how Dr. Sadegh-Zadeh's typology can be utilized for conceptual modeling of medical data. Specifically we illustrate how this typology pertains to concepts and data used in the diagnosis and treatment of sleep disorders. The paper is structured as follows. Section 13.2 defines the key issues in modeling of medical concepts: definition of a concept, characteristics of medical concepts, and computational representation. Section 13.3 describes the fundamental issues in medical data mining and provides the example of operational definition for obstructive sleep apnea. Section 13.4 presents the semiotic approach to conceptual modeling. The final section, Section 13.5, provides the conclusions and future work.

13.2 Modeling of Medical Concepts

In this section, we discuss the notion of a natural concept and the creation of computational models for natural concepts. First, we describe natural concepts and place our description in the context of Dr. Sadegh-Zadeh's categorization of "classical concepts." Second, we discuss concept modeling in the context of knowledge representation. Next, we focus on medical concepts and their characteristics such as context-dependency, changeability and imprecision.

Since our ultimate goal is to create computational data models for data mining, we approach the "concept" definition from a representational perspective. Accordingly, we view "concept" as a principle of classification. Furthermore, we approach the concept definition from a cognitive perspective. First, we ask two fundamental questions: "How do people classify objects into categories?" and "How do people mentally represent categories?" The answers to these questions are fundamental for the creation of computational models. The findings from cognitive psychology about human categorization have demonstrated that category learning and classification in the real world are different from the creation and classification of artificial categories (e.g., mathematical categories) [1, 14, 15]. Furthermore, cognitive psychology describes three approaches to the mental representation of natural categories: *classical*, *prototype*, and *exemplar* [12, 14]. These three categories are described below since the distinction between them is critical for the creation of appropriate computational models in data mining.

- **Classical Approach.** In the classical approach, objects are grouped based on their properties. Objects are either a member of a category or not, and all objects have equal membership in a category. A category can be represented by a set of rules, which can be evaluated as true (object belongs to the category) or false (object does not belong to the category).

- **Prototype Approach.** In the prototype approach, the members are more or less typical of the category; in other words, they belong to the category to a certain degree. The prototype of the category usually represents the central tendency of the category and may be defined as the "average" of all the members of the category. Thus, in the prototype approach, a category is based on a "prototype," which exists as an ideal member of a category, and the other members of the same category may share some of the features with the ideal member [1].

- **Exemplar Approach.** In the exemplar approach, all exemplars of a category are stored in memory, and a new instance is classified based on its similarity to all prior exemplars [10]. This representation requires specification of a similarity measure, as well as storage and retrieval of multiple exemplars. In the exemplar-based representation, the category is defined by all exemplars belonging to a given category.

These three approaches to natural concepts can be mapped into Dr. Sadegh-Zadeh's *classical* and *non-classical* concepts. Dr. Sadegh-Zadeh defines the classical concepts as classes "characterized by the 'common nature' of its members, that is, by a number of properties that are common to all of them." Dr. Sadegh-Zadeh refers to this quality of the classical concepts as the "common-to-all postulate." Thus, a concept is categorized as classical "if it denotes a category that obeys the common-to-all postulate." Dr. Sadegh-Zadeh uses the concept of a square as an example of a classical concept. The concept of a square is characterized by four properties: closed figure, four straight sides, all sides equal in length, and equal angles. In the traditional distinction between artificial and natural concepts, the concept of a square is classified as an artificial concept since the properties have defining nature. Accordingly, all members of the class "square" must meet the four conditions. In contrast, a natural concept has characteristic features rather than defining features. Dr. Sadegh-Zadeh uses the concept of disease as an example of a non-classical concept. The concept of disease "does not denote a category whose members obey the common-to-all postulate."

In general, natural concepts (categories) have two fundamental characteristics: (1) The members of a natural category do not have to share all features; a natural category may have some attributes which are common to many members, and some attributes which may be shared by only a few members; and (2) The members of a natural category may not be equally representative for a category; thus, the members may vary in their typicality.

Most medical concepts can be classified as natural concepts rather than classical concepts. Medical concepts reflect the rapidly expanding and evolving nature of medical knowledge. They are characterized by a high level of changeability, context-dependency, and imprecision. Our discussion on the nature of medical concepts has an introductory character, and merely highlights some major issues in the context of the conceptual models in data mining. Our discussion has also a practical nature, and it presents an operational definition of obstructive sleep apnea.

13.3 Data Mining and Modeling of Medical Concepts

Data mining is based on a secondary use of existing medical data. Thus, the data are not purposely collected for data mining, and the meaning of the data should be interpreted in the context of the original task. In most cases, medical data are collected for three distinct reasons: for an individual patient's care, for medical research, and for patient administration. For the first reason, the data acquisition method is driven by the diagnostic, prognostic, or treatment process and the data are successively obtained, stored, and used by the healthcare practitioners. Since the intended use of the data is the patient's care, the data are often incomplete and have varied granularity. For medical research, data are prospectively acquired through purposely designed clinical trials, epidemiological studies, or studies of healthy populations. These data sets are collected by the researchers to answer specific clinical questions and afterwards are analysed using statistical methods for confirmation or negation of a particular medical hypothesis. For patient administration, data are collected and used for accounting and planning [6]. As a result, data mining uses heterogeneous data and multiple meanings. Integration of these diverse data and metadata requires creation of a unified conceptual data model. This task is complicated by the fact that data are collected using diverse collecting methods, inclusion criteria, sampling methods, sample sizes, types and numbers of measurements, and definitions of outcomes. Therefore, conceptual data modeling for data mining must address the problems related to the secondary use of the data, which means that the data were originally collected for other purposes and may have more or less explicitly defined meaning (metadata).

In his *Handbook of Analytic Philosophy of Medicine*, Dr. Sadegh-Zadeh outlines four main classes of medical concepts: individual, qualitative (classificatory), comparative, and quantitative. Individual concepts denote specific individual objects. For example, John Davey denotes specific patient. Qualitative or classificatory concepts are "either unary predicates or non-comparative, many-place predicates." For example, John Davey has excessive daytime sleepiness, EDS. Comparative concepts express relationship between two or more objects. For example, John Davey has higher EDS than John Smith. Quantitative concepts specify the magnitude of an attribute using a numerical function. For example, John Davey's sleepiness can be "measured" using the Epworth Sleepiness Scale (the standard questionnaire used in sleep clinics). Thus, we can say that John Davey has sleepiness of 20, and John Smith has sleepiness of 15. Dr. Sadegh-Zadeh defines quantitative concepts as "a homomorphism f from a particular empirical, i.e., experiential, structure $\langle R, \geq \rangle$ into a numerical structure R such that R is, usually, the set of positive real numbers, and $\geq$ is the 'is greater than or equals' relation." The quantification f is also called measurement and the homomorphism is referred to as a scale. Quantitative concepts are particularly important in medical diagnosis and treatment evaluation. We will discuss them further in context of the operational definition of concept.

Additionally, Dr. Sadegh-Zadeh discusses the notion of an attribute, and describes two types of attributes: categorical and dispositional. Categorical attributes are permanently present, for example, John Smith's weight. The specific value for weight may change with time, but weight is measurable under all circumstances. On the other hand, dispositional attributes may manifest only under certain circumstances, for example, John Smith's daytime sleepiness. Daytime sleepiness can be observed during day (assuming that John Smith is not a nightshift worker).

13.3.1 Operational Definition of a Concept

Operational definition is based on the notions of conceptualization and operationalization. Conceptualization defines the concept in the context of its use, for example, excessive sleepiness in the context of the clinical diagnosis of a specific sleep disorder. Operationalization states that a concept is defined in terms of the operations by which its referent is measured. For example, sleepiness is "measured" by a subjective measure [16]. The Epworth Sleepiness Scale (ESS) is a self-administered questionnaire composed of eight questions to measure the general level of daytime sleepiness in terms of the probability of falling asleep during daily activities: (1) sitting and reading, (2) watching TV, (3) sitting inactive in public place (e.g. a theatre or a meeting), (4) riding as a passenger in a car for an hour without a break, (5) lying down to rest in the afternoon when circumstances permit, (6) sitting and talking to someone, (7) sitting quietly after lunch without alcohol, and (8) sitting in a car, while stopped for a few minutes in traffic. Each item has a score between $0-3$. The answers are never, slight chance, moderate chance, and high chance. The maximum score is 24. Typically a score of 11 and above is recognized as excessive daytime sleepiness (EDS). Thus, the operational definition of daytime sleepiness allows for a mapping of the concept of sleepiness (in particular context) into an interval between 0 and 24. EDS is defined as a score of ESS ≥ 11. These definitions are essential for standardization of meanings and data integration in medical data mining. However, interpretation of the meaning of a particular ESS value requires thorough contextual analysis. We describe the main issues of the contextual interpretation in the next subsection.

13.3.2 Contextual Interpretation of Excessive Daytime Sleepiness

The symptom of sleepiness, although extensively used in screening and diagnosis, is not easy to describe and, moreover, to quantify. Sleepiness can be measured only indirectly – there is not yet a single laboratory test to identify 'sleepy' individuals. However, excessive daytime sleepiness (somnolence) is one of the most important symptoms of *OSA* used for screening, evaluation, and classification of the severity of *OSA* [3]. Sleepiness is a typical complaint of *OSA* patients (or their family members). But sleepiness is not a universal symptom; about 10% patients with *OSA* do not display excessive sleepiness. Moreover, reported sleepiness may be related to many other problems. Figure 13.1 illustrates the complex dependencies between Excessive Daytime Sleepiness (EDS) and shift work, sleep deprivation, circadian rhythm disruption (CRD), depression, insomnia, and presence of other sleep disorders such as periodic limb movement (PLM) or restless legs syndrome (RLS). The arrows represent possible causal relationships between the factors. For example, depression may increase insomnia and, vice versa, insomnia may increase the feeling of depression.

13.3.3 Operationalization of Obstructive Sleep Apnea

In the case of medical concepts, their meanings are a result of many possible interpretations depending on their context, their specific use, and the particular time of use. Therefore, modeling of medical concepts must be based on the following four premises: (1) only some clinical concepts have clearly defined boundaries; most concepts are fuzzy; (2) clinical concepts are used in specific contexts and are subject to various interpretations; (3) clinical concepts are created and used for specific

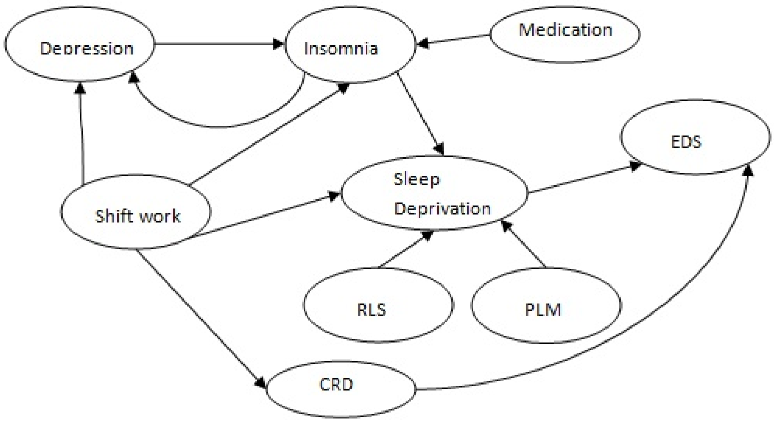

Fig. 13.1 Dependencies between the Excessive Daytime Sleepiness (EDS) and other factors. Daytime sleepiness is closely related to the length of sleep and pattern of sleep/awake hours, and, therefore, cannot be analysed without two additional important factors: duration of sleep and timing of sleep (circadian rhythm disruption).

purposes and must be interpreted in the context of these purposes; and (4) clinical concepts evolve over time.

We describe operationalization in more details using as an example of a particular diagnostic process for obstructive sleep apnea. In our discussion, we use the generic name obstructive sleep apnea (*OSA*) to indicate Obstructive Sleep Apnea Syndrome (*OSAS*) and Obstructive Sleep Apnea/Hypopnea Syndrome (*OSAHS*). *OSA* is a common, serious respiratory disorder afflicting, according to conservative studies, $2-4\%$ of the adult population. The differences in reported *OSA* prevalence values result from different diagnostic methods and varied definitions of *OSA*. “Apnea” means “without breath,” and *OSA* occurs only during sleep, and is, therefore, a condition that might go unnoticed for years. *OSA* is caused by collapse of the soft tissues in the throat as the result of the natural relaxation of muscles during sleep. The soft tissue blocks the air passage and the sleeping person literally stops breathing (apnea event) or experiences a partial obstruction (hypopnea event). The apnea event in adult is defined as at least 10 second breathing pause (complete cessation of air flow) and the hypopnea event is defined as at least 10 second event with reduced air flow by at least 50%.

Although sleep apnea is not a new condition and has been mentioned sporadically in literature (e.g., Charles Dickens provided a description of Joe the fat boy), it was discovered and described only in 1965. Since then, it has been recognized as a serious respiratory disorder. One of the most important daytime symptoms of *OSA* is excessive daytime sleepiness (EDS). One of the important night symptoms is a decrease in the blood oxygen saturation (the percentage of haemoglobin saturated with oxygen) during sleep.

The gold standard for the diagnosis of *OSA* is an overnight in-clinic polysomnography. The sleep study measures the frequency of apnea and hypopnea events. In general, the diagnosis of *OSA* uses two scores: Apnea-Hypopnea Index (*AHI*) and Apnea Index (*AI*). The apnea-hypopnea index (*AHI*), the most commonly used score, is calculated as a number of apnea and hypopnea events per hour of sleep; The apnea index (*AI*) is calculated as a number of apnea events per hour of sleep. Additionally, many definitions of apnea/hypopnea events require one or both of two factors: oxyhemoglobin desaturation of 4% or more and brief arousals from sleep. Thus, the definition of apnea event varies.

The diagnosis of *OSA* can be based on two approaches: (1) a score of apnea/hypopnea events (*AHI*) or (2) a combination of *AHI* scoring and symptoms. In the diagnosis based solely on the *AHI* index, apnea is classified as *mild* for *AHI* between 5 and 14.9, *moderate* for *AHI* between 15 and 29.9, and *severe* for $AHI \geq 30$. The International Classification of Sleep Disorders (ICSD) [16] defines the severity of *OSA* in terms of the frequency of apnea events, the degree of oxygen desaturation, and the severity of daytime sleepiness.

The differences between the scoring and definitions of apnea have important implications for the conceptual data modeling in data mining. Most published medical research studies base the diagnosis of *OSA* on *AHI* or a combination of apnea index (*AI*) and *AHI* obtained from overnight in-clinic PSG. For example, the authors of two articles on craniofacial predictors [9] and [5] define respectively two criteria: (1) *OSA* defined as $AHI \geq 5$ and (2) *OSA* defined as $AI > 5$ or $AHI > 10$. To illustrate the difference between diagnostic criteria based on *AHI* and a combination of *AI* and *AHI*, we applied these two criteria to 233 records from the data set obtained from the authors of the first publication [3]. Figure 13.2 shows the number of records classified into *OSA* and non-*OSA* using two diagnostic criteria: *OSA* defined by $AHI \geq 5$ (column to the left) and *OSA* defined by $AH > 5$ OR $AHI > 10$ (column to the right). The second criterion is more restrictive and excludes 26 records $(26/233)$ from the *OSA* group and classifies them as a non-*OSA*.

The definition of *OSA* is fuzzy. Different studies use different cut-off values to indicate *OSA*, for example $AHI \geq 5$, $AHI \geq 10$, $AHI \geq 15$. To illustrate the differences in prevalence of *OSA*, we applied three cut-off values to 795 records obtained from a sleep clinic [7, 8]. Figure 13.3 shows the changing proportions between non-*OSA* and *OSA* records for $AHI \geq 5$, $AHI \geq 10$, and $AHI \geq 15$.

OSA is operationalized using diverse methods. Thus, the conceptual model for data must define precisely the scoring criteria. The use of *AHI* or *AI* and three cut-off values can result in significant differences in classification of the patients (non-*OSA*, *OSA*), especially for patients with low *AHI* scores [13]. Furthermore, the difficulty with the scoring of *AHI* is compounded by the natural night-to-night variability in the severity of apnea and the differences in diagnostic equipment.

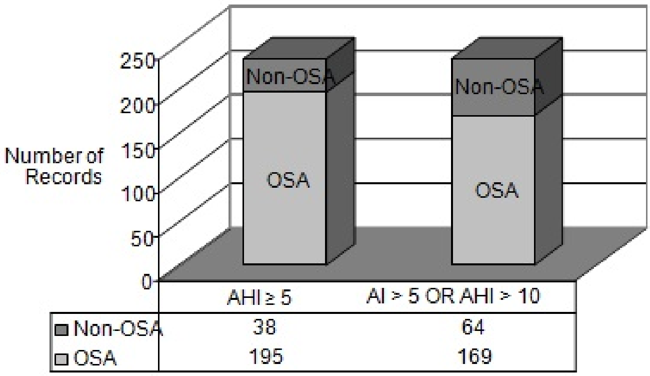

Fig. 13.2 *OSA* prevalence using diagnostic criteria based on two diagnostic criteria

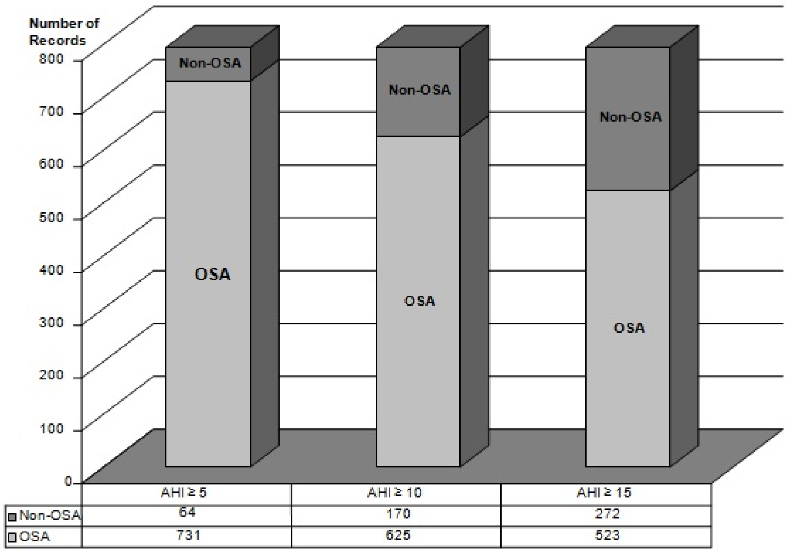

Fig. 13.3 *OSA* prevalence for three levels of *AHI* cut-off values: $AHI \geq 5, 10, 15$

13.4 Semiotic Approach to Concept Modeling

In the case of medical concepts, their meanings are a result of many possible interpretations depending on their context, their specific use, and the particular time of use. Therefore, in most cases, the interpretation is based on the following four premises: (1) only some clinical concepts have clearly defined boundaries; most concepts are fuzzy; (2) clinical concepts are used in specific contexts and are subject to various interpretations; (3) clinical concepts are created and used for specific purposes and must be interpreted in the context of these purposes; and (4) clinical concepts evolve over time. In this section, we describe a semiotic approach to modeling of medical concepts.

Originally, the term ‘semiotics’ (from a Greek word for sign “semaion”) was introduced in the second century by the famous physician and philosopher Galen (129-199), who classified semiotics (the contemporary symptomatology) as a branch of medicine [19]. The use of the term semiotics to describe the study of signs was developed by the Swiss linguist, Ferdinand de Saussure (1857-1913) and the American logician and philosopher Charles Sanders Peirce (1839-1914). Today, semiotics can be broadly defined as the study of signs. Since signs, meaning-making, and representations are all present in every part of human life, the semiotic-based methods have been used in many disciplines, from mathematics through literary studies and library studies to information sciences. A semiotic paradigm is associated with different traditions and with a variety of empirical methodologies. Our intention in this paper is to define the basic terminology needed to present the need for the semiotic approach to the modeling of medical concepts. The examples of the operationalization of *OSA* illustrate that the meaning of a sign arises in its interpretation or even in multiple possible interpretations.

Peirce defined “sign” as any entity carrying some information and being used in a communication process. Peirce, and later Charles Morris, divided semiotics into three categories [2] : syntax (the study of relations between signs), semantics (the study of relations between signs and the referred objects), and pragmatics (the study of relations between the signs and the agents who use the signs to refer to objects in the world). This triadic distinction is represented by Peirce’s semiotic triangle: the representamen (the form which the sign takes), an interpretant (the sense made of the sign), and an object (an object to which the sign refers).

We base our approach to concept modeling on the traditional Piece’s triangle. However, we have used Peirce’s triadic model (object, representamen, interpretant) rather loosely since our emphasis is on the process of sense-making (process of semiosis in Peirce’s theory). Furthermore, we emphasize the communication process and the role of the interpreter in the creation (construction) of meaning. Thus, in semiotic terminology, the meaning of the sign (representamen) arises in its interpretation. Our model uses the notions of conceptualization, operationalization, and utilization. The notions of conceptualization and operationalization have their roots in social sciences. The notion of utilization of measures has been added by us to model the pragmatic aspects. We introduce a triplet for the representation of the concept and use the terms: conceptualization (what to measure), operationalization (how to measure), and utilization (how to use the measure). We observe a strong parallelism between the semiotic triangle (object, representamen, interpretant) and the triplet: conceptualization, operationalization, and utilization. In many ways, the semiotic approach presented here is oversimplified and does not reflect the richness of multiple semiotic traditions. However, we believe that a semiotic approach allows for modeling of complex medical concepts. We focus on the pragmatic aspects (utilization) of the models. Therefore, we have introduced a notion of “pragmatic modifiers," which represent the notion of "interpretant" or utilization. This notion focuses on the role of interpretation (interpreter) in the creation of meaning. We called them modifiers since in a way, they "modify" the meaning. We identified four groups of modifiers (this is not an exhaustive list of possible modifiers): goals (e.g., diagnosis, assessment, treatment evaluation), agents (e.g., patients, health professionals, medical sensors, computer systems), perspectives (e.g., health care costs, accessibility, ethics), and views (e.g., variations in the diagnostic criteria used by different experts or clinics). The names of the modifiers are used here in an arbitrary fashion, and they are introduced here to indicate various sources of modifications. For example, the important modifiers are health-care costs and accessibility of specialists.

The modeling of the concept of *OSA* in the context of various goals involves three aspects: conceptualization (what to measure), operationalization (how to measure), and utilization (how the measure is applied). We mapped these aspects using the semiotic triangle, shown in Figure 13.4.

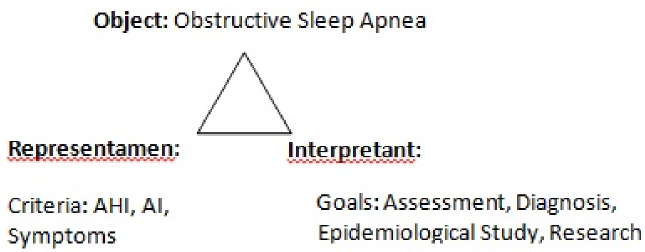

Fig. 13.4 Peircean semiotic triangle for contextual modeling of *OSA*

The multiplicity of diagnostic criteria and ongoing modifications clearly indicate the need for a flexible computational representation, which must model the various criteria and approaches to the diagnosis, assessment, and treatment of *OSA* (and other sleep disorders).

13.5 Conclusions and Future Work

In this paper, we have discussed the typology of medical concepts introduced by Dr. Kazem Sadegh-Zadeh in his *Handbook of Analytic Philosophy of Medicine*. Specifically, we have concentrated on three issues: non-classical concepts, qualitative concepts, and operational definition of concepts. We have used an example of obstructive sleep apnea to demonstrate that (1) medical concepts are created and used for specific purposes and must be interpreted in the context of these purposes; (2) medical concepts rarely have clearly defined boundaries; most concepts are fuzzy; (3) clinical concepts are used in specific contexts and are subject to various interpretations; and (4) clinical concepts evolve over time. To address these issues, we have used a semiotic approach to modeling of medical concepts. Semiotics provides the modeling constructs for the description of the concept, its representation, interpretation, and utilization. Furthermore, we have observed that (1) data mining requires explicit conceptual models based on operational definitions of medical concepts; and (2) the quality of the data-mining process depends on the quality of the conceptual data models and the quality of the data.

We are planning to further expand the proposed framework and to build a comprehensive computational model for the medical concept of obstructive sleep apnea and its symptoms. We will apply this model in a clinical decision support system for the diagnosis and treatment of *OSA*, as well as in a support system for the treatment of sleep disorders. Furthermore, we plan to utilize the proposed computational model for the analysis of patients' data from clinics which use diverse diagnostic criteria. The explicit modeling will allow us to compare treatment results from various clinics.

References

1. Bruner, J.S., Goodnow, J.J., Austin, G.A.: A Study of Thinking. Wiley, New York (1956)
2. Chandler, D.: Semiotics: The Basics. Routledge, London (2002)
3. Curcio, G., Casagrande, M., Bertini, M.: Sleepiness: Evaluating and Quantifying Methods. International Journal of Psychophysiology 41, 251–263 (2001)
4. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery: An Overview. In: Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (eds.) Advances in Knowledge Discovery and Data Mining, pp. 37–54. The MIT Press, Menlo Park (1996)
5. Ferguson, K.A., Ono, T., Lowe, A.A., Ryan, C.F., Fleetham, J.A.: The Relationship Between Obesity and Craniofacial Structure in Obstructive Sleep Apnea. Chest 108(2), 375–381 (1995)
6. Goble, C.A., Glowinski, A.J., Nowlan, W.A., Rector, A.L.: A Descriptive Semantic Formalism for Medicine. In: Proceedings of the Ninth International Conference on Data Engineering, pp. 624–631 (1993)
7. Kwiatkowska, M., Ayas, N.T., Ryan, C.F.: Evaluation of Clinical Prediction Rules Using a Convergence of Knowledge-driven and Data-driven Methods: A Semio-fuzzy Approach. In: Zanasi, A., Brebbia, C.A., Ebecken, N.F.F. (eds.) Data Mining VI: Data Mining, Text Mining and their Business Applications, pp. 411–420. WIT Press, Southampton (2005)
8. Kwiatkowska, M., Atkins, M.S., Rollans, S., Ryan, C.F., Ayas, N.T.: Decision Tree Induction in the Creation of Prediction Models for Obstructive Sleep Apnea (*OSA*): A Pilot Study (Abstract). In: International Conference of American Thoracic Society, San Diego (2006)
9. Lam, B., Ip, M.S.M., Tench, E., Frank Ryan, C.: Craniofacial Profile in Asian and white Subjects with Obstructive Sleep Apnea. Thorax 60(6), 504–510 (2005)
10. Medin, D.L., Marguerite, M., Medin, D.L., Schaffer, M.M.: Context Theory of Classification Learning. Psychological Review 85, 207–238 (1978)
11. Minda, J.P., David Smith, J.: The Effects of Category Size, Category Structure and Stimulus Complexity. Journal of Experimental Psychology: Learning, Memory and Cognition 27, 755–799 (2001)

12. Nosofsky, R.M.: Exemplars, Prototypes, and Similarity Rules. In: Healy, A.F., Kosslyn, S.M., Shiffrin, R.M. (eds.) From Learning Theory to Connectionist Theory: Essays in Honour of William K. Estes, vol. 1. Erlbaum, Hillsdale (1992)
13. Redline, S., Kapur, V.K., Sanders, M.H., Quan, S.F., Gottlieb, D.J., Rapoport, D.M., Bonekat, W.H., Smith, P.L., Kiley, J.P., Iber, C.: Effects of Varying Approaches for Identifying Respiratory Disturbances on Sleep Apnea Assessment. American Journal of Respiratory and Critical Care Medicine 161, 369–374 (2000)
14. Reed, S.K.: Cognition. Theory and Applications, 4th edn. Brooks/Cole Publishing, Pacific Grove (1996)
15. Rosch, E., Mervis, C.B.: Family Resemblances: Studies in the Internal Structure of Categories. Cognitive Psychology 7, 573–605 (1975)
16. Rosenberg, R.S., Mickelson, S.A.: Obstructive Sleep Apnea: Evaluation by History and Polysomnography. In: Fairbanks, D.N.F., Mickelson, S.A., Tucker Woodson, B. (eds.) Snoring and Obstructive Sleep Apnea, 3rd edn. Lippincott Williams & Wilkins, Philadelphia (2003)
17. Sadegh-Zadeh, K.: Handbook of Analytic Philosophy of Medicine. Springer, Dordrecht (2012)
18. Sebeok, T.A., Danesi, M.: The Forms of Meaning: Modeling Systems Theory and Semiotic Analysis. Mounton de Gruyter, Berlin (2000)
19. Sebeok, T.A.: Signs: An Introduction to Semiotics. University of Toronto Press, Toronto (1999)