Wojtusiak J. (2014) Rule Learning in Healthcare and Health Services Research. In: Dua S., Acharya U., Dua P. (eds) Machine Learning in Healthcare Informatics. Intelligent Systems Reference Library, vol 56. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40017-9_7

Chapter 8 Rule Learning in Healthcare and Health Services Research

Janusz Wojtusiak

Machine Learning and Inference Laboratory
Department of Health Administration and Policy
George Mason University

Abstract. Successful application of machine learning in healthcare requires accuracy, transparency, acceptability, ability to deal with complex data, ability to deal with background knowledge, efficiency, and exportability. Rule learning is known to satisfy the above criteria. This chapter introduces rule learning in healthcare, presents very expressive attributional rules, briefly describes the AQ21 rule learning system, and discusses three application areas in healthcare and health services research.

Keywords. Rule learning, attributional calculus, AQ21 system, health services research, aggregated data, healthcare billing data

8.1 Introduction

Healthcare requires modern computational tools to handle the complexity of data and workflows. The healthcare environment is dynamic and frequently changing: New knowledge is published on a daily basis, new drugs are constantly available, and the best practice guidelines change. Moreover, healthcare is a critical area in which success is measured by patient survival and wellbeing. Unfortunately, many existing treatment and reimbursement systems used in healthcare treat individual patients as “average” cases without tailoring to patient characteristics.

The above reasons call for machine learning methods to manage the complexity and automatically adapt to frequent changes. This chapter focuses on one of the best known and most important methods in machine learning in healthcare: rule learning. It briefly describes rule learning methods, discusses their use in healthcare delivery, research, administration and management, and presents advantages of using rule learning rather than traditional computational approaches and other machine learning methods.

In order to fully justify the use of rule learning in healthcare, the following sections briefly outline aspects of machine learning that are particularly important in this application area.

8.1.1 What is Needed in Healthcare and Health Services Research?

Machine learning methods have a wide range of applications in healthcare delivery, research, administration, and management. Many of these applications are slowly emerging as the healthcare community becomes more familiar with machine learning and its immense potential. On the other hand, most machine learning researchers are not familiar with healthcare settings and over-trivialize them. This mutual lack of understanding between healthcare and machine learning communities results in the lack of advanced machine learning methods adoption.

Among the healthcare areas that benefit the most from machine learning are those that rely on automated processes or that can be automated. The ability of machine learning methods to adapt to dynamically changing environments, previously unseen situations, and new challenges make them ideal for these types of applications. Two of the most common applications of machine learning in healthcare are: decision support systems and knowledge discovery. Decision support systems rely on computational models that aid decision makers in a variety of situations. These models can be constructed and maintained using machine learning. In addition, knowledge discovery, which primarily derives from medical datasets, can be used to study patterns of healthcare delivery systems, management, billing, etc. Machine learning has, thus, great potential when correctly applied to hard problems that cannot be solved with more traditional computational methods or manually without the use of computers.

However, for machine learning to be adopted in healthcare, methods need to fulfill several requirements. These requirements are eminent and applicable to virtually all domains in which machine learning is or can be used. However, some of these requirements are particularly important in healthcare when the adoption of new technologies and results are exceptionally challenging.

- **Accuracy.** Models have to provide reliable predictions and/or reliably describe data, which is, in most cases, their main function. Multiple measures of accuracy are available, all of which perform some form of counting/scoring of correct and incorrect predictions and combinations thereof. Some commonly used measures of accuracy include precision, recall, sensitivity, specificity, F-score, and others.

- **Transparency.** Medical and healthcare studies require models to be easily understood by people not trained in machine learning, statistics, and other advanced data analysis methods. In this sense, providing just the reliable predictions is not sufficient, as models should also "explain" why a specific prediction is made and what the model actually does. This corresponds not only to methods that lead to creation of new knowledge, but also to autonomous systems that because of their critical role need to leave an "audit trail" and be analyzed/verified periodically.

  The concept of understandability and interpretability has been well known since early work on expert systems and artificial intelligence, but has been largely ignored by many modern machine learning methods. One reason for this is that it is very hard to measure the complexity of created models and hypotheses, and use that measurement as one of knowledge representation selection criteria. It is virtually impossible to consistently measure and compare the transparency of models learned in different representations. (How do we compare transparency of specific SVM-based, NN-based, and rule-based models for diagnosing liver diseases? How do we generalize the measure?) Moreover, compound knowledge representations, which are natural to people, tend to be difficult to learn through machine learning methods. One such representation, called attributional calculus consists of attributional rules, which are briefly outlined in Section 8.3.1.

- **Acceptability.** Models need to be accepted by their potential users. While partially related to transparency, acceptability requires that the models that do not contradict the knowledge of existing experts are otherwise "reasonably" congruent with what is currently being done, and correspond to existing workflows. Acceptability is a key issue in healthcare, more than in any other industry. Clinicians, administrators, and supporting staff do not want to change the way they work, even if the developed models being used are accurate and superior to methods currently being used. The use of ML algorithms should immediately lead to improved work and provide incentives to participants; otherwise results may not be adopted.

- **Ability to handle complex types of data.** Healthcare data are complex. Even relatively simple applications of machine learning to healthcare data require making numerous conversions, data pre-processing, encoding of variables, and so on. In order to have widespread acceptance in healthcare, machine learning methods should be able to operate directly with healthcare data without the need to artificially encode. Healthcare data are not, and should not be, treated by ML tools as a collection of numbers without meaning. Although more advanced ML methods recognize a wide range of data types (nominal, structured, ordinal, interval, ratio, absolute, compound, etc.), prevalent standards such as ICD-9, ICD-10, CPT, SNOMED, and HL7 are currently not directly supported by ML tools.

- **Ability to handle background knowledge.** Computers require massive amounts of data to make simple decisions or discover simple facts. Humans do exactly the opposite -- we are able to make important decisions and discover important facts based on minimal information. Although there are many differences in human and computer inference/learning processes, one of the most important is the ability to use background knowledge to place problems into the appropriate context. Similarly, machine learning algorithms that are provided with large knowledge bases and a wealth of background knowledge need not have access to huge amounts of data. This allows machine learning algorithms to focus on the discovery of novel facts and not what is already known to experts. Extremely large repositories of medical and healthcare knowledge (is often not coded and in many cases only available as text of published manuscripts) can be incorporated into the machine learning process.

- **Efficiency.** Both model induction and model application algorithms need to be efficient. Machine learning algorithms applied in healthcare should be able to cope with very large amounts of data. The data may have many examples (sometimes called records or datapoints), attributes (sometimes called variables or features), or both. The theoretical estimates of algorithm complexity are often available for many methods. More importantly users want the methods to be executed in a specific period of time, even if it means that results are only approximate or "good enough."

- **Exportability.** Results of machine learning should be directly transferable to decision support and other systems where they can be immediately applied. It is not unusual that the learned models will work along with already existing models and thus need to be compatible. For example, learned models can be translated or directly learned in the form of rules in Arden Syntax, a popular representation language in clinical decision support systems. If models are learned in completely different representations, they need to be translated (usually approximately) to the target form.

This chapter focuses on the use of rules and rule learning methods in different healthcare areas. Rules are known to be one of the most transparent knowledge representations that also conform to other criteria outlined above.

8.2 Rule Learning

Over the past few decades multiple rule learning algorithms and software have been developed. Multiple types of rules are considered in machine learning research depending on their use and form, including: association rules (which are used to represent regularities in data), decision rules (which are used to support decisions) and their subtype classification rules (used to classify examples into concepts), rules with exceptions (that include part describing when the rule does not apply), *m*-of-*n* rules (used to count true values or statements), and attributional rules (the most expressive form of rules considered here).

The AQ21 system is particularly suitable for problematic healthcare situations because of its flexibility, ability to deal with multiple types of attributes, handle both large and small datasets, use background knowledge in different forms, learn from individual and aggregated data, manage meta-values, cope with noise, perform constructive induction, generate alternative hypotheses, and many other features. AQ21 uses attributional rules as the main form of knowledge representation. The following subsections briefly introduce attributional rules, and outlines AQ21 main algorithms.

8.2.1 Attributional Rules

Healthcare applications require rules that are more expressive than typically used

$$\textit{CLASS IF CONDITION} \tag{8.1}$$

Most software creates rules in which *CONDITION* is a conjunction of simple conditions in the form *ATTRIBUTE* = *VALUE*. Many such rules are needed to describe even simple concepts. Attributional rules are currently the most expressive form of rules induced by machine learning algorithms. They are the main knowledge representation in a formal language called *attributional calculus*, AC [9]. AC has been created to support *natural induction*, an inductive learning process which has results that are natural to people because of their form and content.

Natural induction requires that knowledge be equivalent to statements in natural language (i.e. English), so those who are not experts in machine learning or knowledge mining, or do not have a technical background may understand it. Thus, medical doctors, healthcare administrators, nurses, and researchers should be able to understand, interpret, modify, and apply knowledge learned by computer systems. Such a goal requires that knowledge discovery programs use a language that can either be automatically translated to natural language or easily understood on its own.

Learned knowledge is represented in attributional calculus in the form of *attributional rules,* which consist of *attributional conditions*. An attributional condition takes the form:

$$[L\ rel\ R{:}A], \tag{8.2}$$

where L is an attribute, an internal conjunction or disjunction of attributes, a compound attribute, a counting attribute, or an expression. *rel* is one of =, >, <, ≤, ≥, :, or ≠. R is an attribute value, an internal disjunction of attribute values, an attribute, an internal conjunction of values of attributes that are constituents of a compound attribute, or an expression. A is an optional annotation that may list statistical information describing the condition. The annotation often includes $|p|$ and $|n|$ values for the condition, defined as the numbers of positive and negative examples, respectively, that satisfy the condition, and the condition's consistency defined as $|p|/(|p|+|n|)$.

There are several forms of attributional rules allowed by attributional calculus. Three important forms of attributional rules are presented below:

$$CONSEQUENT <== PREMISE \tag{8.3}$$

$$CONSEQUENT <== PREMISE\ \llcorner\ EXCEPTION \tag{8.4}$$

$$CONSEQUENT <== PREMISE\ \ulcorner\ PRECONDITION \tag{8.5}$$

where *PREMISE*, *CONSEQUENT*, *EXCEPTION*, and *PRECONDITION* are complexes, that is, conjunctions of attributional conditions. An *EXCEPTION* can also be an explicit list of examples that constitute exceptions to the rule. The rules without exception or preconditions are interpreted as the *CONSEQUENT* is true whenever the PREMISE is true. The rules with exceptions are interpreted that the *CONSEQUENT* is true whenever the PREMISE is true, except for when the *EXCEPTION* is true. The rules with preconditions are interpreted that the *CONSEQUENT* is true whenever the *PREMISE* is true, provided that the *PRECONDITION* is true. The symbols $\llcorner$ and $\ulcorner$ are used to denote exception and precondition, respectively. Each rule may be optionally annotated with several parameters such as numbers of covered examples (positive and negative), the rule complexity, etc.

One class of the data is usually described using several rules, called a *ruleset*. Rules considered here are independent, i.e., the truth status of one rule does not affect interpretation of other rules. This is in contrast to many other rule learning programs that learn sequential rules that need to be evaluated in a specific order. A set of rulesets that describe all considered classes in the data (often defined by possible values of an output/dependent attribute) is called a *ruleset family*, a.k.a. classifier. Depending on the problem at hand, the goal may be learn a complete classifier, a ruleset for one class of interest, or individual rules representing regularities/patterns in the data.

Table 8.1 Table with example conditions and rules

[Length>7.3]
The length of an entity is greater than 7.3 units (as defined in the attribute's domain).
[Color=red ∨ blue: 40,2]
The color of an entity is red or blue. The condition is satisfied by forty positive and two negative examples.
[Length & Height≤12]
An entity's length and height are both smaller or equal to 12 units. The units are defined in the attributes' domains.
[Weather: sunny & windy]
The weather is sunny and windy. This is an example of a condition that

includes a compound attribute Weather.

[Part=acceptable] <== [Width=7..12] &[Length<3] &[Material=steel ∨ plastic]

A part is acceptable if its width is between 7 and 12, its length is less than 3 and its material is steel or plastic.

[Activity=play] <== [Condition=cloudy ∨ sunny: 7,8] & [Temp= medium ∨ high]

⎣ [Condition=cloudy] & [Wind=yes] & [Temp= high] : p=7,n=0,q=1

An activity is play if the condition is cloudy or sunny and temperature is medium or high, except for when the condition is cloudy, there is wind and temperature is high. The rule covers 7 positive and no negative examples. Its quality of the rule is 1.

8.2.2 AQ21

The well-known family of AQ programs originated with the simple version of the A^q algorithm for solving the general covering problem used at the core of rule learning [5]. Numerous implementations and extensions of the method were developed over the years. Among the best known AQ implementations are AQ7 [6], AQ11 [7], AQ15c [11], AQ17 [1], AQ19 [8], and most recently AQ21 [12][13][14].

The AQ21 system consists of two main modules for learning attributional rules, and for their application (Figure 8.1). The learning module consists of data and background knowledge, a pre-processing module, a rule generation module, and a post-processing module. Similarly, the testing module consists of a pre-processing module which converts data and rules to common representation, a rule application module which matches examples against rules, and a post-processing which calculates summaries and statistics.

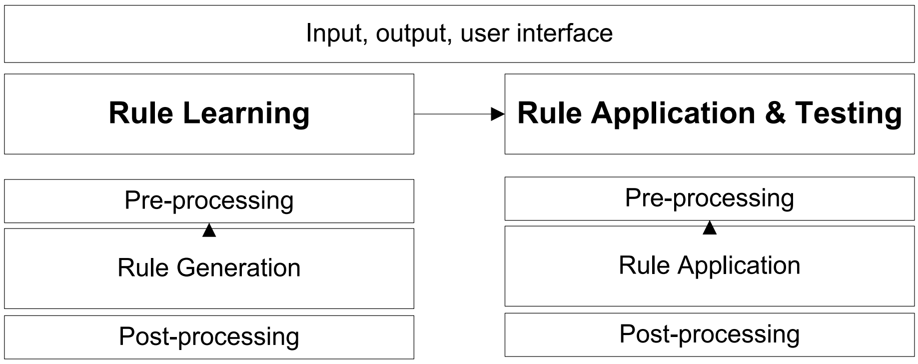

Fig. 8.1 AQ21 system architecture

Rule learning starts with the pre-processing of data and background knowledge which both need to be converted into the right representation and then prepared for rule generation. The process may involve simple steps such as encoding of attribute values, and/or more complex ones including constructive induction. The goal for the latter is to automatically determine the representation space (a set of attributes, their types, and domains). This method is best suitable for the learning problem at hand. AQ21 implements two of three known classes of constructive induction (data-driven (DCI) [3], knowledge-driven (KCI), hypothesis-driven (HCI) [10], and multi-strategy [2]), DCI and KCI. The methods include operators such as attribute selection, attribute generation, and attribute modification.

At the core of the AQ learning is its rule generation module. The method pioneered the separate-and-conquer approach to rule learning, in which data representing a target class being learned are sequentially covered in a way that avoids negative examples. The AQ21 rule generation module starts by focusing on a single example and generates possible generalizations of that example that are consistent or partially consistent with the data and background knowledge. This process, called *star generation*, results in a rule or set of rules that describe part of the data. Multiple stars may be generated in parallel, in order to prevent erroneous generalizations due to noise in the data. The process of star generation is repeated until all data or a significant portion of data are covered (explained) by generated rules. The quality of rules in AQ21 is evaluated using lexicographical evaluation functional (LEF), a method which sequentially evaluates rules through multiple criteria. Numerous variants of the AQ rule generation algorithm have been investigated over the years and are widely described in literature.

The rule post-processing method includes rule optimization, selection of the final rules to be used in a hypothesis or a set of alternative hypotheses, and calculation of statistical parameters describing these rules. The final rules are presented to the user or transferred to the testing and application module.

The rule testing and application module starts with the pre-processing of hypotheses and examples in order to match their representation and prepare for the actual application process. Each considered example (application case) is evaluated against rules. In the case of application of rules in decision support, only one example is usually considered. Rules can be evaluated strictly (when an example either matches a rule or not) and flexibly (when a degree of match, DM, ranging from zero to one is calculated). Multiple schemas [9] are available on how to flexibly evaluate individual conditions, rules, and entire rule sets.

Unlike most classifiers that always give one definitive answer, the AQ21 application module may either provide multiple possible answers, or simply answer “don’t know.” In this philosophy, it is better to provide users with more than one plausible answer with high confidence, or not answer at all, than give a likely incorrect definitive answer.

8.3 From Rule Learning to Decision Support

Decision support systems are broadly defined as computer systems that aid decision makers. This definition can include everything from simple spreadsheet applications, through simulation models, to rule-based expert systems. In this chapter, we focus on knowledge-based decision support systems in which computers provide support to their users based on the content of their knowledge bases.

Traditionally, decision support systems are static in the sense that their knowledge does not change over time without explicit intervention by the user. Machine learning-based decision support systems can, however, evolve and adapt to dynamically changing environments in which they operate. Adaptability is, thus, one of two important areas in which machine learning can help in decision support.

Consider an alert system which provides clinicians with messages informing them about important events related to a specific patient, i.e., allergies, drug-drug interactions, abnormal results. An oversensitive alert system that displays too many messages causes a well-known phenomenon called alert fatigue. In such a case, physicians no longer read alerts, but rather ignore all of them. A typical approach to the problem is to create a system-wide policy/threshold so that alerts do not overwhelm users. This one-size-fits-all approach ignores all the differences between physicians and the way they practice. A machine learning-based solution is able to adapt to specific users (physicians) and show only alerts that have the lowest chance of not being overwritten.

The second important area in which machine learning can be used in healthcare is knowledge generation. The majority of decision support systems are based on rules. These rules, sometimes called Medical Logic Modules (MLMs), are prepared by panels of experts based on the best practice and known evidence. Their creation is a long and difficult process. One of the important applications of machine learning is knowledge generation – the knowledge if present in the right forms can help in preparation of MLMs.

Because rules created by the AQ21 system are independent (i.e. unordered), they can be easily incorporated into decision support systems. For example, attributional rules described above can be directly written in ARDEN syntax [4]. The actual rules are written in the "logic" slot of MLMs while the "data" slot is used to derive attribute values and translate then into the required format. Because one MLM corresponds to a complete decision, it includes multiple rules forming a complete ruleset family. Attributional rules can be also manually inspected by experts and modified as rules and compliance requirements change.

8.4 Review of Selected Applications

This section describes three recent studies that applied rule learning in diverse areas of healthcare. They span over medical, comparative effectiveness, and managerial datasets.

8.4.1 Hospital Bills Classification

The purpose of the described study is to improve billing by advancing healthcare provider operations and performance through the use of machine learning methods [16]. Across the country, healthcare providers are experiencing ongoing pressure from declining revenues. Payers are under increasing pressure to contain costs. The implementation of healthcare reform through the Patient Protection and Affordable Care Act (Public Law 111-148) will further exacerbate this issue. These and additional demands to combat waste, fraud, and abuse are creating mounting pressures to achieve 'perfection' in all phases of healthcare billing and reimbursement authorization for hospitals and independent healthcare providers (e.g. physicians and medical group practices). In order to ensure that payments are appropriate, payers must ascertain that there is proper documentation of care prior to reimbursement. Providers must be diligent in maintaining proper documentation to receive the correct payment and avoid loss of revenue.

The opposing pressures from payers and providers call for the use of decision support/screening methods, to better manage the billing and revenue cycle and detect inconsistencies in coverage, care/service documentation and payments, and to guide financial and clinical personnel through this process. Specifically, we are using machine learning to create models for screening billing information for inconsistency. The initial, proof-of-concept, study presented here is based on the batch processing of obstetrics data collected from a one year period in 2008.

In the first step, the data are pre-processed to match requirements of the machine learning application used. Data available in multiple tables in the hospital information system need to be converted into a flat file. Additional processing of variables needs to be done. In the second step, the AQ21 machine learning system [13], which creates predictive models in the form of highly transparent attributional rules, is used. In order to apply the method to create models, the data is classified as "normal payment" and "abnormal payment" which correspond to payments consistent and not consistent with contractual agreements, respectively. Finally, after the rule learning phase, the models are used to predict whether a specific bill is likely to receive normal payment in advance to its submission to the payer.

Initial application of the method in analyzing billing information for obstetrics patients covered by Medicaid achieved promising results. The presented method provides two strong benefits in analyzing billing information. First, the use of machine learning allows one to automatically create models for predicting bill payments before their submission. The models allow screening of billing information before the bill is sent to payees, therefore maximizing the chance of receiving full payments, and reducing unnecessary denials. Second, the use of highly transparent representations of models in the form of attributional rules, allows for the detection of regularities in bill denials which may lead to potential workflow improvement.

8.4.2 Comparative Effectiveness Research

The gold standard for biomedical research is randomized clinical trials (RCT). In many cases, RCTs are impossible or unethical to perform, and only secondary analysis of existing data form clinical records is possible. Rule learning is an attractive approach to comparative effectiveness research of alternative treatments or medications. The latter are often prescribed based on trial and error.

The problem considered in comparative effectiveness research is substantially different from one considered in typical concept learning in which examples are labeled with classes. Here, the data are in the form of rows including C_i, T_i, and O_i, where C_i are the i_{th} patient case characteristics, T_i is the applied treatment or combination of treatments, and O_i indicates outcomes [15]. Models are created and tested using the following three steps, also illustrated in Figure 8.2.

1) For each treatment or combination of treatments, T, select P_T cases from the database for which therapy T was successful and N_T cases for which therapy T was unsuccessful.

2) Apply rule learning to induce general models, M_T, based on P_T and N_T to predict whether therapy T will be successful given a patient's characteristics. A collection of such models for all considered combinations of treatments will be the final model M. Similarly, create models M_{NT} to predict that a given therapy will not be successful. The reason for creating both positive and negative models is that using both models allows for better control of the level of generalization, and thus increases the confidence in the final models.

3) Given a set of patient characteristics $<c_1, \ldots c_k>$, model M will return a set of possible combinations of treatments $\{T_1, \ldots T_n\}$ that are likely to be successful, $M(<c_1, \ldots c_k>) = \{T_1, \ldots T_n\}$. It is possible that for a given case more than one combination of treatments is returned, i.e. $n>1$, or no considered combination of treatments is returned, i.e. $n=0$. Similarly, models M_{NT} are applied, to create a list of potentially improper combinations of treatments.

4) Test model M on a subset of "unused" data consisting of P "successful" cases and N "unsuccessful" cases. Results of the testing are reported in terms of specificity, selectivity, and statistical significance of individual models and all models together.

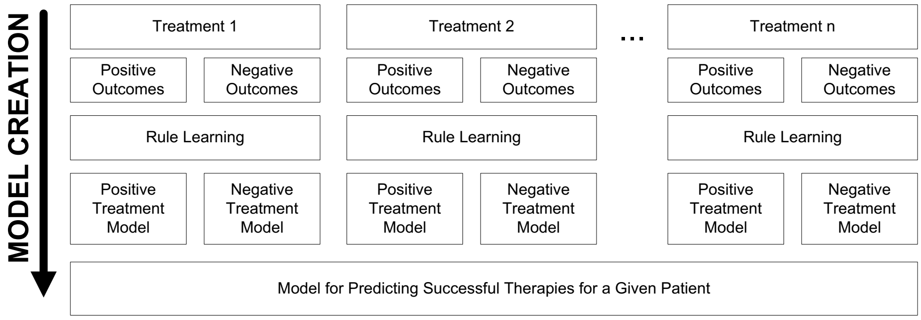

Fig. 8.2 Creation of models for comparative effectiveness research

The created models define groups (or clusters) of patient characteristics that are likely to have positive or negative outcomes. Note that the groups may be intersecting i.e., more than one combination of treatments may appropriate in a specific case, and not exhaustive, i.e., there may be cases for which none of the examined combinations of treatments is predicted to be successful. In the latter case, a *flexible interpretation* of rules may be used to select the closest potentially successful combination of therapies. Within groups of patients selected by machine learning, traditional comparative effectiveness can be performed.

8.1.1 Aggregated Data

There is a growing need to combine data originated from multiple clinical studies. A majority of published studies describe relatively small cohorts and produce platform-dependent results that often lack consistency. Individual measurements of the clinical parameters are protected by The Health Insurance Portability and Accountability Act (HIPAA), thus precluding a combination of multiple cohorts into the large database to perform secondary analyses. A combination of multiple studies, which is the goal of systematic reviews, relies on meta-analysis methods to statistically combine results of the studies. Traditional meta-analysis, however, does not perform knowledge discovery or build predictive/classification models from aggregated data [14].

The problem addressed here is how to learn rules from aggregated data published from multiple studies, rather than from individual examples (subjects). The goal of the method is to discover a model *M* for diagnosing diseases *D*, from published results in which data satisfy a set of criteria *C*. One important characteristic of the method is that the studies do not need to describe diagnostic methods for diseases *D*, but to only include relevant data summaries. Common inclusion criteria that are prerequisites for the traditional meta-analysis methods are not required either. It is sufficient that the criteria are disclosed, so they can serve as inputs to the model along with the aggregated data. The process of the model development is depicted in Figure 8.3.

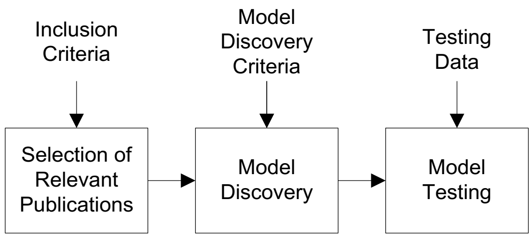

Fig. 8.3. Steps in rule learning from published aggregated data

The rule learning problem considered here induces a rule-based classifier $M(X) \rightarrow D$ that can be used to diagnose *X* patients into diseases from *D*. The model is induced using aggregated data describing groups of patients, not individual datapoints as typically handled by machine learning algorithms. Specifically, the method uses aggregated data *A*, inclusion criteria *C*, and other groups' information *G* to create model *M*. This process extends learning from aggregated data that deals with multiple cohorts of patients described as mean +/- standard deviation of each clinical parameter.

The method has been applied to deriving diagnostic models for metabolic-syndrome related liver complications from summarized (aggregated) descriptions of the small cohorts of patients available from published manuscripts. The significance of this topic is large because approximately 47 million people in the United States have metabolic syndrome (MS) and this number is on the rise. The aggregated clinical data were retrieved from articles published in leading peer-reviewed journals. By applying the developed rule learning methodology, we arrived at several different possible rulesets (sets of rules that together form a model to make a specific diagnosis) that can be used to predict three considered complications of MS, namely non-alcoholic fatty liver disease (NAFLD), simple steatosis (SS), and nonalcoholic steatohepatitis (NASH). It should be noted that the NAFLD group comprises both SS and NASH cases, which means that values of the output attribute form a hierarchy.

Seven NAFLD or NASH predicting rulesets were generated using the AQ21 system executed with different parameters. Resultant rulesets predicting NAFLD or NASH were blindly validated using a well- defined NAFLD database containing 489 patients with biopsy-proven NAFLD, NASH or SS with extensive clinical and laboratory data.

An example of typical automatically learned rule states that patients with BMI>26.85 are likely to have NAFLD, except for when AST is at most 27.2 and adiponectin level are at least 7.25 [14]. The rule is formally shown as:

$$\begin{aligned}[\text{Group=NAFLD}] \quad \text{<==} \quad & [\text{BMI} > 26.85] \\ & \lfloor \; [\text{AST} \le 27.2] \;\&\; [\text{Adiponectin} \ge 7.25]. \end{aligned} \tag{8.6}$$

Validation of this rule for predicting NAFLD resulted in a positive predictive value (PPV) of 85-87%, reflecting relatively high “rule-in” characteristic of the algorithm. The best rule for the prediction of NASH relied on combination of fasting insulin, HOMA and adiponectin values with an accuracy of 78%, with PPV of 71% and negative predictive value (NPV) of 37%.

8.5 Summary

This chapter briefly presented rule learning and its uses in healthcare and health services research. The focus of this paper was on the AQ21 rule learning and testing system because of the system’s applicability to healthcare problems. AQ21 can be viewed more like a laboratory for experimentation with healthcare data rather that a single computer program, which can be executed on data and produce rules. Rule learning performed by AQ21 is particularly suitable for healthcare applications because its high transparency increases the chance that models will be accepted by users.

Acceptability of machine learning methods is a central criterion among those listed in Section 8.2. Other criteria (accuracy, transparency, etc.) lead to the acceptability of models, which in healthcare community is very hard to achieve. While other types of models, such as decision trees and Bayesian networks, are known to be highly transparent, attributional rules follow most of the criteria listed in Section 8.2.

Among the numerous current and potential applications of rule learning in healthcare and health services research, three diverse applications were briefly presented in this chapter. Each application demonstrates that rule learning has great potential and can give good results. The application of rule learning is, however, always straight-forward, and significant work and preparations need to be done before rule learning can be effectively/efficiently used.

Future work on rule learning should focus on four directions. (1) Richer and more natural (to people) rule-based knowledge representations can be created by extending attributional calculus to capture concepts that are natural to healthcare practitioners and researchers. (2) Easy to use tools that deal directly with healthcare data can be developed. One attempt to make computational intelligence and machine learning (CIML) tools accessible to the healthcare community was through CIML Virtual Organization [17]. The VO's goal is to provide the healthcare community with access to CIML tools, advice, educational materials, and networking. (3) Efficiency of rule learning methods can be improved. High complexity or rule based representations require long computation times, particularly when advanced methods, such as constructive induction, are used. (4) Machine learning, in particular rule learning, can be popularized as an attractive approach to data analysis and systems' adaptability, to healthcare community.

References

1. Bloedorn, E.; Wnek, J.; Michalski, R. S.; Kaufman, K. AQ17 A Multistrategy Learning System The Method and Users Guide. Reports of the Machine Learning and Inference Laboratory : MLI 93-12. School of Information Technology and Engineering. George Mason University, Fairfax (1993).
2. Bloedorn, E. Multistrategy Constructive Induction. Ph.D. Dissertation, Reports of the Machine Learning and Inference Laboratory : MLI 96-7. School of Information Technology and Engineering. George Mason University, Fairfax (1996).
3. Bloedorn, E.; Michalski, R. S. Data-Driven Constructive Induction. IEEE Intelligent Systems : Special Issue on Feature Transformation and Subset Selection, pp. 30-37. IEEE Press, (1998).
4. Hripcsak, G. Writing Arden Syntax medical logic modules. Computers in Biology and Medicine, iss. 24, vol. 5, pp. 331-363. (1994).
5. Michalski, R. S. On the Quasi-Minimal Solution of the General Covering Problem. Proceedings of the V International Symposium on Information Processing (FCIP 69)(Switching Circuits), Vol. A3, pp. 125-128. Yugoslavia, Bled (1969).
6. Michalski, R. S; Larson, J. AQVAL/1 (AQ7) User's Guide and Program Description. Department of Computer Science. University of Illinois, Urbana (1975).
7. Michalski, R. S.; Larson, J. Incremental Generation of VL1 Hypotheses: The Underlying Methodology and the Description of Program AQ11. Department of Computer Science : Reports of the Intelligent Systems Group, ISG 83-5, UIUCDCS-F-83-905. University of Illinois, Urbana (1983).
8. Michalski, R. S.; Kaufman, K. The AQ19 System for Machine Learning and Pattern Discovery: A General Description and User's Guide. Reports of the Machine Learning and Inference Laboratory : MLI 01-2. George Mason University, Fairfax (2001).
9. Michalski, R. S. ATTRIBUTIONAL CALCULUS: A Logic and Representation Language for Natural Induction. Reports of the

Machine Learning and Inference Laboratory : MLI 04-2. George Mason University, Fairfax (2004).
10. Wnek, J.; Michalski, R. S. Hypothesis-driven Constructive Induction in AQ17-HCI: A Method and Experiments. Machine Learning. Vol. 14, no. 2, pp. 139-168. (1994).
11. Wnek, J.; Kaufman, K.; Bloedorn, E.; Michalski, R. S. Inductive Learning System AQ15c: The Method and User's Guide. Reports of the Machine Learning and Inference Laboratory : MLI 96-6. George Mason University, Fairfax (1996).
12. Wojtusiak, J. AQ21 User's Guide. Reports of the Machine Learning and Inference Laboratory : MLI 04-3. George Mason University, Fairfax (2004).
13. Wojtusiak, J.; Michalski, R. S.; Kaufman, K.; Pietrzykowski, J. The AQ21 Natural Induction Program for Pattern Discovery: Initial Version and its Novel Features. Proceedings of The 18th IEEE International Conference on Tools with Artificial Intelligence. Washington, D.C. (2006).
14. Wojtusiak, J.; Michalski, R. S.; Simanivanh, T.; Baranova, A. V. Towards application of rule learning to the meta-analysis of clinical data: An example of the metabolic syndrome. International Journal of Medical Informatics, vol. 78, iss. 12, pp. 104-111. (2009).
15. Wojtusiak, J.; Alemi, F. Analyzing Decisions Using Datasets with Multiple Attributes: A Machine Learning Approach. Handbook of Healthcare Delivery Systems. CRC Press, (2010).
16. Wojtusiak, J.; Shiver, J.; Ngufor, C; Ewald, R. Machine Learning in Hospital Billing Management. Presentation : HIMSS 2011 Acedemic Forum. Hosted by AUPHA, Orlando, (2011).
17. Zurada, J. M.; Mazurowski, M.A.; Abdullin, A.; Ragade, R.; Wojtusiak, J.; Gentle, J. E. Building Virtual Community in Computational Intelligence and Machine Learning. Computational Intelligence Magazine, vol. 4, iss. 1, pp. 43-54. (2009).