Chapter 4.14
Beyond Classification:
Challenges of Data Mining for Credit Scoring

Anna Olecka
D&B, USA

ABSTRACT

This chapter will focus on challenges in modeling credit risk for new accounts acquisition process in the credit card industry. First section provides an overview and a brief history of credit scoring. The second section looks at some of the challenges specific to the credit industry. In many of these applications business objective is tied only indirectly to the classification scheme. Opposing objectives, such as response, profit and risk, often play a tug of war with each other. Solving a business problem of such complex nature often requires a multiple of models working jointly. Challenges to data mining lie in exploring solutions that go beyond traditional, well-documented methodology and need for simplifying assumptions; often necessitated by the reality of dataset sizes and/or implementation issues. Examples of such challenges form an illustrative example of a compromise between data mining theory and applications.

INTRODUCTION: PRACTITIONER'S LOOK AT DATA MINING

"Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

This basic KDD definition has served well as a foundation of this field during its early explosive growth. For today's practitioner, however, let us consider some small modifications: *novel* is not a necessity, but the patterns must be not only *valid* and *understandable* but also *explainable*.

*[...] process of identifying **valid, useful, understandable and explainable** patterns in data.*

A *data mining* practitioner does not set out to look for patterns hoping their practitioners discoveries might become useful. A goal is typi-

cally defined beforehand, and is usually driven by an existing business problem. Once the goal is known, a search begins. This search is guided by a need to best solve the business problem. Any patterns discovered, as well as any subsequent solutions, need to be understandable in the context of the business domain. Furthermore, they need to be acceptable to the owner of the business problem.

Successes of data mining over the last decade, paired with a rapid growth of commercially available tools as well as a supportive IT infrastructure, have created a hunger in the business community for employing data mining techniques to solve complex problems.

Problems that were once the sole domain of top researchers and experts can be now solved by a lay practitioner with the aid of commercially available software packages. With this new ability to tackle modeling problems in-house, our appetites and ambitions have grown; we now want to undertake increasingly complex business issues using data mining tools. In many of these applications, a business objective is tied to the classification scheme only indirectly. Solving these complex problems often requires multiple models working jointly or other solutions that go beyond traditional, well documented techniques. Business realities, such as data availability, implementation issues, and so forth, often dictate simplifying assumptions. Under these conditions, data mining becomes a more empirical than scientific field: in the absence of a supporting theory, a rigorous proof is replaced with pragmatic, data driven analysis and meticulous monitoring and tracking of the subsequent results.

This chapter will focus on business needs of risk assessment for new accounts acquisition. It presents an illustrative example of a compromise between data mining theory and its real life challenges. The section "Data Mining for Credit Decisioning" outlines credit scoring background and common practice in the U.S. financial industry. The section titled "Challenges for Data Miner"

addresses some of the specific challenges in credit model development.

DATA MINING FOR CREDIT DECISIONING

In today's competitive world of financial services, companies strive to derive every possible advantage by mining information from vast amounts of data. Account level scores become drivers of a strong analytic environment. Within a financial institution, there are several areas of data mining applications:

- Response modeling applied to potential prospects can optimize marketing campaign results, while controlling acquisition costs.
- Customer's propensity to accept a new product offer (cross-sell) aids business growth.
- Predicting risk, profitability, attrition and behavior of existing customers can boost portfolio performance.
- Behavioral models are used to classify credit usage patterns. Revolvers are customers who carry balances from month to month, Rate Surfers shop for introductory rates to park their balance and move on once the intro period ends. Convenience Users tend to pay their balances every month. Each type of customer behavior has a very different impact on profitability. Recognizing those patterns from actual usage data is important. But the real trick is in predicting which pattern a potential new customer is likely to adopt.
- Custom scores are also developed for fraud detection, collections, recovery, and so forth.
- Among the most complex are models predicting risk level of prospective customers. Credit cards lose billions of dollars annually in credit losses incurred by defaulted accounts. There

Related Content

Supporting Participation in Online Social Networks

Agostino Poggi, Paolo Fornacciari, Gianfranco Lombardo, Monica Mordoniniand Michele Tomaiuolo (2019). *New Perspectives on Information Systems Modeling and Design (pp. 122-142).*

www.irma-international.org/chapter/supporting-participation-in-online-social-networks/216335

Data Warehouse Maintenance, Evolution and Versioning

Johann Ederand Karl Wiggisser (2010). *Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction (pp. 171-188).*

www.irma-international.org/chapter/data-warehouse-maintenance-evolution-versioning/36614

A Multidimensional Model for Correct Aggregation of Geographic Measures

Sandro Bimonte, Marlène Villanova-Oliverand Jerome Gensel (2010). *Evolving Application Domains of Data Warehousing and Mining: Trends and Solutions (pp. 162-183).*

www.irma-international.org/chapter/multidimensional-model-correct-aggregation-geographic/38223

Real-Time Big Data Warehousing

Francisca Vale Lima, Carlos Costaand Maribel Yasmina Santos (2019). *Emerging Perspectives in Big Data Warehousing (pp. 28-57).*

www.irma-international.org/chapter/real-time-big-data-warehousing/231007

Continuous Auditing and Data Mining

Edward J. Garrity, Joseph B. O'Donnelland G. Lawrence Sanders (2005). *Encyclopedia of Data Warehousing and Mining (pp. 217-222).*

www.irma-international.org/chapter/continuous-auditing-data-mining/10596