(IJRST) 2012, Vol. No. 2, Issue No. I, Jan-Mar **ISSN: 2249-0604**

FILTRATION TECHNIQUES IN DATA MINING

***Ajay, $Rakesh Kumar, #Dr. P.K. Jakhar, @Dr. Anuj Kumar**
**, $Research Scholar, CMJ University*
#, @Research Guide

ABSTRACT

In this paper different subjects related to filtration are addressed. Although no completeness is aimed to, the following description deals with the most important subjects of filtration. At first an overview about filtration, particularly the principles of filtration, some aspects of beer filtration and filterability is given, followed by an introduction of Data Mining.

Keywords: *- Filtration, aspects*

FILTRATION

Principles of Filtration

The principle task of filtration is to separate a suspension in its solid and fluid components. Although several principles can be applied for filtration (see for example [1]) the databases available for DM have been collected for a kieselguhr filtration. Thus, in this context, filtration can be considered as a mechanical process which is used for the separation of suspensions with a wide distribution of particle sizes. Compared to other separation processes, filtration stands out with good separation characteristics and low energy demand at the same time [2]. The filter is permeable to the fluid but due to its pore structure the solid phase is not able to pass. The particulate material, depending on its interactions, deposits on the filter and forms a growing porous layer, which takes over the function of the filteraid (cake filtration), or attaches within the pores (deep bed / precoat filtration). Deep bed filtration is applied for the clarification of fluids with a marginal amount of solid particles [3]. Recently, cross flow filtration were at first implemented in brewery industries.

Description of methods

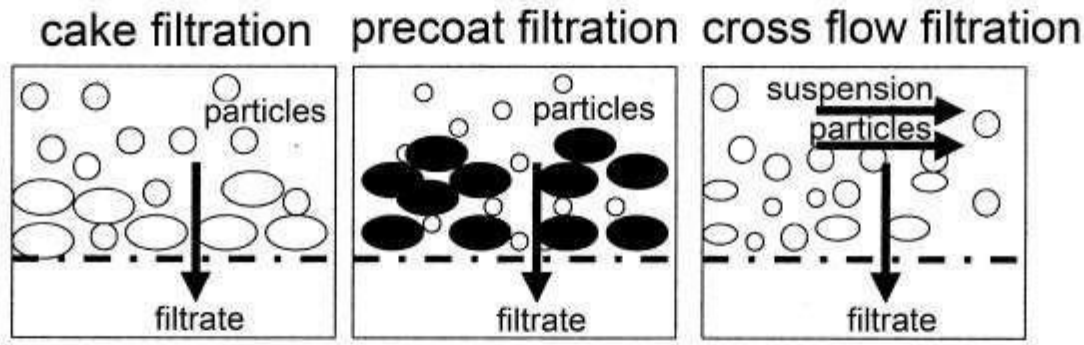

Figure 2.1: Scheme of cake, precoat and cross flow filtration

Filtration systems are subdivided by several criteria. One of these is the size of retained particles. Thereby, four main groups are defined:

• Filtration with pore diameter up to 10µm,
• Micro filtration with pore diameter up to 1µm,
• Ultra filtration with pore diameter up to 10−3µm,
• reverse osmosis.

Table 2.1: Filteraids: fields of application

application	clarification of	filteraid
food industry	beer, wine, juice, edible oil, treacle, starch hydrolysate	kieselguhr, perlite, silicic acid, wood flour
petroleum industry	petroleum, organic liquids	kieselguhr
metallurgy, mechanical engineering	rolling oil, pickling bath, hardening bath, galvanic bath	kieselguhr

Because of the deposit on or in the filter layer the efficiency of the process fades, either with a loss of pressure difference or, in case of a constant pressure filtration, with a reducing volumetric flow.

FILTRATION

Limited, a regeneration phase is required when reaching the limit. The use of filteraids provides a longer runtime of the filter. Common filteraids are diatomite, silicic acid and perlite. Their fields of application are described in table 2.1. At the precoat filtration two layers are established up on the filter with the help of filteraids . The first one, having a coarser structure, forms a supporting layer which is called filter cake. The second layer is applied continuously by dosaging kieselguhr to the suspension and therefore growing steadily. Thereby, the available filtration surface is assured, because the continuing addition of kieselguhr inhibits a blocking of surface.

Table 2.2: Diameter of particles in beer

Particle	Diameter [μm]
Yeast	$5 - 10$
Bacteria	$0{,}2 - 2$
Tanning protein	$0{,}2 - 1{,}5$
Polyphenol	$- 3$
β-glucans	$10^{-3} - 10^{-2}$

Consumer can evaluate the haze of beer, so all particles, which are visible for the human eye, have to be removed. Due to this, the scope of beer filtration is clarification of the product and providing a bright beer during its shelf life. The haze is mainly formed by micro organisms, like yeast cells or bacteria, and high molecular ingredients, like proteins or polysaccharides. Table 2.2 shows some of the relevant substances and their diameters. To assure the brightness of beer until end of shelf life, not only these particles have to be removed, but also those substances which can agglomerate within the time period in question. To monitor the clarification during filtration, the product is analyzed after passing the filter with optical methods. The larger particles, like yeast cells and other micro organisms, reflect applied light, so the reflection is measured in an angle of 25^0. The diffusion, caused by smaller particles, like high molecular substances, is measured in an angle of 90^0 to the light source. The result of this method is presented in form of haze values (h_{25}, h_{90}) and denoted in EBC (European Brewery Convention) units, an unit derived from a formalin haze standard [3]. An overview on this method is given by [4]. With respect to practical relevance, the h_{90} is the more regarded attribute, because the consumer realizes a turbid beer, caused by many small pending particles, more easily [5]. Even though new filtration methods like cross flow filtration have to be considered, today the common technology to achieve a bright, stable beer is precoat filtration with the filteraid diatomite. Schmidt [5] gives an overview on kieselguhr, its winning, refinement and disposal as well as on its attributes. Another aspect is the chemical and physical stability of beer. The recent research of Papp [6] and Kusche [7] addresses those problems in detail. The filtration of beer is influenced by various factors. Schur [8] classifies these influences in four categories:

- Filtration plant,
- filteraid,
- Mode of operation and
- Filterability.

Filtration Plant

Normally, the filtration process is split up into individual steps of defined partial targets. The appropriate adjustment of the subsequent steps is of crucial importance for the filtration success. These steps are commonly separation, precoat filtration, stabilization and finally a particle filter. Filteraid The first common filteraids were asbestos and cotton, today these are kieselguhr (diatomite), perlite or cellulose [9]. Diatomite is purified, dried and grinned skeleton of silicic algae. Diatomite is classified by its permeability and particle size distribution (fine, medium, coarse). The precise combination and amount of these fractions of kieselguhr influences the formation of the filtration cake and the characteristics of filtration. A finer kieselguhr causes a better clarification but effects the permeability of the cake. So the choice and combination of kieselguhr types are a compromise between clarification and pressure increase [3].

Mode of Operation

The level of automation and the control strategies of the plant influence the filtration. Sudden yeast surges, pressure differences or batches with different filterability change the conditions during filtration. Also the demands on shelf life differ between breweries. Modern plants are equipped with haze and pressure meters. Thus, volume flow of beer or the filteraid dosage can be regulated.

Filterability

The filterability has an important influence on filtration. The obtained filtration target should be reached with costs below 0.5 euro/hl. Only sufficient knowledge about the process enables the brewer to react in time. As this is one of the main aspects, filterability is discussed in detail in the following.

Some Aspects of Filterability

After clarification of beer by means of filtration was accepted at the beginning of the last century, it has been noticed that new problems are linked with this process. The common problems are on one side runtime of the filter, because of pressure difference reaching the critical value, and on the other side haze exceeding limit values. Due to this, there are numerous approaches to analyze the process [3].

The term"filterability" has been used in literature many times, but still there is no commonly accepted definition. It is often linked to pressure difference and filtration runtime. In rare cases the clarification has been considered. Kreisz [3] proposes therefore an objective evaluation, which is based on pressure difference, volumetric flow and clarification. Generally, these attributes together with kieselguhr and its characteristics are used to evaluate filtration and its quality. The basic problem of filterability is to identify the significant upstream influences on this important process type. In literature several approaches try to analyze a variety of parameters. Often these parameters are considered and analyzed as independent attributes. On one side influences of ingredients (polysaccharides, proteins, polyphones, melanoidines, minerals, yeast and bacteria) are mentioned and on the other side several brewing processes from mill to filtration are noted [10, 11, 12]. But not for all of these criteria scientific evidence is shown. The results can be summarized as follows [3]:

• **Influences by ingredients:**

– Raw materials

_ Polysaccharides

_ Proteins

_ Polyphonies

_ Mineral materials

– Yeast

_ Cell count

_ Products of autolysis

_ Polysaccharides

– Microorganisms

_ Polysaccharides

_ Mucous substances

• **Influences by Technology**

– malt quality

– milling

– mashing

– wort clarification

– boiling

– Whirlpool sedimentation

– Fermentation

– storing

Polysaccharides are high molecular carbohydrates. Among these macro molecules, with a molecular weight in a range of several thousand up to some million dalton, the beta- and alpha-glucans are outstanding. Concerning filtration, they are the most discussed beer ingredients. Their appearance in beer depends as well on malt as on technology. Kreisz [3] gives a detailed overview about polysaccharides and their influences on filtration.

Approaches to Prediction of Filterability

There are several systems to predict filterability. Some of these are described in this section. The tests can be divided in two groups: filter tests with constant pressure and tests with constant volume flow. According to Kreisz [3] the following list gives an overview of most important tests:

• **Test with constant pressure**
– Membrane filter test by Esser
– Filtration test by Siebert
– kieselguhr filter test by Raible
– kieselguhr filter test by Raible, Heinrich and Niemsch
– Filtration test by Webster and Molzahn

• **Test with constant volume flow**
– Zuercher test
– kieselguhr candle filter by Reed

Esser [13] took notice of correlation of filtration in practice and membrane filtration of the same solution. He developed a system based on a filtration of unfiltered beer with a 0, 25 µm pores in laboratory scale. The volumes V_1 and V_2, which have passed the filter at the timestamps t_1 and t_2 are measured and applied to a V/t, t - diagram. The gradient of the connecting line of t_1, V_1 and t_2, V_2 is used as a description of filterability. In contrast to this, Raible et al. [14] introduced a filtration layer of steel canvas with 15 µm pores. Kieselguhr is precoated on this layer. The filtration time and volume are measured as well as the haze. Based on this volume and runtime the filtration cake coefficient a can be determined as follows:

$$a = \frac{t}{V^2} \left[\frac{s}{m^2}\right] \tag{2.1}$$

This coefficient, Kreisz [15] calls it the filter cake factor, is influenced by the haze of unfiltered beer and the applied kieselguhr. To guarantee the reproducibility the same kieselguhr has to be used. With the cake factor the specific filtration volume can be calculated according to Kreisz [3]:

$$F_{spez} = \sqrt{\frac{3600}{a}} \cdot 0,1 \frac{hl}{m^2 h} \tag{2.2}$$

Another approach, which combines several aspects for a"filterability and stability check", is supplied by Annemueller and Schnick.

(IJRST) 2012, Vol. No. 1, Issue No. IV, Jan-Mar **ISSN: 2249-0604**

DATA MINING

Principles of Data Mining

"The convergence of computing and communication has produced a society that feeds on information. Yet most information is in its raw form: data. If data is characterized as recorded facts, then information is the set of pattern, or expectations, that underline the data. There is a huge amount of information locked up in databases — information that is potentially important but has not yet been discovered or articulated" [15]. In literature Data Mining (DM) is described as the core of"Knowledge Discovery in Databases" (KDD), but also both terms are used synonymic. DM is the process of extracting implicit, previously unknown, and potentially useful information from data. It is also defined as the automatic or (more commonly) semi-automatic process of discovering patterns in data. The discovered patterns must be meaningful in that they lead to some advantage, usually an economical one. KDD has intersecting aims like statistics, but can be separated from this by several features. The datasets examined by KDD investigators are commonly larger than those of statisticians and statistics is mostly concerned with static data in opposition to the evolving data. But also, DM overlaps in many cases with statistics. Hand [17] discusses the similarities and differences of statistics and KDD. To be used for the non trivial prediction of new data, the patterns must be expressed. This expression can take place in two different ways, as a black box model or a transparent box model. The transparent box reveals the structure of the pattern, whereas the interior of the black box is effectively incomprehensible. Both modeling techniques yield good prediction, but the results of the transparent box are represented in terms of a structure that can be examined, reasoned about, and used to inform future decisions. These patterns are called structural patterns because they capture the decision structure in an explicit way [16]. Thus, with transparent box models not only good prediction is realizable, but also gaining knowledge of the examined process.

DM is a multi disciplinary field that includes database theory, statistics, artificial intelligence (machine learning, pattern recognition) and visualization methods. Nakhaeizadeh et l. describe the target of DM with a set of sub targets. Segmentation. Divisioning of data in small, homogeneous, interesting and reasonable subsets or classes. This step is used for preparation of raw data, to create subsets which can be analyzed easier. Classification. Finding common attributes of database objects is the target of classification. These attributes are assigned to defined classes of a classification model. The class names can be predefined or result from segmentation. Concept description. A concept includes all basic characteristics of a class. Prognosis. Scope of prognosis is to supplement missing, numeric attributes to objects. In contrast to classification the target variable is a numeric value. Data description and aggregation. Data description and aggregation help the user to understand data. They allow a description of the basic characteristics of data in compact form. The data aggregation provides a clear structure and helps the user to get an exact image of the targets and data structure. Anomaly detection. If an object differs from an expected value or norm, then it is an anomaly. This behavior is significant for a problem which has to be solved; an anomaly can also refer to a yet unknown problem

(IJRST) 2012, Vol. No. 1, Issue No. IV, Jan-Mar ISSN: 2249-0604

which has to be examined. Interconnection analysis. The Interconnection analysis searches for models that describe a correlation between attributes of an object. These interconnections are used to describe the probability, for the appearance of a value with given attributes. Fayyad et al. also separate the DM-process in sub processes, given by figure 2.2.

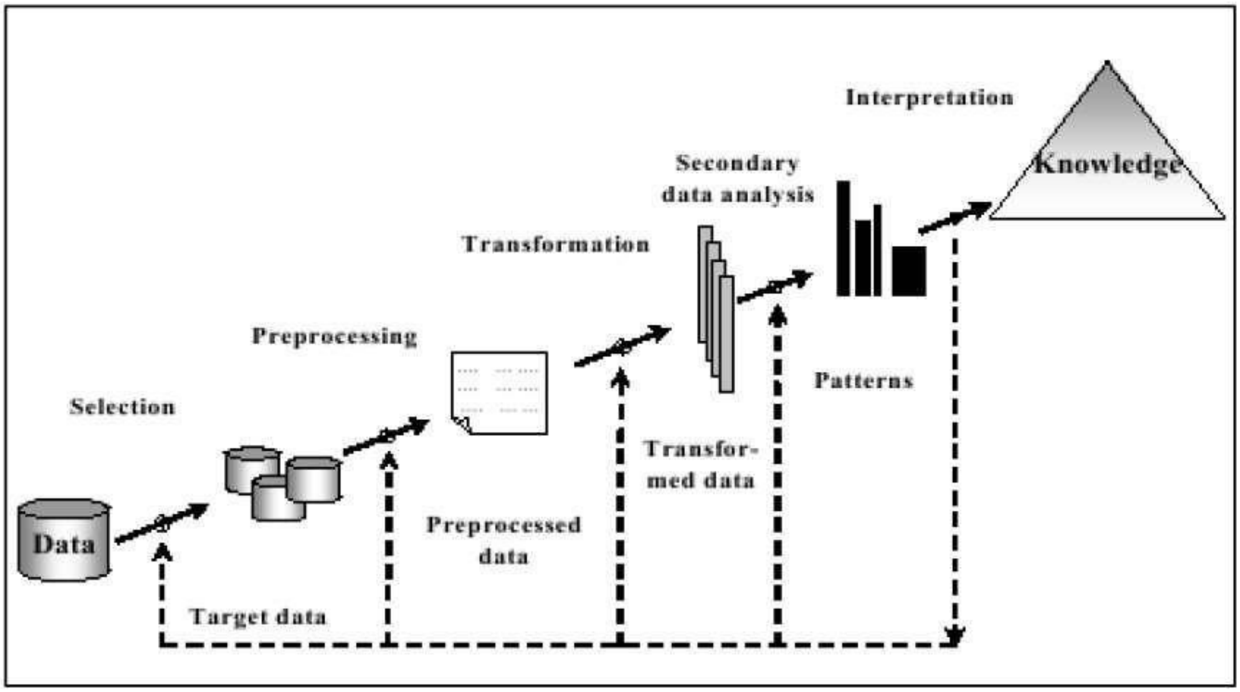

Figure 2.2: Overview of the steps comprising the KDD process

Application of Data Mining Methods

This section gives an overview on fielded applications of Data Mining, according to the papers of Provost et al. [17] and Langley et al. [19] and recapitulating textbooks [19]. Image recognition. Burl et al. [20] and Kubat et al. [21] present applications to image classification, for cataloguing volcanoes on the planet Venus and for detecting oil spills at sea. Fayyad et al. [22] used also a Decision Tree at the Second Palomar Observatory Sky Survey for classifying sky objects like stars or galaxies. Medicine. Lee et al. [24] and Finn et al. [23] used these techniques in scientific analysis and discovery, for predicting chemical carcinogenicity and for pharmacological discovery. Chemical process control. Leech [25] predicts the quality of uranium dioxide powder pellets, which are used in nuclear plants, with the help of a Decision Tree method. Credit decision. Loan companies regularly use questionnaires to collect information about people applying for a credit. Based on 1,014 training cases and 18 descriptive attributes, Michie [26] implements a Decision Tree at American Marketing. DM methods are often used to analyse market penetration, market development and product development [27].

CONCLUSION

The preceding two parts of this section present the state of the art of the two fields Data Mining and Filtration. Together, they provide the motivation and target of this paper, as they show on the one hand the problems in predicting filtration and on the other hand the aptitude and successful implementation of Data Mining methods for Knowledge Discovery and prediction tasks. Previous investigations carried out at the Chair of Fluidmechanics and Process automation of the Technical University of Munich prove those motivation and target. Under the leadership of Professor Delgado the investigation on aptitude of cognitive algorithms started in 1996. Moreover, methods like Fuzzy Logic or Artificial Neural Networks are used for recognition of damages on beverage crates [28, 29, 30], modeling of multi-stage high-pressure inactivation of micro organisms [31, 32, 33], state detection and feedback control of anaerobic wastewater treatment , modeling and optimization of fermentations and the prediction of flow fields.

REFERENCES

1. Kl¨osgen, W., Z˙ ytkow, J. M. (2002). Handbook of Data Mining and Knowledge Discovery.Oxford University Press.
2. Glass, K., Karopka, T. (2002). Genomic Data Explosion The Challenge for Bioinformatics?Advances in Data Mining, S.80-98.
3. Hand, D. J. (2002). Statistics. Handbook of Data Mining and Knowledge Discovery. OxfordUniversity Press.
4. Langley, P. and Simon, H. (1995). Applications of Machine Learning an Rule Induction. Communcications of th ACM. Nov 1995/ Vol. 38, No.11.
5. Provost, F., Kohavi, R. (1998). On applied research in Machine Learning. Machine Learning,vv, 1-6.
6. Ye, N. (2003). Handbook of Data Mining. Lawrence Erlbaum Associates.
7. Kubat, M., Holte, R. C., Matwin, S. (1998). Machine Learning for the detection of oil spillsin satellite radar images. Maschine Learning.
8. Grabmeier, J., Rudolph, A. (2001). Techniques of Cluster Algorithms in Data Mining. DataMining and Knowledge Discovery.
9. Lee, Y., Buchanan, B. G., Aornis, J: M: (1998). Knowledge-based learning in exploratoryscience: Learning rules to predict rodent carcinogenicity. Maschine Learning.
10. Finn, P., Muggleton, S., Page, D., Srinivasan, A. (1998). Pharamacophore discovery using the inductive logic programming system PROGOL. Maschine Learning.
    a. Za¨ıt, M., Messatfa, H. (1996). A Comparative study of clustering methods. FutureGeneration Computer Systems.

11. Michaud, P. (1996). Clustering techniques. Future Generation Computer Systems.
    a. Bacon, L. D. (2002). Handbook of Data Mining and Knowledge Discovery. Oxford University Press.
12. Zacharias, J., Hartmann, Ch., Delgado, A. (2001). Recognition of Damages on Crates of Beverages by an Artificial Neural Network. Accepted in Proceedings of EUNITE 2001 (European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems).
13. Kilimann, K. V., Hartmann, C., Delgado, A., Vogel, R. F., G¨anzle, M. G. (2005). A Fuzzy logic based model for the multi-stage high-pressure inactivation of Lactococcus lactis ssp. Cremoris MG 1363, International Journal of Food Microbiology 98 (1), pp. 89-108.
14. Kilimann, K. V., Hartmann, C., Delgado, A., G¨anzle, M. G. (2004). Data mining and fuzzy modelling of high pressure inactivation pathways of Lactococcus lactis. 3rd International Conferences on High Pressure Bioscience and Biotechnology, Rio de Janeiro, Brazil.
    a. Kilimann, K. V., Hartmann, C., G¨anzle, M. G.(2004). Application of fuzzy logic and statistical analysis for modelling high-pressure induced microbial inactivation. (P. Neittaanm¨"aki, T. Rossi, K. Majava, O. Pironneau) 4th European Congress on Computational Methods in Applied Sciences and Engineering (Eccomas 2004), Jyv¨askyl¨a/Finnland.