Data Mining and Data Gathering in a Refinery

MAHMOUD REZA SAYBANI[a], TEH YING WAH[b]

[a,b] Department of Computer Science and Information Technology

University of Malaya

50603 Kuala Lumpur

MALAYSIA

a) saybani@siswa.um.edu.my , b) tehyw@um.edu.my   http://fsktm.um.edu.my

Abstract:

This article handles one of critical steps of data mining, which is data collection. It will show how the researcher could get access to the valuable data of a refinery. And it explains the procedures of refining criteria for data collection. It also briefly explains the oil refining procedures to make the concept of data gathering at the refinery easier to understand. Each manufacturing company has its own specifications and rules that are needed to be considered when collecting data. As such the result of data gathering is almost always different for different manufacturing companies.

Key-Words: Data gathering, data collection, data mining, oil refinery

1. Introduction:

Data mining algorithms play an important and successful role in many manufacturing companies including oil refineries. Profit management, quality and process control in a refinery are other emerging and interesting areas where data mining methodologies find useful applications.

The complex workflow during oil refining automatically produces huge amount of data, therefore data mining is becoming increasingly important in oil refining industries. Data gathering is thus a vital issue for refineries. There are many articles about the importance and need of data gathering[1, 2], these particular articles deal with the way of data collection from sensors. According to [3], data gathering is an inherent part of improving services, tools and techniques. Without it (data gathering), it would be very difficult to measure demand, capacity and capability.

2. Challenges of data collection in a refinery

There has been no research on how challenging data collection at a refinery can be. Refineries are economically and politically very important, and there is a big competition among them. Therefore getting

access to their numbers, i.e. data and strategies is almost impossible.

Refineries are not only places, where strategic goods are produced, they are also places for new innovations; and in many places around the world they are major employers of their regions. What happens at refineries, it affects people and the economy. This article describes ways of data gathering at a refinery.

The most important criteria in data collection is building and/or creating trust between refinery management and the researcher. The management wants to make sure that their data are not misused and that they will beneficial to the progress and improvement of the refinery. Without the trust, no one can collect the data. One of the researchers of this paper used to work for a refinery for many years, and therefore this major step was already taken care of. But still getting the data needed more efforts.

To better understand the procedures of data gathering at a refinery, it is essential to understand the process of oil refining. At least one should get a wide spectrum of what is going on there.

3. Understanding Oil Refining Process:

Crude oil and natural gas are the bases for petroleum; they occur naturally in the ground and were formed millions of years ago. Different oil fields have different crude oil; it varies in color, composition, from a pale yellow low viscosity liquid to heavy black 'treacle' consistencies. After extracting crude oil and natural gas from the ground, they are transported by pipeline and or ship to refineries, where they get processed into valuable products such as petrol, diesel, lubricating oil, fuel oil, petrochemicals, asphalt, waxes and natural gas. In order to get to these products a coordinated arrangement of manufacturing procedures are carried out to make physical and chemical changes in crude oil, the result of this conversion are the above mentioned everyday products.

Crude oil consists of a mixture of hydrocarbon compounds and other materials such as oxygen, nitrogen, sulphur, salt and water. In the process of refining, most of the non-hydrocarbon substances are removed and oil is broken down into its components and then mixed into new products. [4]

For the actual refining, the stored heavy crude oil has to be cleaned of contaminants such as water and sand, after this is done, the clean feedstock gets pre-heated, before reaching the furnace. In furnace, because of extreme heat chemical splitting of the crude oil into fuel gas, liquid products and residuum occurs. The longer the hydro-carbon molecule, the higher temperature it requires boiling it out of crude oil. Based on different boiling points for different products, following products or substances are produced, from lightest to heaviest or from the top of the tower to the bottom: gasoline, kerosene distillate, light gas oil, heavy gas oil (used for lubrication oils). Large amount of heavy residuum is subject to even more refining. This residuum receives additional heating in vacuum tower, where light vacuum gas oil and heavy vacuum light gas oil are extracted. The remaining solids can either become asphalt or can be sent to another heat exchanger and furnace, the hydrocracker unit. In hydrocracker unit, the residuum will be subject to heavy and intense pressure, heat, catalysts and hydrogen gas. Other Gases such as hydrogen sulfide, carbon dioxide, carbon monoxide, and liquids and some solids are extracted. The liquids are then sent to fraction unit for further processing. The unconverted residuum is sent to cocking unit, where further gases and cock is extracted.[5]

Figure 1 illustrates production of some products in a typical refinery.
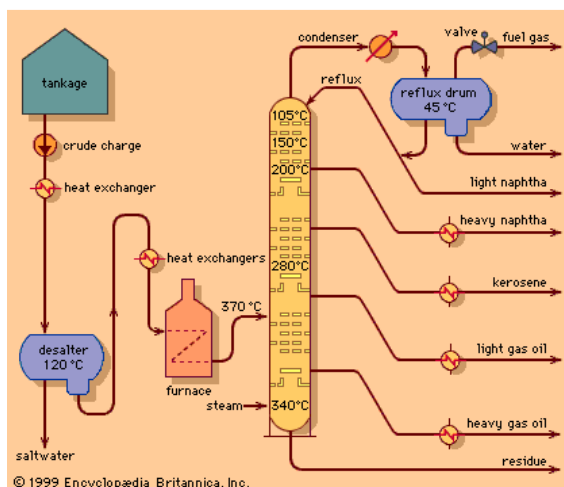
63

Fig. 1: Distillation Column [6]

4. Methodology of data gathering

The process of refining is very complex and alongside the very precise activities a lot of data are produced.

Initial data gathering in a refinery is like conducting a survey, similar to the way Modell suggests[7], researchers of this paper contacted responsible and knowledgeable engineers at the refinery for a variety of purposes and with some goals in mind. Those talks were conducted for:

1. Becoming familiar with process of oil refining
2. Fact gathering
3. Uncovering potential use of data mining
4. Finding out about refinery's need
5. Evaluation of previously collected facts
6. Verification of information collected from other resources
7. Confirmation of information gathered from different departments
8. Following up, amplification, and clarification
9. Gathering viewpoints and opinions of the engineers on how to improve the process of data management.
10. Providing information about this research

11. Getting leads for further data collection.

At the refinery the data are collected automatically based on the OPC technology; one of the researchers was a team member responsible for developing of applications, which enabled various refinery units to read data collected from many sensors across the refinery. Each unit collects its own data automatically, and the engineers use some of the data for their analysis or reporting purposes, for predicting purposes there was a need for data mining. To collect data for a specific goal, researchers had to go through many steps. Usually collecting data in a refinery is not easy; the first step was to introduce the concept of data mining and the need for data gathering, researcher had to sit down for many weeks with different engineers to explain the process of the research. Researcher found it useful to introduce the concept of the data mining by bringing samples that are taught in most data mining books.

Talks were necessary to understand the process of data flow and to get a broad understanding of production in a refinery. Talking to experts was an opportunity to gain knowledge at least to that extent that one can get a relative good overview of what is going on, what are requirements and what are possible usages of the gathered data.

Next step was to put a plan, how to start the data collection. In a refinery there are many places, where data could be collected, so after identifying some important sources of the data, due to its complexity, we decided to limit the research to some major products and to some important features.

After initial talks and discussion, the researcher asked experts to present their data. Some data were collected in Microsoft

Excel sheets, and others were collected automatically using the above mentioned applications based on OPC client/server technology.

Data entered in Microsoft Excel sheets were results of each days' activity. One issue that initially researchers were confronted with was the fact that the formats of Excel sheets were not unique. Over the years the format of data was changed based on the different needs and requirements of the managers. So first thing the researchers did was to extract data from hundreds of Microsoft Excel sheets and put them in a specific format in a single Excel sheet. After data transfer, the next step that was carried out was to clean the data, which is also called data cleansing.

Table 1 shows how initially data were collected in an Excel sheet.

SAMPLE TEST		L.S.R.N	H.S.R.N	B.N	KERO.	GAS OIL	L.V.G.O
Density @15°C			739.0		810.0	857.0	886.0
I.B.P	°C		77		184	257	267
5%	°C		90		190	269	279
10%	°C		94		197	279	292
30%	°C		105		207	295	312
50%	°C		115		215	307	332
70%	°C		125		225	320	350
90%	°C		139		240	343	376
95%	°C		144		245	354	383
F.B.P	°C		160		265	362	391
COLOR	M	30	30	30	30	1.0	1.5
	A	30	30	30	30	1.0	1.5
	N	30	30	30	30	1.0	1.5
H₂S ppm	M	---	NIL				
	A	---	NIL				
	N	---	NIL		----	----	
R.V.P (Kpa)	M	98	FL.P T °C M	43	75	130	136
	A	96	A	43	75	128	122
	N	96	N	42	68	132	128

Table 1: A scheme of data collection of one day for many products.

Table 2 shows a scheme of extracting daily data and putting them into one table or Excel sheet.

Pr. Name	Density @ 15C	IBP	5%	10%	30%	50%	70%	90%	95%	FBP	Color-M
LSRN	675.00	35	46	48	54	61	81	87	90	98	30
LSRN	675.00	34	44	46	53	61	71	86	93	97	30
LSRN	674.50	36	47	48	54	61	72	88	95	99	30
LSRN	677.50	35	47	49	55	63	73	86	92	96	30
LSRN	667.00	35	41	44	48	54	62	77	87	96	30
LSRN	670.00	35	45	48	54	62	73	88	95	101	30
LSRN	676.30	35	46	48	54	61	71	87	94	99	30
LSRN	671.00	36	44	46	50	57	66	83	93	99	30
LSRN	668.00	31	39	40	44	49	57	73	80	90	30
LSRN	668.00	31	39	40	44	49	57	73	80	90	30
LSRN	668.00	34	42	44	49	56	67	84	89	96	30
LSRN	670.00	36	42	45	50	56	64	79	88	96	30
LSRN	682.00	39	50	52	57	63	72	87	94	97	30
LSRN	667.00	36	42	44	50	57	68	85	93	96	30
LSRN	669.00	34	44	45	51	58	69	86	93	98	30

Table 2: Consolidated data collected for one product during a specific period of time, after data cleansing

Sometimes the experts didn't need to record all parameters involved in the production, sometimes they collected data on a weekly basis, and therefore the researcher had to separate those data from others that were more or less regular. Choosing the features in data mining is very essential for the outcome. With the help of experts at the refinery, the researchers identified some useful features that have impacts on the quality of some of the mid and or final products. Of course this does not eliminate the need for mathematically identifying relevant and less relevant features. Next step is to use or develop methods, algorithms and data mining techniques to verify whether or not the goals are achievable, and if yes, how.

5. Conclusion and discussions:

One major factor for data gathering is creating and/or building trust between the refinery management and the researcher a researcher has to convince the management of the need for such a project. After building trust, finding common language of understanding between the researchers and the experts is very essential, researcher must be prepared to learn new things and digest the knowledge. Going through numbers and

figures, and by looking at charts and graphs with the help of experts, makes it easier to understand the production process, which in turn helps to identify features that are most likely to be important for a data mining analysis. Getting the actual data is one major step to work on the models, but before that the data needs to be cleansed and put in a format that is plausible and useful for data mining algorithm. There are few thoughts that researchers would like to document here for future research candidates. Start everything by documenting all relevant actions. Having a check list to verify the goals of data collection is definitely very helpful, especially when one collects a lot of information about many features. Collecting

products' specifications, tables, charts and graphics help to understand things much easier, and it helps to ask qualified questions, while conducting data collection.

6. Acknowledgements:

References:

[1] V. N. Himanshu Gupta, Samir Das, and Vishnal Chowdhary, "Efficient Gathering of Correlated Data in Sensor Networks," *ACM Transactions on Sensor Networks,* vol. 4, no. 1, pp. 31, 2008.

[2] J. Kohonen, *Data Gathering in Sensor Networks*, University of Helsinki, Finland., Helsinki, 2004.

[3] NHS. "Data Gathering," http://www.tin.nhs.uk/index.asp?pgid=1133 .

[4] "Refining of Petroleum," http://www.world-petroleum.org/education/petref/index.html.

[5] "Refining Process," http://www.lloydminsterheavyoil.com/refining.htm.

[6] E. Britanica, "Schematic diagram of a crude-oil fractional distillation column.," 1999.

[7] M. E. Modell, *A Professional's Guide to Systems Analysis*, 2nd ed., 2007.