APPLICATION OF DATAMINING METHOD (ID3) TO CORRELATION OF FEED PROPERTIES AND REACTIVITY FOR HDS OF DIESEL FUEL

Narinobu Kagami, Ryuichiro Iwamoto, Masaru Amano, and Tetsuji Tani

Manufacturing Deparment, Idemitsu Kosan Co., Ltd.
26 Anesakikaigan Ichihara-city Chiba, 299-0107, Japan

Introduction

In recent years, much attention has been focused on the deep hydrodesulfurisation (HDS) of diesel fuel oil because environmental legislation has become more and more strict. Then it becomes more important to estimate the reaction temperature correctly to meet the target of sulfur at various conditions using various feedstocks. Because too deep desulfurisation makes the catalyst life shorter than producing correct target sulfur diesel[1)], determining the accurate corrected WAT (weight average temperature) is very important for the management of the diesel production and the catalyst life span.

It is well known that corrected WAT is used to evaluate the catalyst life. And WAT is a function of the operation condition factors such as temperature, LHSV, Hydrogen pressure, Hydrogen-oil ratio and of the feed properties. Operation condition factors can be normalized from the formula based on n-order (pseudo) reaction and Ahrrenius plots through tuning the constant. These operation factors can be changed independently using pilot tests, however, it is almost impossible with a real feed to vary just one property independently. Therefore, the suitable selection of key factors of feed properties might be difficult using real feed.

The formula of the corrected WAT normalized to the standard from various feed properties was made by trial and error method somehow, using the factors chosen by the information of literature[2)]. However, we still need a convenient way to evaluate the validity of the chosen factors.

Recently, datamining methods tend to be used more commonly in data processing field. Datamining is one of the technologies for building knowledge-based systems by inductive inference from examples. Therefore, it can also be the method to search the valuable information from large data sets. In dataming methods, the classification analysis called ID3 (Iterative dichotomiser) is applicable.

ID3 algorithm was suggested by J.R Quinlan[3)] and extended to deal with numerical values[4)]. Using entropy of information that shows the degree of chaotic state, ID3 establishes the importance of the factor automatically.

Then ID3 was applied to evaluate the validity of the chosen factors from various feed properties for the corrected WAT normalized to the standard.

Experimental

Catalyst and Pretreatment. 100cc of commercial catalyst of CoMo type catalyst was filled in a high-pressure micro-flow reactor. Before the evaluation of SRLGO, the catalyst was presulfided in situ at 250 °C for 2h and then 300 °C for 2h in stream of a dimethyl-disulfide/SRLGO/hydrogen mixture (sulfur content, adjusted as 2.5wt%).

Pilot Test. Pilot tests were carried out using high-pressure micro-flow reactor. The temperature of the catalyst bed was kept constant to avoid temperature difference over the catalyst bed by heater blocks. Thermocouples were located in the center of the catalyst beds. In these range of the operation conditions, the results of sulfur content of the product were from 500 to 5 ppm.

Analysis of feeds and products. Six kinds of straight run light gas oils (SRLGO) from Middle East crude were tested in high pressure flow reactor. The various properties of feedstocks were measured as shown in Table 1.

The sulfur content of product was also measured. The product oils were stripped by nitrogen to get rid of the H_2S in the liquid before the measurement.

ID3 data analysis. ID3 software package were used for the analysis of the data of pilot tests (XpertRule (Lascombe and Company, Ltd.)). Figure.1 shows the algorithm of ID3. A classification rule was expressed as a decision tree. The relationship between object Item A and attribute Items X, Y, Z were examined using the entropy of information. As step 0 shows, the entropy of the object item (I) was calculated. Next object item is divided according to the attribute items, and each entropy (E (X, Y, Z)) was calculated as step 1 shows. Then the subtract values of I – E (X, Y, Z) were information gains. And the largest information gain means that the information becomes less chaotic. Then, Item X was selected as first attribute and this routine was reiterated. In this way, ID3 can select the simplest decision tree.

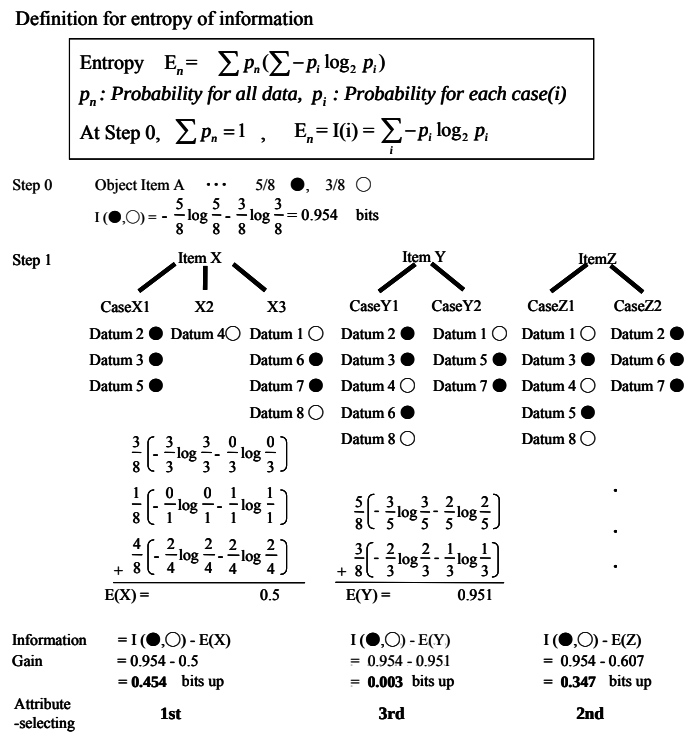

Figure 1. Outline algorithm of ID3(Iterative Dichotomiser).

Results and Discussion

Operation condition factors can be corrected independently, because each operation factor can be changed independently using the same feed stocks. It is well known that corrected WAT is used to evaluate the catalyst activity and its deactivation rate. The reaction temperature under the normalized conditions can be calculated as follows (Ahrrenius eq.1, n-order reaction eq.2, hydrogen-oil ratio eq.3, hydrogen pressure eq.4).

$$\ln k = \ln A - \frac{Ea}{RT}, \ln k'' = \ln A - \frac{Ea}{RT''} \qquad (eq.1)$$

$$k = \frac{LHSV}{n-1} \times \left(\frac{1}{S_p^{\,n-1}} - \frac{1}{S_f^{\,n-1}} \right) \qquad (eq.2)$$

$$k' = k \times \left(\frac{Hoil\text{-}Ratio_p}{Hoil\text{-}Ratio_S} \right)^{\alpha} \qquad (eq.3)$$

$$k'' = \exp(\ln k') \times \left(\frac{HP_p}{HP_S} \right)^{\beta} \qquad (eq.4)$$

In Fig.2, the actual WAT and corrected WAT standardized at a certain operation condition were shown. After the normalization of operation condition factors, the data scattering range becomes small. However, the scattering is still above 15 °C difference, depending on the properties on the kind of feed.

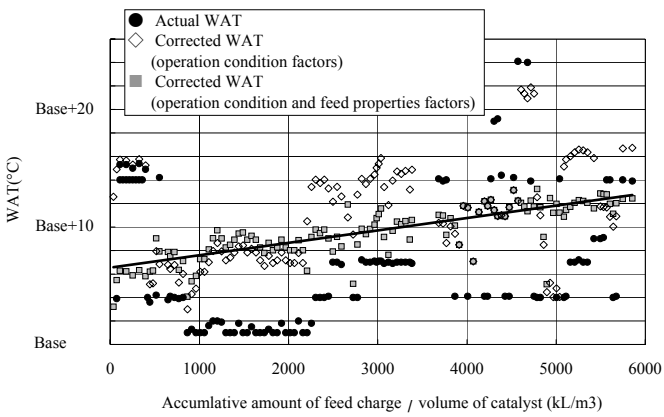

Figure 2. Actual and Corrected WAT for HDS reaction.

The WAT during the deep desulfurisation of the gas oil could be regarded to increase linearly with time on stream as shown in some literature[5]. When LHSVs were changed, accumulative amount of feed throughput per catalyst volume should be better to use instead of time on stream, to evaluate the deactivation pattern of catalyst.

Recently the inhibition of HDS reaction by nitrogen compounds was evaluated[6]. Especially the basic nitrogen compounds are thought to have the strongest effects to HDS reaction.

Moreover, it is well known that the heavier sulfur compounds (Dibenzothiophene (DBT) type sulfur compounds, especially 4,6-substituted DBT) has a lower activity than lighter sulfur compound such as benzothiophene type sulfur compounds.

Therefore, the following factors of feed properties were chosen for WAT normalized feed properties as eq.5 shows. "SpGr" is density of feed stocks and it is also used for reside hydrodesulfurization reaction. "Base N" is the content of basic nitrogen (Potentiometric titration). And "T90" is the temperature of 90% distillation (ASTM D-86), which means how heavy the feed is.

Then the trial and error method was used to determine the constant value of a, b, c (eq.5).

$$\ln A' = \ln A + a \cdot \ln\left(\frac{SpGr'}{SpGr}\right) + b \cdot \ln\left(\frac{BaseN'}{BaseN}\right) + c \cdot \ln\left(\frac{T90'}{T90}\right) \qquad (eq.5)$$

As deactivation pattern was regarded to be linear line, the least square method was applied and the standard deviation of all data was minimized changing each constant values. In this way, the values of constants were determined, and the results of normalization of feed properties are shown in Fig.2 Then, the difference from the linear line calculated by the least square became around 4 °C and much smaller than the one normalized only operation condition factors.

As described above, corrected WAT for feed properties was obtained. However, there is still the question whether the factors of feed properties chosen for normalization of the WAT are the best or not. To evaluate the validity of the chosen factors, ID3 method was used. Each difference between the data of corrected WAT normalized operation condition factors and the calculated deactivation line was examined.

Therefore, the data of subtraction from the diamond plot (◊) to the line in Fig.2 is the object item "A" and, "A" is examined by the feed properties factors, items "X, Y, Z …….. ". Operation factors are also used to convince that the normalization of them is. The one applied method is ID3, and the other is the correlation coefficient for the analysis. However, object item "A" should not be continuum but step number for the ID3 analysis. Then subtraction was classified per 2 °C.

Table 1 shows the results of the analysis. The effective factors for classification were arranged in order of increasing entropy (Class I to XII), using ID3 method. Clearly, the smaller the entropy, the more effective the factor is. However, in order to select the valid factors, they should be independent of each other. Therefore, the correlation coefficient analysis was also used, and the factors which have a coefficient above 0.8 are shown as the same pattern-code.

Table 1. The Results of ID3 and Correlation Coefficient.

Ranking of entropy (step 1, E(X,Y...))	**Effective factor for classification**					
I	Type of feed	.	.	.	.	.
II	**Density**	Nitrogen	**Basic nitroger**	.	.	.
III	**Saturate**	1-ring Aroma.	T50VOL%.	T60VOL%.	T70VOL%.	Viscosity
IV	Feed Sulfur	T80VOL%...	**T90VOL%...**	T95VOL%...	EP.	.
:	**:**	**:**	**:**	**:**	**:**	**:**
VII	Accumulative amount of feed	.	.	.	.	.
VIII	WAT	.	.	.	.	.
IX	LHSV	.	.	.	.	.
X	Product Sulfur	.	.	.	.	.
XI	H2/Oil	.	.	.	.	.
XII	H2 Press.	.	.	.	.	.

The most effective factor is type of feed, and it means that the feed properties factors normalization should be done. The second ranking factors are density, and nitrogen, and basic nitrogen. The correlation coefficient of nitrogen and basic nitrogen is higher, then it means that they had a monotonic relation. Therefore, only the basic nitrogen was selected considering the information described above. The third effective ranking is saturate. It was examined in the same way (eq.5) to minimize the standard deviation. However, it has almost no effect to reduce the standard deviation. The forth are Feed sulfur, T90 vol%, T95vol%, and EP, then the feed sulfur has already been used to calculate the reaction rate constant. T90 vol%, T95vol%, and EP could be considered to be the same factor, then, T90vol% should be taken.

As described above, the factors determined by trial and error method are shown in the bold type, and these are validated by ID3 and correlation coefficient analysis.

Acknowledgment. We thank Prof. Jacob A. Moulijn for generous support to publish this work.

References

(1) Knudsen, K. G.; Cooper, B. H.; Topsoe, H. *Applied Catalysis A-General,* **1999**, *189*, 205.
(2) Laredo, G. C.; los Reyes, J. A.; Cano, J. L.; Castillo, J. J. *Applied Catalysis A-General,* **2001**, *207*, 103.
(3) Quinlan, J. R. *Machine Learning*, **1986**, *1(1)*, 81.
(4) Araki, Kojima *Jinkou Chinou Gakkaishi*, **1992**, *7(6)*, 992.
(5) Gerritsen, L. A.; Vogt , K.; Huismans, A. Quinlan, *Summary At the ERTC in Berlin,* in November **1998**.
(6) Whitehurst, D. D.; Knudsen, K.; Nielsen, I. V.; Wiwel, P.; Zeuthen, P. *Abstracts of Papers of the American Chemical Society,* **2000**, *219*, 23-PETR.