

STUDY AND EVALUATION OF CLUSTERING ALGORITHM FOR SOLUBILITY AND THERMODYNAMIC DATA OF GLYCEROL DERIVATIVES

Shaofei Wu^{1,2,*}

^{*1} Hubei Province Key Laboratory of Intelligent Robots, Wuhan Institute of Technology, Wuhan, P.R.China

² School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, P.R.China

* Corresponding author; E-mail: wasbfc@yeah.net

The dissolution characteristics of glycerol derivatives in solvents at different temperatures and pressures were studied. The effects of solvent structure on gas absorption capacity and separation selectivity were analyzed. The absorption thermodynamics and kinetics were discussed. The experimental results show that as the temperature increases, the solubility of glycerin derivatives decreases, and the separation selectivity between gases also decreases. And at the same temperature, the more carbon atoms of the glycerin derivative, the easier the dissolution process in the ionic liquid. The thermodynamic parameters of each gas do not change much with increasing temperature. At the same time, the spectral clustering algorithm can be used to obtain the characteristics of the global optimal solution, which solves the problem that the traditional hybrid data clustering algorithm is easy to fall into the local optimal solution.

Keywords: *glycerol derivatives, solubility, thermodynamics, clustering algorithm*

1 Introduction

Esterified products of glycerol and fatty acids can be used in food, medicine, cosmetics and materials. Monoglycerides and diglycerides are excellent additives in baked goods, margarines, dairy products and sauces [1]. Monoglyceride is an amphiphilic molecule, so it is also a good nonionic surfactant and emulsifier [2-4]. Raw materials such as oil and water that are difficult to mix will mix well after adding monoglyceride. It can also improve the consistency of the product. It makes the skin care lotion feel better, so it is widely used in the cosmetics industry, especially in shampoo and bath products. In addition, monoglyceride has good lubricity and plasticity, also textiles and plastics. They are internal and external lubricants for manufactured mechanical systems.

Ferulic acid glyceride is also an important derivative of ferulic acid. Ferulic acid glyceride includes ferulic acid monoglyceride, diferulic acid glyceride and tri-ferulic acid glyceride, and its hydrophilic-lipophilic balance range is wide, so functional lipids of different nature can be synthesized as needed [5]. And its various functions can be exerted in a wider field. For example, it has a strong ultraviolet absorbing ability at 280 to 340 nm, and is used as a new sunscreen product in various cosmetics. Therefore, the study of glycerol ferulic acid has attracted people's interest. However, at present, the research on glycerol is still relatively rare at home and abroad [6-8]. The research is mainly related to the enzymatic synthesis of glycerol ferulic acid, and there are few studies on

chemical synthesis. It can be seen from the above that the reaction time of the synthesis of glycerol ferulic acid catalyzed is longer and the conversion rate is not high [9,10]. At present, the research on enzymatic synthesis only stays in the laboratory stage [11-14]. There are also many problems in the use of lipase. For example, when lipase is used in an organic solvent reaction system, the organic solvent is highly demanded, and the organic solvent is inactivated by the small enzyme.

In this paper, ferulic acid and glycerol were used as raw materials, and p-toluenesulfonic acid was used as a catalyst to synthesize ferulic acid monoglyceride by direct esterification. The synthesis conditions of the synthesis of ferulic acid monoglyceride were optimized by single factor experiment and orthogonal experiment. The kinetics of the synthesis reaction was studied. The reaction product, ferulic acid monoglyceride, was separated and purified. It was identified by UV and NMR, and the ability of ferulic acid monoglyceride to scavenge free radicals was investigated. A series of ionic liquids and eutectic cerium solvents were designed and synthesized, and the dissolution and absorption of glycerol derivatives in ionic liquids and eutectic solvents were investigated. At the same time, a spectral clustering algorithm that improves the mixed data distance metric was introduced, which was used to evaluate the thermodynamic data.

2. Experimental content

2.1. Experimental methods

(1) Synthesis of polyglycerol

Mix 40 g of glycerin with powdered NaOH, and slowly raise the temperature to 250 ~ 260 °C by N₂. Observe the synthesis according to the amount of water. After the reaction is finished, evacuate (250 ~ 260 °C / 1333 Pa) for 0.5 h, and cool down to 90 ~ 100 °C. Add acetic acid to neutralize the catalyst, so that the product pH = 6 ~ 7.

(2) Synthesis of acrylic acid modified ferulic acid monoglyceride

In the esterification reactor, polyglycerol and catalyst ZnO (according to 8% of the mass of acrylic modified rosin) were slowly heated to 240 °C, and acrylic modified rosin was added to the reactor in batches. The rosin is added within 1.5 ~ 2 h, the generated water is removed by nitrogen during the reaction, and the acid value is continuously sampled and analyzed. When the acid value < 10, the reaction is finished. The vacuum was heated at 240 ~ 250 °C and 1333 Pa for 0.5 h, and the acrylic modified rosin polyglycerol ester was obtained by cooling.

(3) Solubility of acrylic acid-modified ferulic acid monoglyceride in water: The solubility in water was measured at room temperature and under heating, respectively.

2.2. Reaction thermodynamics study

Accurately weigh 0.5 g of ferulic acid, the ratio of alkyd is to 30, and the amount of catalyst is 8%. It was placed at 65, 70, 75, 80, 85 °C, and the reaction time is 30min. After the reaction is completed with ethanol, the solution was made up to 50 mL. 1 mL was taken to a volume of 25 mL, filtered, and the monoester content was determined by liquid chromatography.

The concentration of ferulic acid monoglyceride was quantified by liquid chromatography, and the concentration of ferulic acid monoglyceride was obtained by a standard curve. The rate of esterification reaction R_g can be calculated as follows:

$$Rg(\text{mol} / \text{L} / \text{min}) = \frac{\text{Ferulic acid monoglyceride concentration}(\text{g/L})}{\text{Ferulic acid monoglyceride molar mass}(\text{g} / \text{mol}) \times \text{Reaction time}(\text{min})}$$

3. Results and discussion

3.1. Solubility of ferulic acid monoglyceride in sodium lauryl sulfate

The solubility of citric acid in ethanol aqueous solutions of 0.7741, 0.8804 and 0.9386 ethanol solutions were determined under normal pressure. The experimental results are shown in Fig. 1.

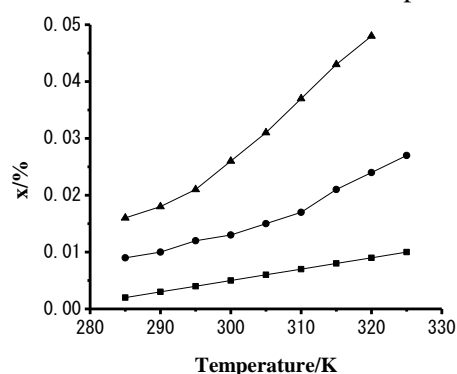


Figure 1 Solubility of ferulic acid monoglyceride in aqueous ethanol

As can be seen from Fig. 1, the hydroxyl conversion of glycerol and the selectivity of the target product start to increase as the catalyst loading increases. This is because the catalyst can be increased in activity by appropriately increasing the loading of the catalyst, so that the catalytic effect is better. When the catalyst is continuously added, the hydroxyl conversion rate of glycerol decreases, which may be due to the increase of the active site of the catalyst. It causes some side reactions other than etherification of glycerol, resulting in different by-products, and thus the selectivity is also reduced. The acrylic acid-modified ferulic acid monoglyceride and rosin were respectively in water (80 °C water bath) under the action of sodium lauryl sulfate (LAS), sodium dodecylbenzenesulfonate (K12) and OP-10, the results are shown in Tab. 1.

Table 1 Dissolution of acrylic acid modified ferulic acid monoglyceride in LAS/K12/OP-10

	mol/L	Acrylic modified ferulic acid monoglyceride/g	Rosin/g	Phenomenon
LAS	0.05	0.011	0.015	Insoluble after adding rosin, insoluble after cooling
K12	0.05	0.39	0.020	Insoluble after adding rosin, crystal precipitation during cooling
OP-10	0.05	0.63	0.041	After heating, it is milky white, the weight of dissolved rosin increases, and the cooling is yellow

As can be seen from Tab.1, the solubility of acrylic modified rosin polyglycerol and rosin is different in the same concentration of LAS, K12, OP-10 aqueous solution. It is larger in OP-10 and smaller in LAS. In addition, in K12, the solubility of acrylic acid-modified ferulic acid monoglyceride increased with temperature, but it was not obvious. From the molecular structure of ferulic acid monoglyceride, it is known that there is a phenolic hydroxyl group in the molecule of ferulic acid monoglyceride. The conjugation effect occurs due to the unpaired single electron interaction between the ytterbium electrons on the benzene ring and the oxygen atom. As a result, the unpaired single

electrons are directed toward the benzene ring, and the hydrogen-oxygen bond is weakened. Therefore, the phenol-containing substance has a more active hydroxyl hydrogen, which can provide hydrogen ions and compete for active oxygen. it can combine with free radicals, convert free radicals into inert compounds, and stop the chain reaction of free radicals.

3.2. Reaction process curve

Accurately weigh the ferulic acid monoglyceride standard sample with 95% ethanol to prepare different concentrations of solution. Determine the peak area of different concentration solutions by high performance liquid chromatography. Prepare the peak area of ferulic acid monoglyceride according standard curve of the concentration of ferulic acid monoglyceride, and the results are shown in Fig. 2.

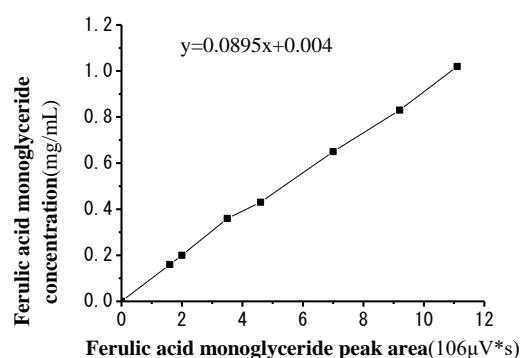


Figure 2 Standard curve of concentration to peak area of glycerol ferulate

The esterification reaction rate is expressed as a change in the concentration of ferulic acid monoglyceride produced by the reaction per unit time. During the esterification reaction, the concentration of ferulic acid monoglyceride increases with the reaction time, and the reaction rate can be obtained from the relationship between the two. The concentration of ferulic acid monoglyceride in the reaction system at different reaction times was determined, and the reaction progress curve was obtained as shown in Fig. 3.

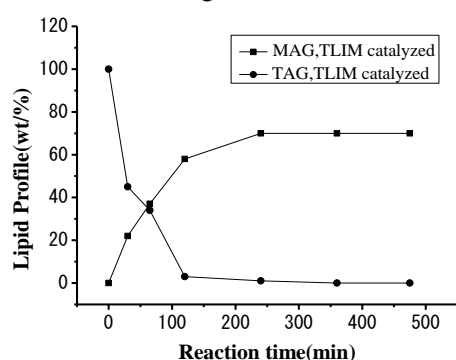


Figure 3 Curve of reaction course

As the slope of the curve in the graph gradually decreases with time, the reaction rate decreases with the increase of reaction time. The experiment shows that the concentration of ferulic acid monoglyceride in the initial stage of the reaction (within 30 min) has a linear relationship with time, which can be approximated. The initial reaction rate is constant. The reaction time of this experiment was 30 min, and the approximate initial reaction rate was obtained, so 30 min was selected as the

kinetic study time. As the reaction time increases, the catalyst catalyzes the glycerol etherification reaction more fully, and the indicators also increase. When the reaction time is extended from 8 h to 9 h, the conversion and selectivity of glycerol hydroxyl groups are not much changed, because as the reaction progresses, many by-products, such as water, have a very great impact to the reaction. And so the removal of water from the etherification process is very important.

3.3. Dissolution thermodynamic analysis

3.3.1. Relationship between esterification rate and ferulic acid concentration

The concentration of glycerol and p-toluenesulfonic acid was kept constant, and the reaction was carried out at 80 °C for 30 min. When the ferulic acid concentration is in the range of 0 ~ 0.3609 mol/L, the relationship between the esterification rate R_g and the ferulic acid concentration ($C_{\text{ferulic acid}}$) is shown in Tab. 2.

Table 2 Relationship between esterification reaction velocity and concentration of ferulic acid

Ferulic acid concentration c ($\times 10^{-1}$ mol / L)	1.032	1.550	2.066	2.587	3.096	3.609
Reaction speed ($\times 10^{-4}$ mol / L / min)	2.220	4.864	7.229	9.976	14.72	19.86

From the above table, the mathematics tool Matlab is used for mathematical analysis, and the least squares method is used for fitting. The results are as follows:

$$R_g = 1.9364 * [c_{\text{Ferulic acid}}]^{1.802} \quad (1)$$

The relationship between the ferulic acid concentration and the reaction rate is shown in Fig. 4.

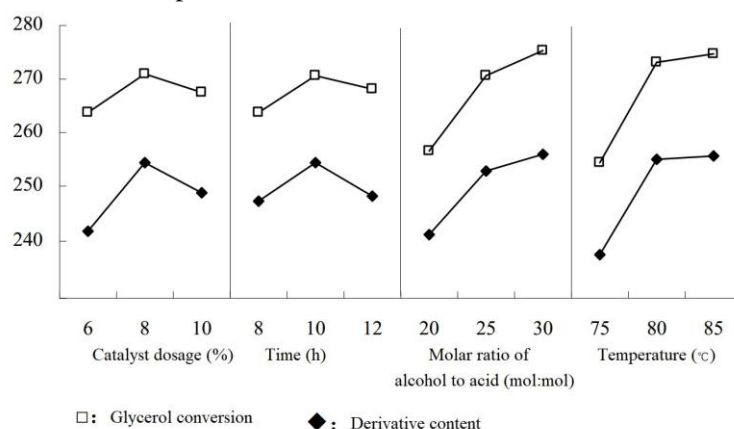


Figure 4 Influence of various factors at different levels on the reaction rate of esterification

It can be seen from Fig. 4 that the rate of esterification increases with the increase of ferulic acid concentration, which is due to the increase of ferulic acid concentration, the increase of the concentration of strontium salt, and the acceleration of esterification. In the range of ferulic acid concentration studied, within the range, the esterification reaction rate is substantially linear with the ferulic acid concentration. The reaction for ferulic acid has a reaction order of 1.802.

3.3.2. Relationship between esterification reaction rate and glycerol concentration

The concentration of ferulic acid and p-toluenesulfonic acid was kept unchanged, and the

reaction was carried out at 80 °C for 30 min, and the glycerol content was changed from 0.20 to 0.60 mol / mL. The results of the experiment which show the glycerin content and the reaction rate are shown in Fig. 5.

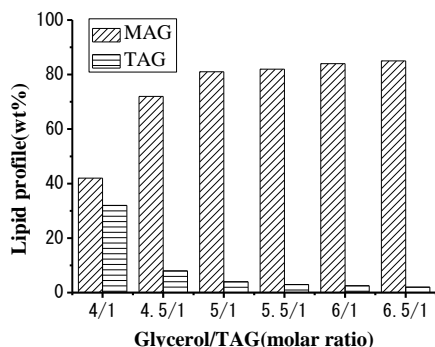


Figure 5 The effect of glycerol on the reaction rate of esterification

As can be seen from Fig. 5, the glycerol content has little effect on the reaction rate in the experimental range, because the glycerol is greatly excessive in the experimental range, so the esterification reaction rate is not significantly increased when the amount of glycerin is increased. The glycerol content is therefore combined with the reaction kinetic constant. From the above research results, it can be concluded that the relationship between the esterification reaction rate and the reactant concentration in the initial stage of the reaction is:

$$R_g = K [C_{\text{Ferulic acid}}]^{1.802} \quad (2)$$

K is a constant. From the reaction rate equation, the total reaction order of the reaction in the initial stage of the reaction was 1.802.

It can be known from the reaction rate equation that the concentration of ferulic acid has a great influence on the reaction rate. In order to shorten the reaction time, the concentration of ferulic acid can be increased, and the change of the concentration of glycerol has little effect on the reaction rate, and the concentration of glycerol cannot be increased. It cannot shorten the reaction time by increasing the reaction rate. This has guiding significance for the actual production process.

3.3.3. Effect of reaction temperature on reaction rate

The alkyd ratio, the catalyst addition amount and the reaction time were kept unchanged, and reacted at 35, 40, 45, 50, and 55 °C, respectively, to investigate the effect of temperature on the reaction rate. The results are shown in Fig. 6.

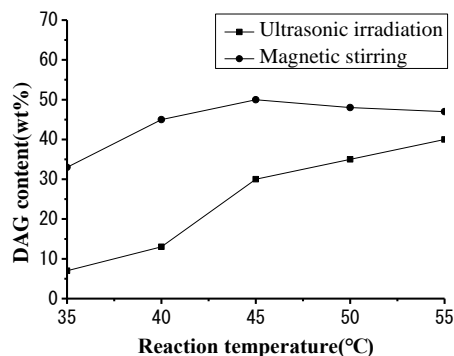


Figure 6 The effect of temperature on the reaction rate of esterification

From the above figure, the initial reaction rate of the synthesis reaction in the range of 45 to 55 °C can be obtained, and it can be seen that the initial reaction rate rapidly increases with temperature rise. Regarding the quantitative relationship between the rate constant K and the reaction temperature T , Arrhenius summed up a large amount of experimental data and proposed some empirical formulas, which are expressed as:

$$\frac{d \ln K}{dT} = \frac{E_a}{RT^2} \quad (3)$$

Where: E_a is called experimental activation energy, $\text{J}\cdot\text{mol}^{-1}$.

If you use the above formula as an indefinite integral, you can get:

$$\ln K = \frac{-E_a}{RT} + B \quad (4)$$

Where: B is the integral constant.

It can be seen from the above equation that the plot of $1/T$ with $\ln K$ should be a straight line with a slope of $-E_a/R$.

The relationship between $\ln K$ and $1/T$ obtained from the data in the table is shown in Fig. 7:

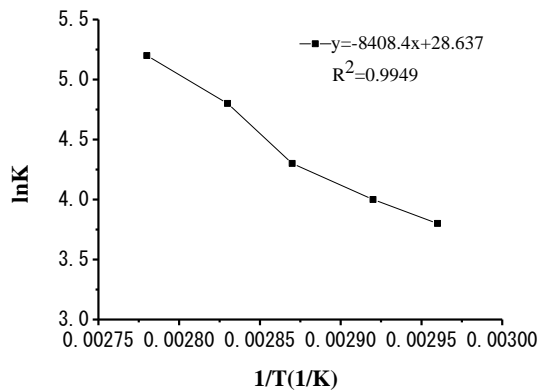


Figure 7 Curve of $\ln K$ to $1/T$

The experimental activation energy E_a can be obtained from the slope of the above line as -8408.4.

$$E_a = -R \times (-8408.4) = 69907.44 \cdot \text{mol}^{-1}$$

The absorption process is an exothermic process over the temperature range of this experiment. It can also be found from the calculation results that the gas absorption is a process in which the entropy value is reduced because the absorption process turns the disordered gas molecules into relatively ordered liquid molecules. Comparing the changes of thermodynamic parameters at different temperatures, it was found that the thermodynamic parameters of each gas did not change much with the increase of temperature.

3.4. Evaluation of thermodynamic data by clustering algorithm

When we establish the similarity matrix, we think that when the number of the same classification attribute is the same, the similarity between the data is larger. When the number of the same classification attribute is smaller, the similarity between the data is smaller, but the difference is similar to the class. We use the Gaussian kernel function to establish a similarity relationship for the classification attribute. The similarity function is defined as follows:

$$w_s(x_i, x_j) = m * \exp\left(-\left(\frac{d_s(x_i, x_j)}{\sigma_1}\right)^2\right) + a * \exp\left(-\left(\frac{d_s(x_i, x_j) - b}{\sigma_2}\right)^2\right) \quad (5)$$

We consider that the traditional spectral clustering algorithm can achieve good results in numerical clustering. For the numerical property data in the mixed data, we still use the general Gaussian kernel function to establish the similarity matrix:

$$w_s(x_i, x_j) = \exp\left(-\frac{d_n(x_i, x_j)}{2\sigma^2}\right) \quad (6)$$

For mixed data, since we assume that the classification attribute is as important as the numerical attribute, we define the mixed data similarity matrix as follows:

$$w_m(x_i, x_j) = \frac{w_s(x_i, x_j) + w_d(x_i, x_j)}{2} \quad (7)$$

The more obvious the block structure of the similar matrix, the closer the similarity matrix is to the ideal clustering result, and the more similar the similarity measure function can represent the clustering distribution of the original data. By comparing the values of the similarity function described above, EPC assigns the data point to the most similar center point, which is the most suitable cluster.

```

1 Call Procedure 1
2 k ← 0
3 for each  $x_i \in D^p$  do
4   if  $x_i$  labeled then
5     continue
6   take the  $x_i$  as the centre of the cluster, i.e.  $\text{centre}_k = x_i$ 
7   k+ +
8   for each  $x_j \in D^p$  do
9     while  $x_j \neq \text{centre}_k$  do
10      if  $x_j$  is unlabeled then
11        if  $\frac{\text{lesser}(\text{centre}_k, x_j)}{\text{greater}(\text{centre}_k, x_j)} < \zeta$  then
12           $\text{centre}_k$  and  $x_j$  is dissimilar;
13        else
14          add the  $x_j$  to the cluster $_k$ 
15          assign the label of cluster $_k$  to  $x_j$ 
16        else
17          if  $\text{sim}(\text{centre}_k, x_j) \geq \text{sim}(\text{centre}_{\text{original}}, x_j)$  then
18            remove  $x_j$  from the original cluster
19            add  $x_j$  to cluster $_k$ 
20 Call Procedure 2

```

In order to verify the validity of the algorithm, a manual data set is selected, which has 150 data items, was divided into 3 categories. The first 4 attributes are numeric in 7 attributes, and the last three

attributes are subtypes. There is a class tag attribute. The results obtained by using the thermodynamic data for clustering algorithm evaluation are shown in Tab. 3.

Table 3 Clustering Results of EPC for Tagged Receptor Point Data Sets

ζ	Number of points in the cluster				Number of clusters
[0,0.40]	3200				1
0.5	1346	1054	800		3
[0.53,0.70]	800	800	800	800	4
0.75	377 243 180	471 329	633 167	800	8
0.8	167 124 151 89 89 57 82 41	137 160 86 138 41 51 75 66 46	154 71 46 111 117 50 46 105 58 42	227 128 99 64 39 96 72 51 24	36

As can be seen from Tab. 3, when the threshold is 0.5, the data set is divided into 3 clusters; when the threshold is between 0.53 and 0.7, the data set is divided into 4 clusters. The number of data points is 800 in each cluster, and the class labels of the points in each cluster are the same, and the labels of the points in different clusters are different. In other words, each point is placed in its corresponding class. This phenomenon just illustrates the essential feature of the EPC algorithm, that is, it divides the data set according to the degree of similarity between data points. Another feature of the EPC algorithm has also been verified in detail in this experiment, that is, the higher the threshold, the more the number of clusters, the more similar the points within the cluster. Moreover, in this experiment, the clustering results are stable in 100 experiments, and the k-prototypes algorithm is random and unstable, showing the stability and accuracy of the experimental thermodynamic values.

4 Conclusion

In this paper, the process conditions for the synthesis of ferulic acid monoglyceride were studied, and its solubility under different conditions was studied. The process conditions were optimized by single factor experiment and orthogonal experiment. There was a large excess of glycerol during the experiment, so there was no significant increase in the rate of esterification when the amount of glycerol was increased. The glycerol content is therefore combined with the reaction kinetic constant. Data processing with Matlab gives the relationship between the esterification reaction rate and the reactant concentration at the initial stage of the reaction: $R_g = K [C_{\text{Ferulic acid}}]^{1.802}$, where K is a constant. From the reaction rate equation, the total reaction order of the reaction in the initial stage of the reaction was 1.802. The experimental activation energy of the esterification reaction was 69,907.44 J·mol⁻¹. Finally, the spectral clustering algorithm of mixed data is used to find the optimal clustering solution for the thermodynamic data through graph division. After experimental analysis, the rationality and effectiveness of thermodynamic data are proved.

Acknowledgements

This work was supported by Natural Science Foundation of Hubei Province of China (Grant No.2018CFB681).

References

- [1] Jin L , Hu H , Sun B , et al. Three-point bending fatigue behavior of 3D angle-interlock woven composite[J]. Journal of Composite Materials, 2012, 46(8), pp.883-894.
- [2] Yan L . Sorption of humic acid to layered double hydroxides prepared through ion thermal method [J]. DESALINATION AND WATER TREATMENT, 2017, 93, pp.109-119.
- [3] Zhang Xiaotao, Wang Jiqiang, Wang Wei, et al. Elastic-plastic-creep response of multilayered systems under cyclic thermo-mechanical loadings [J]. JOURNAL OF MECHANICAL SCIENCE AND TECHNOLOGY, 2018,32(3), pp. 1227-1234.
- [4] Hong Meiling, Li Jie, Zhang Wenfeng, et al. Semimetallic 1T' WTe₂ Nanorods as Anode Material for the Sodium Ion Battery [J]. ENERGY & FUELS, 2018, 32(5), pp. 6371-6377.
- [5] Zou Xuejun, Dong Yuying, Li Sijia, et al. Facile anion exchange to construct uniform AgX (X = Cl, Br, I)/Ag₂CrO₄ NR hybrids for efficient visible light driven photocatalytic activity [J]. SOLAR ENERGY, 2018, 169, pp.392-400.
- [6] Chen Shufang, Xu Zejun, Zhang Daohong, et al. Synthesis and application of epoxy-ended hyperbranched polymers [J]. CHEMICAL ENGINEERING JOURNAL, 2018,343,pp. 283-302.
- [7] Wen Zhipan, Ke Jun, Xu Jinlei, et al. One-step facile hydrothermal synthesis of flowerlike Ce/Fe bimetallic oxides for efficient As(V) and Cr(VI) remediation: Performance and mechanism [J]. CHEMICAL ENGINEERING JOURNAL, 2018, 343, pp.416-426.
- [8] Miao Ling, Zhu Dazhang, Liu Mingxian, et al. Cooking carbon with protic salt: Nitrogen and sulfur self-doped porous carbon nanosheets for supercapacitors [J]. CHEMICAL ENGINEERING JOURNAL, 2018, 347, pp.223-242.
- [9] Hong Hanyu, Shi Yu. Fast Deconvolution for Motion Blur Along the Blurring Paths [J]. CANADIAN JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING-REVUE CANADIENNE DE GENIE ELECTRIQUE ET INFORMATIQUE, 2018, 40(4), pp.226-274.
- [10] Zhang Xiaotao, Chen Haofeng, Ma Zhiyuan, et al. Shakedown boundaries of multilayered thermal barrier systems considering interface imperfections[J]. INTERNATIONAL JOURNAL OF MECHANICAL SCIENCES, 2018,144,pp.33-40
- [11] Gao Zhisheng, Xie Shenglong, Zhang bo, et al. Ultrathin Mg-Al layered double hydroxide prepared by ionothermal synthesis in a deep eutectic solvent for highly effective boron removal[J]. CHEMICAL ENGINEERING JOURNAL,2017,319,pp.108-118
- [12] **Chen Fengxi, Xie Shenglong, Huang XL, et al. Ionothermal synthesis of Fe₃O₄ magnetic nanoparticles as efficient heterogeneous Fenton-like catalysts for degradation of organic pollutants with H₂O₂ [J]. JOURNAL OF HAZARDOUS MATERIALS, 2017, 322(Si), pp.152-162**
- [13] **Xiang L, Li Z, Cheng J. Construction asphalt prepared by chemical treatment of deoiled asphalt [J]. PETROLEUM SCIENCE AND TECHNOLOGY, 2016, 34(10), pp.920-926**
- [14] Dong Shuhong, Zhou Jianqiu, Hui David. A quantitative understanding on the mechanical behaviors of carbon nanotube reinforced nano/ultrafine-grained composites [J]. INTERNATIONAL JOURNAL OF MECHANICAL SCIENCES, 2015, 101, pp.29-37