Summer Research 2020

Predict Solubility

(Dated: May 11, 2020)

This document outlines the problem and the steps we will be taking to tackle it. It will be updated as we go along during this summer.

I. THE BACKGROUND

The Free Energy of Mixing

We will be working to try to predict the solubility of a substance in a liquid. The starting point is a free energy of fluid mixtures (derived by me based on statistical field theory – it is a 33-page derivation, so I won't bore you with the details here). The expression of the free energy is

$$\beta f_{\text{mix}} = \frac{1}{2}\sum_S \frac{\phi_S}{v_S}\left[\frac{y_S}{y}\ln(1+y) - \ln(1+y_S)\right] + \sum_S \frac{\phi_S}{v_S}\ln\phi_S \tag{1}$$

where

- f_{mix} is the free energy per volume
- $\beta = \frac{1}{k_B T}$ is the inverse temperature scaled by the Boltzmann constant k_B
- ϕ_S is the volume fraction of component S
- v_S is the volume of a molecule of component S

and the y and y_S are dimensionless measures of the effective dipole moments of the molecules, given by

$$y_S = \frac{\beta \bar{\mu}_S^2}{3\varepsilon_0 v_S} \tag{2}$$

with $\bar{\mu}$ being the dipole moment of a molecule of the component S, and

$$y = \sum_S \phi_S y_S \tag{3}$$

Looking at the equation itself probably won't make much sense. So let's get some perspectives by plugging in numbers.

Let's first consider a binary mixture of trichloromethane and water. We will label trichloromethane as component A and water as component B. Because this is a mixture of just two components, we have $\phi_A + \phi_B = 1$ – so we can plot the free energy as a function of ϕ_A. We have parameters $v_A = 134.05$ Å^3, $v_B = 30.0$ Å^3 (you may get these based on

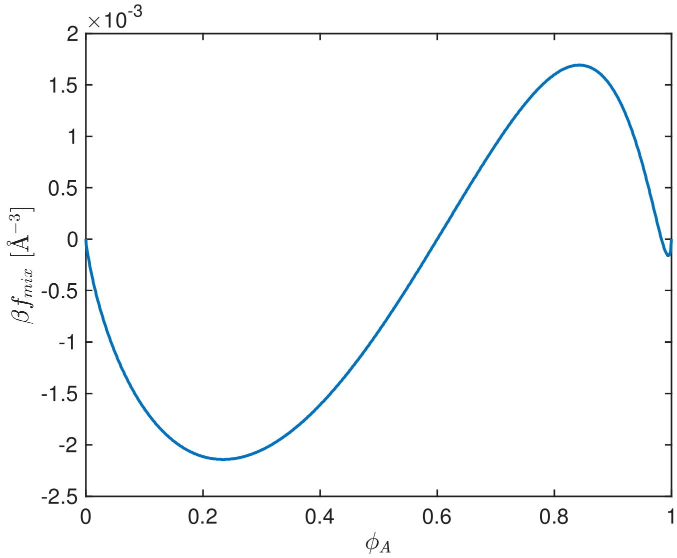

FIG. 1. βf_{mix} vs. ϕ_A for trichloromethane(A)-water(B) system

their molar volumes), $y_A = 2.84$, and $y_B = 29.8$ (these can be calculated from the dielectric constant of the liquid – more on what this is later). For this, the plot of βf_{mix} vs. ϕ_A is shown in Figure 1.

Let us also consider a different system with A being methanol and B being water. For this system, $v_A = 67.23$ Å^3, $v_B = 30.0$ Å^3, $y_A = 14.27$, and $y_B = 29.8$. For this, the plot of βf_{mix} vs. ϕ_A is shown in Figure 2.

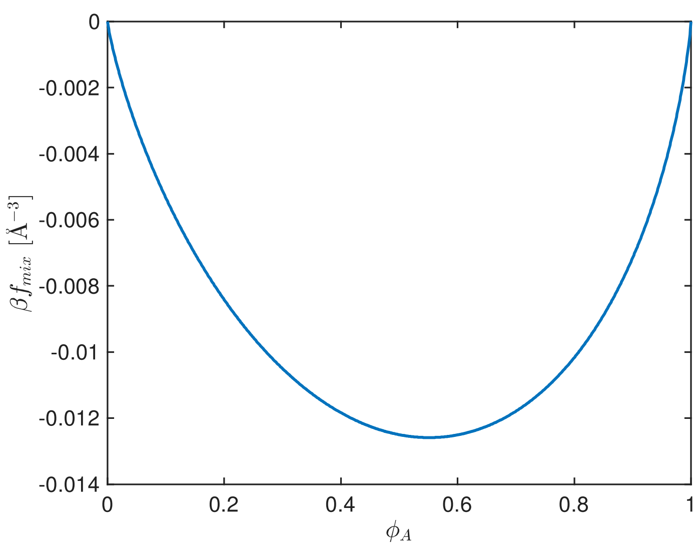

FIG. 2. βf_{mix} vs. ϕ_A for methanol(A)-water(B) system

Notice that, for two immiscible liquids as in the case of trichloromethane and water, a concave region exists in the plot βf_{mix} vs. ϕ_A, and for two miscible liquids as in the case of methanol and water, the plot βf_{mix} vs. ϕ_A is convex everywhere. This gives us a way to predict whether two liquids are miscible with each other.

The dipole moment and the dielectric constant

The dipole moment on a molecule is defined as

$$\boldsymbol{\mu} = \sum_i q_i \mathbf{d}_i \tag{4}$$

where q_i is the partial charge on the ith atom, and $\mathbf{d}_i$ is the displacement vector of the ith atom.

Because of the molecular dipole moments, substances have a property called the dielectric constant, ε, which is its ability to screen electrostatic interactions. That is, while the energy of interaction for two charges q_1 and q_2 separated by a distance r in vacuum is given by the Coulomb Law as $\frac{q_1 q_2}{4\pi\varepsilon_0 r}$, when the two charges are in water, the energy of interaction is $\frac{q_1 q_2}{4\pi\varepsilon_0 \varepsilon r}$ (The interaction is screened by a factor of ε). The dielectric constant can be measured experimentally. For water, the value is approximately 80.

It is non trivial to relate the dielectric constant of a liquid to the dipole moment of the molecules. One of the possible expression is (based on the same 33-page derivation that I have):

$$\frac{(\varepsilon - 1)(2\varepsilon + 1)}{\varepsilon} = 3y \left(\frac{2y^2 + 3y + 9}{(y+3)^2} \right) \tag{5}$$

II. THE PROBLEM STATEMENT

With the help of the theory, can we better predict the solubility of small molecules?

III. THE PLAN FOR THIS SUMMER

Learning Machine Learning (Week 1–10)

It is probably the best that you learn machine learning from the experts. One of the best place to learn this is a Cousera course by Andrew Ng at https://www.coursera.org/learn/machine-learning. Spend some time on this each week (perhaps one day a week), and I am hoping that you may get through Week 1–7 of the course during this 10 weeks of summer research. It is also important that you spend some time to work on the assignments in the course.