Efficient ant colony optimization for computer aided molecular design: Case study solvent selection problem

Berhane H. Gebreslassie, Urmila M. Diwekar*

Center for Uncertain Systems: Tools for Optimization & Management (CUSTOM), Vishwamitra Research Institute, Crystal Lake, IL 60012 United States

ARTICLE INFO

ABSTRACT

In this paper, we propose a novel computer-aided molecular design (CAMD) methodology for the design of optimal solvents based on an efficient ant colony optimization (EACO) algorithm. The molecular design problem is formulated as a mixed integer nonlinear programming (MINLP) model in which a solvent performance measure is maximized (solute distribution coefficient) subject to structural feasibility, property, and process constraints. In developing the EACO algorithm, the better uniformity property of Hammersley sequence sampling (HSS) is exploited. The capabilities of the proposed methodology are illustrated using a real world case study for the design of an optimal solvent for extraction of acetic acid from waste process stream using liquid–liquid extraction. The UNIFAC model based on the infinite dilution activity coefficient is used to estimate the mixture properties. New solvents with better targeted properties are proposed.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Solvents are used for a variety of purposes in process industries. They are extensively used as process materials, as extracting agents, and as process liquids in process industries, pharmaceutical industries, and solvent based industries. Waste solvents are main source of pollution to air, water, and soil. Therefore, it is empirical to use environmentally benign solvents without compromising the process performance. Moreover, there are some solvents that must be eliminated because of environmental and health effects and regulatory requirements (Karunanithi et al., 2005; Kim and Diwekar, 2002b,c; Xu and Diwekar, 2005). The environmental awareness, the strict legislation and the need for high performance solvents have resulted in the search for less toxic and environmentally benign solvents and solvent formulations that have improved performance characteristics. Several methodologies have been developed for solvent selection over the years. The first approach uses traditional laboratory synthesis and test methodology to find promising solvents. This method can provide reliable and accurate results, but in many cases this approach is limited by cost, safety, and time constraints. The second approach is to screen the property database. Though, the screening of the database is the most common and

simple method, it is limited by size and accuracy of the database. These methods are usually expensive and time-consuming.

Solvent selection based on computer aided molecular design (CAMD) is fast emerging systematic tool for efficient and reliable design of candidate solvents from their fundamental building blocks (Marrero and Gani, 2001; Karunanithi et al., 2005; Kim and Diwekar, 2002b,c; Xu and Diwekar, 2005). Beyond the solvent selection, the CAMD technique is practiced with great success in different disciplines such as pharmaceutical process designs (Gernaey and Gani, 2010), polymer design (Satyanarayana et al., 2009), and bioethanol production (Alvarado-Morales et al., 2009). CAMD is generating large number of structural molecules with desired properties from a small set of structural groups (building blocks). CAMD is the reverse use of the group contribution method. Different solution strategies are implemented to solve CAMD techniques: heuristic numeration (Hostrup et al., 1999; Li et al., 2002), knowledge based technique (Harper and Gani, 2000; Yamamoto and Tochigi, 2008), molecular property clusters with algebraic equations (Chemmangattuvalappil et al., 2009; Eljack and Eden, 2008; Kazantzi et al., 2007) and optimization-based methods (Karunanithi et al., 2005; Samudra and Sahinidis, 2013; Diwekar and Shastri, 2011; Ostrovsky et al., 2002).

In the optimization approaches, because of the nonlinearity behavior of the UNIFAC model, the CAMD for solvent selection is formulated as a mixed integer nonlinear programming problem (MINLP) that seeks to optimize the desired properties of the solvent molecules subject to molecular design feasibility rules. To solve the MINLP formulation of the CAMD problems, different optimization
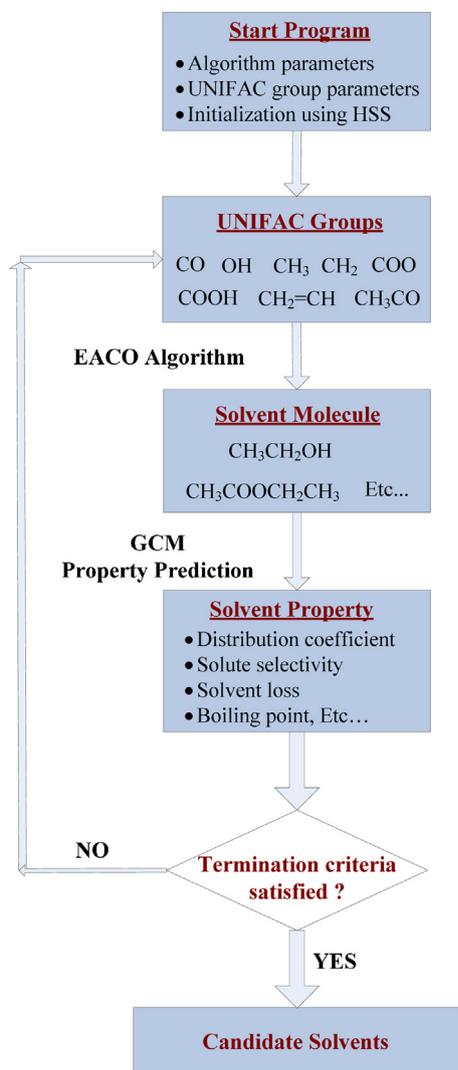
Fig. 1. A basic algorithm for solvent selection using EACO.

methods have been proposed: decomposition methods that use local optimizers for the NLP sub-problem (Odele and Macchietto, 1993; Harper et al., 1999; Karunanithi et al., 2005, 2006), global optimization (Ostrovsky et al., 2002; Samudra and Sahinidis, 2013), interval analysis (Achenie and Sinha, 2003), and dynamic optimization (Giovanoglou et al., 2003). Recently, probabilistic methods such as simulated annealing (Kim and Diwekar, 2002a,b,c), genetic algorithms (Cheng and Wang, 2008; Diwekar and Xu, 2005; Xu and Diwekar, 2005) are adopted as an alternative to the local optimization strategies to find better solutions.

Ant colony optimization (ACO) algorithm proposed by (Dorigo, 1992) is a metaheuristic optimization strategy that can provide a viable alternative to solve the MINLP of the CAMD problem. Although in recent years, there has been a significant research interest in developing and implementing for different applications, to the authors knowledge, there is no work in the literature that implement efficient ant colony (EACO) algorithm to solve the MINLP CAMD problems.

The algorithm implemented in this work to solve the solvent selection problem combines the CAMD and the EACO algorithm proposed by (Gebreslassie and Diwekar, 2015) as shown in Fig. 1. The algorithm parameters, the UNIFAC building block groups and their properties such as the volume and surface area parameters, and the interaction parameters between groups are first introduced. These UNIFAC groups are uniquely designed to

generate all possible molecules by exploring all possible combinations. Using the group combinations from this set of groups, solvent molecules are generated. For example, as shown in the algorithm, ethanol (CH_3CH_2OH) is generated from the CH_3, CH_2, and OH groups. The number of combinations can be reduced by introducing constraints from physical and chemical properties, structural feasibility as well as those from regulatory restrictions. Once molecules are generated, the properties of the molecules are predicted based on the properties of their groups using the UNIFAC model. This method can generate candidate solvents with a reasonable accuracy. The rest of the paper is organized as follows. Section 2 describes the CAMD model formulation. The optimization problem of the solvent selection under study is formulated as an MINLP model in Section 3. Ant colony optimization and the proposed EACO algorithm to solve the MINLP optimization problem is presented in Section 4 followed by Section 5 that discusses the result. Finally, Section 6 presents the concluding remarks of the paper.

2. Solvent selection model formulation

To replace the current solvent or formulate a new one, there are several criteria and solvent properties that can be reviewed such as the solute distribution coefficient m, solvent selectivity β, solvent loss S_L, and physical properties like boiling point, density, viscosity, and so on.

2.1. Distribution coefficient (m)

It is a measure of solvent capacity and it represents the solute distribution between the solvent and the raffinate phases. A high value of m reduces the size of extraction equipment and the amount of recycling solvent. The solute distribution coefficient is estimated as shown in Eq. (1).

$$m = \frac{\text{concentration of solute in extractive phase}}{\text{concentration of solute in raffinate phase}} \cong \frac{\gamma_{BA}^\infty}{\gamma_{BS}^\infty} \frac{MW_A}{MW_S} \quad (1)$$

where A, B, and S represent the raffinate (a nonpolluting molecule e.g. water), solute (polluting molecule e.g. acetic acid), and solvent (e.g. ethanol) phases, respectively. MW denotes the molecular weight, and γ^∞ is the infinite dilution activity coefficient.

2.2. Solvent selectivity (β)

It is the ratio between the solute distribution coefficient and raffinate. It estimates the ability of the solvent to selectively dissolve a solute (polluting molecule). A high solvent selectivity value thus can reduce the cost of solute recovery and it is defined as shown in Eq. (2).

$$\beta = \frac{\text{distribution coefficient of solute}}{\text{distribution coefficient of solvent}} = \frac{m_B}{m_A} \cong \frac{\gamma_{AS}^\infty}{\gamma_{BS}^\infty} \frac{MW_B}{MW_A} \quad (2)$$

2.3. Solvent loss (S_L)

It is a measure of the concentration of solvent in raffinate phase. It is the measure of the solvent loss tendency; a low solvent loss value means high selectivity toward solute and determines the immiscibility between the solvent and the raffinate. It is defined as shown in Eq. (3).

$$S_L = \text{concentration of solvent in raffinate phase} \cong \frac{1}{\gamma_{SA}^\infty} \frac{MW_S}{MW_A} \quad (3)$$

In Eqs. (1)–(3), the distribution coefficient, the solvent selectivity and the solvent loss properties are given as a function of the infinite dilution activity coefficients (γ^∞). It shows the non-ideality of the mixtures (A–B, A–S, and B–S). If the mixture is ideal, γ^∞ is