L-lysine Neuro-Dynamic Optimal Control

Dr. Tatiana Ilkova[1] and Dr. Mitko Petrov[2]

[1] Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences

Received: 27 October 2011 Accepted: 20 November 2011 Published: 4 December 2011

Abstract

In this paper Neuro-dynamic programming (NDP) is proposed as an alternative to alleviate the ?curse of dimensionality? of the Dynamic programming (DP) for optimal control of a fed-batch fermentation process in the L-lysine production. The most effective and cheapest method for the Llysine biosynthesis (in biological active form) is the microbiological method via a direct fermentation. In this paper an optimization method of the L-lysine production from strain Brevibacterium flavum 22LD is used and that is NDP. The results show that the quality of L-lysine enhances at the end of the process. The proposed method is particularly simple to implement and can be applied for on-line optimization.

Index terms— dynamic programming, neural network, Llysine fermentation, optimal control

1 INTRODUCTION

-lysine is an essential amino acid, which means that it is essential to human health, but cannot be produced by the body. For this reason L-lysine must be obtained from food. Amino acids are the building blocks of the protein. Lysine is important for proper growth and it plays an essential role in the production of carnitine, which is a nutrient responsible for converting fatty acids into energy and helping to lower cholesterol.

The insufficient L-lysine quantity in the fodders reduces the biological value of the fodder doses, it also reduces the weight increase and the further productiveness of the agricultural animals, decreases the fodder quality, used for a kilogram growth and decreases the product quantity from animal origin. Llysine is also used in the food industry for farming, in the medicine as a component of the infusion solution (blood substitutes) and as generally strengthening patent medicines. Lysine appears to help the body absorb and conserve calcium and it plays an important role in the formation of collagen, a substance which is important for the bones and connective tissues including skin, tendon, and cartilage (Anastassiadis, 2007).

Amino acids are the basic bioelements of proteins, which are the most important macromolecules for the functions of humans and animals. Out of the 20 L-amino acids, which are found worldwide in most of the living organisms, L-lysine is one of the nine essential amino acids for human and animal nutrition (Anastassiadis, 2007).

Neuro-dynamic programming (NDP) is proposed as an alternative to alleviate the "curse of dimensionality" of the Dynamic programming (DP). The term NDP expresses the reliance of the methods, described in this article with respect to both the DP and the neural network concepts. The term reinforcement learning is also used in the artificial intelligence community where the methods originated from. Using common artificial intelligence terms, the methods help the systems "learn how to make good decisions by observing their own behavior and use built-in mechanisms for improving their actions through a reinforcement mechanism" (Bertsekas & Tsitsiklis, 1996).

The key idea is to use a scoring function to select decisions in complex dynamic systems, arising from a broad variety of applications for engineering design, operations research, resource allocation, finance, etc. This is much similar to a computer chess, where positions are evaluated by means of a scoring function and the move that leads to the position with the best score is chosen. NDP provides a class of systematic methods for computing the appropriate scoring functions using approximation schemes and simulation/evaluation of the system's performance (Driessens & Dzeroski, 2004).

Using common artificial intelligence terms, the methods allow the systems to "learn how to make good decisions by observing their own behavior and use builtin mechanisms for improving their actions through a reinforcement mechanism".

In more mathematical meaning "observing their own behavior" relates to simulation and "improving their actions through a reinforcement mechanism" relates to the iterative schemes for improving the quality of approximation of the optimal cost function, the Q-factors or the optimal policy. There has been a gradual realization that the reinforcement learning techniques can be fruitfully motivated and interpreted in terms of classical DP concepts such as the value and policy iteration **??**Barto et. all, 1995;Sutton, 1988).

NDP is a relatively new class of the dynamic programming methods for control and sequential decision making under uncertainty. These methods have the potential of dealing with some problems that were thought to be intractable for a long time due to either a large state space or the lack of an accurate model. They combine ideas from the fields of neural networks, artificial intelligence, cognitive science, simulation, and approximation theory. In recent years L the method has been applied successfully for an optimal control of fermentation process (FP). The literature sources show that the calculating time is significantly reduced, while the desired products quantity is increased **??**Kaisare et. all, 2003; **??**kova & Petrov, 2008).

The aim of this study is to develop optimal feed rate strategy of biotechnological process in L-lysine production using Neuro-dynamic control.

2 II.

3 PROCESS SPECIFICS AND L-LYSINE PRODUCTION MATHEMATICAL MODEL

The development of a multi-step biotechnological process requires three steps, comprising of:

? Identification and characterization of a suitable biological system (microorganism, biocatalyst). (Anastassiadis, 2007).

In addition to parameters like pH, agitation and aeration rate, air saturation, temperature, dissolved carbon dioxide and foaming, the medium composition is a very important factor highly influencing fermentation processes, which are often a subject of extensive process development and optimization studies. The culture medium has to satisfy the requirements of microbial growth and production in a suitable manner. L-lysine can be produced either using a chemical or a biochemical method, which is economic, even though relatively low yields are obtained during the extraction of L-lysine, requiring specific installations and the use of expensive products. The stereospecificity of amino acids and the steadily increasing L-lysine demand necessitates indispensably their fermentative production (the L isomer) over synthetic processes.

The experimental investigations are done in a 15 L bioreactor that is included in an Automatic Control System. The Automatic Control System is flexible and includes control of the following parameters of the process: rotation speed, oxygen partial pressure, temperature, pH, foam level, gas flow rate, flow rates of the main substance. The process is led in the next conditions:? Temperature T=300C; ? pH pH=6.8-7.6; ? pO 2 pO 2 =20-30%; ? Gas flow rate Q G =60 L h-1; ? Rotation speed n=450 min-1; ? Maximum bioreactor volume 15 L.

For the L-lysine fermentation defined media is used which acquires nutrients that require pure growth and essential additives or alternatively undefined media containing natural organic substances such as soybeanhydrolyzate, corn steep liquor, yeast extract or peptone is used. Common fermentation media for Llysine production contain various carbon and nitrogen sources, inorganic ions and trace elements (Fe++, Mn++), amino acids, vitamins (biotin, thiamine-HCl, Nicothin amide) and numerous complex organic compounds. An upper expression of genes is also achieved by optimizing the composition of the media and the culture technique in addition to the physiological and genetic parameters (Anastassiadis, 2007).

The model of the fed-batch processes includes the dependences between the concentrations of the basic variables of the process: cell mass concentration (bacteria Brevibacterium flavum), substrate concentration, L-lysine, Threonine concentration and oxygen concentration in the liquid phase. The general scheme of the L-lysine is shown in Figure 1. The mathematical model of the process is based on the mass balance equations as a perfect mixing in the bioreactor is adopted. The model of the process has the following type (Anastassiadis, 2007):(1) (2) X V F X dt dX ? = µ X k X k X k S S V F dt dS in ? µ 7 6 5) (? ? ? ? = Tr V F X k Tr Tr V F dt dTr in ? ? ? = µ 13) (L L l L C V F X k X k X k C C a k dt dC ? ? ? ? ? = ? µ 16 15 14 *) ((3) (4) (5) (6)

The specific rate of L-lysine synthesis and specific consumption rate have the following form: (7) (8) where: specific rate of L-lysine synthesis, h specific consumption rate of L-lysine, h

The initial conditions in the model (**??**) -(**??**) have the follows values:

The model coefficients in (1) -(**??**) have the following values:

4 III. NEURO-DYNAMIC OPTIMAL CONTROL OF THE PROCESS

The objective of this work is to find the optimal feed flow rate (F(t)) of a fed-batch process, such as the L-lysine production that will raise L-lysine at the end of the process, i.e.: (9) where: t 0 -initial time, t f -final time of the fermentation. Therefore, the control objective is to drive the reactor from the low product steady state to the desirable high product rate. It may be considered as a step change in the set point at time t=0 from the low product concentration to the high product concentration steady state.

In the systems the decisions are made in stages. The outcome of each decision is not fully predictable but can be anticipated to some extent before the next decision is made. Each decision results in some immediate cost, but it also affects the context in which the future decisions are to be made and thus it affects the cost incurred in future stages. DP provides a mathematical formalization of the tradeoff between the immediate and future costs. Generally, in DP formulations there is a discrete-time dynamic system whose state evolves according to the given transition probabilities that depend on the decision/control u.

DP is an elegant way to solve the introduced optimization problem (9). It involves a stagewise calculation of the cost-to-go function to arrive at the solution not just for a specific initial state, but for a general initial state. Once obtained the cost-to-go function, represents a convenient vehicle to obtain the solution for a general state. In very few cases the stagewise optimization to obtain analytically a closedform expression for the cost-to-go function has been solved. The conventional approach to the problem involves gridding the state space, calculating and storing the cost-to-go for each grid points as one marches backward from the first stage to the last. For an infinite horizon problem the number of iteration required for convergence can be very big. Such an approach is seldom practically feasible due to the exponential growth of the computation with respect to the state dimension. Unfortunately, from the very beginning it was apparent that an increase of the dimensionality of the problem, i.e. an addition of reservoirs, caused an exponential increase in the time required to find a solution. This is referred to as the ”curse of dimensionality”, which must be removed so that this approach can find a widespread use.

NDP aims to develop a methodological foundation for combining dynamic programming, compact representations, and simulation to provide the basis for a rational approach to complex stochastic decision problems (Bertsekas & Tsitsiklis, 1996; **??**aisare et. all, 2003).

Two fundamental DP algorithms, policy iteration and value iteration, are the starting points for the NDP methodology. The most straightforward adaptation of the policy iteration method operates as follows: we start with a given policy (a rule for choosing a decision u at each possible state i), and we approximately evaluate the cost of that policy (as a function of the current state) by least-squares-fitting a scoring function to the resultsX dt dL ? = F dt dV = ( )( )( ) [ ] L L C k S S k Tr k C Tr k + ? + + = 4 0 3 2 1 μ ( )( )( )( ) [ ] L L L C k C k S k S k C S k + + + + = 12 11 10 9 8 ? μ - -1 ; ? - -1 ; X -

biomass concentration, g l -; L -L-lysine concentration, g l -1 ; S -glucose concentration, g l -1 ; V -working liquid volume, l; F -feed flow rate, l h -1 ; Tr -Threonine concentration, mg l -1 ; t -process time, h; C L -dissolved oxygen concentration, g l -1 ; C* -equilibrium dissolved oxygen concentration, g l -1 ; S in -input feed substrate concentration, g l -1 ; Tr in -input feed Threonine concentration, g l -1 ; k l a -volumetric liquid mass transfer coefficient, h -1 . X(0) = X 0 = 3.00 g l -1 ; S(0) = S 0 = S i = 100.00 g l -1 ; Tr(0) = Tr 0 = Tr in = 100.00 mg l -1 ; L(0) = 0.00 g l -1 ; C L (0) = C * = C 0 = 6.1x10 -3 g l -1 ; V(0) = V 0 = 10.00 l. k 1 = 20.8, k 2 = 42.0, k 3 = 28.0, k 4 = 1.1, k 5 = 1.01, k 6 = 0.07, k 7 = 0.51, k 8 = 62.0, k 9 = 28.0, k 10 = 37.0, k 11 = 4.0, k 12 = 0.12, k 13 = 6.10, k 14 = 448.0, k 15 = 22.0, k 16 = 209.0, k l a = 120.? = f t t dt t V t L Q 0) ( ) (max

u of many simulated system trajectories using that policy. A new policy is then defined by minimization in Bellman’s equation where the optimal cost is replaced by the calculated scoring function and the process is repeated. This type of algorithm typically generates a sequence of policies that eventually oscillates in a surrounding of an optimal policy. The resulting deviation from optimality depends on a variety of factors, principal among which is the ability of the architecture of scoring function to accurately approximate the cost functions of the various policies.

NDP uses simulated process data received under suboptimal policies to fit an approximate cost-togo function -generally by fitting artificial network. With the value iteration approach NDP the initial approximate cost-to-go function in the future was improved by an iteration procedure based on Bellman equation. In this way the simulation role has two points. First, by simulation the process under a reasonably chosen suboptimal policy and all possible operating parameters it provides set data points that define the relevant ”working” region in the state space. Second, the simulation provides the cost-to-go value under the suboptimal policy for each state visited, which iteration of the Bellman equation can be initialed with **??**Kaisare et. all, 2003).

In this paper we will demonstrate NDP approach not only for reducing the computational demand, but also for improving the controller performance through the use of the cost-to-go approximator. A neural network is chosen as an approximator to obtain cost-to-go as a function of system states. While a properly trained neural network has good interpolation capabilities, one may not be used to extrapolate over the regions of state space that are not covered during its training. Extrapolation by neural network results in deteriorated performance of the controller.

The policy improvement theorem states that a new policy that is greedy (a greedy policy is one whose current cost is the least) with respect to the cost-to-go function of the original policy is as good as or better than the original policy.

When the new policy is as good as the original policy the above equation becomes the same as Bellman equation.

The relevant regions of the state space are identified by simulation of NDP control and the initial suboptimal cost-to-go function is calculated from the simulation data. In this survey a functional approximator is used to interpolate between this data. The improvement is obtained through the iteration of the Bellman equation. When the iteration converge this offline computed cost-to-go function can be used for an on-line optimal control calculation for the bioreactor (Xiong & Zhang, 2005).

NDP uses neural network approximations for the approximation of cost-to-go function. The cost-to-go function was not used to generate an explicit control law; instead, it was used in an on-line optimization to reduce the large (or infinite) horizon problem to a relatively short horizon problem. The method was found to be robust to approximation errors. Both deterministic (step changes in kinetic parameters) and stochastic problems (random variations in kinetic parameters and feed composition) were explored Lee J.M., & Lee, J. H., 2009; Tosukhowong & Lee J. H., 2009).

5 a) NDP algorithm

The following notations are used for description of the algorithm:

The general simulation-approximation scheme involves computation of the converged cost-to-go approximation off-line. The architecture of the scheme is shown in Figure **??** The simulation-based approach involves computation of the converged profit-to-go approximation off-line. The following steps describe the general procedure of NDP algorithm: 1. Performing of simulations of the process with chosen suboptimal policies under all representative operating conditions. Starting with a given policy (a rule for choosing a decision u at each possible state i), and approximately evaluate the cost of that policy (as a function of the current state) by least-squaresfitting a scoring function to the results of the many simulated system trajectories using that policy. 2. Calculation of the -horizon cost-to-go for each state visited during the simulation, using the simulation data. The solution of one-stage-ahead cost plus cost-to-go problem results in the improving of the cost values. Cost-to-go is the sum of the single state cost from the next point to the end of the horizon. 3. The deviation, which is a result of the optimality, depends on a variety of factors, principal among which is the ability of the architecture of Bellman function to approximate accurately the cost functions of the various policies. 4. A new policy is then defined by minimizing Bellman's equation where the optimal cost is replaced by the calculated scoring function and the process repeats. This type of algorithm typically generates a sequence of policies that eventually oscillate in a surrdounding of an optimal policy. 5. Fitting a neural network function approximator to the data to approximate the cost-to-go function as a smooth function of the states. 6. As described above the improved costs are again fitted to a neural network, to obtain subsequent ? iterations of Bellman functions, and so on ..., until convergence is accomplished. 7. Policy update may sometimes be necessary to increase the coverage of the state space. In this case more suboptimal simulations with the updated policy are used to increase the coverage or the number of the data points in certain region of the state space.

The NDP algorithm block-scheme is shown in Fig. **??**.

Fig. **??** : NDP algorithm block-scheme.

Take into consideration that when starting with a fairly good approximation of the cost-to-go (which has to be a result of using a good suboptimal policy), the cost iteration has to converge fairly fast -faster than the conventional stagewise cost-to-go calculation.

The next values of F are examined 0.2, 0.4, 0.5, 0.7, that can cover the possible rang of variations. The bioreactor was started at three different W(0) values for each of the parameter values around the low product yield steady state.

A functional approximation relating cost-to-go with augmented state was obtained by the neural network -with five hidden nodes, six input nodes and two output nodes. The neural network presented a good fit with a mean error of 10-3 after training for 1000 epoch.

Improvement of the cost-to-go is obtained through the iterations of the Bellman equation. This method is known as a value iteration (or value iteration). The solution of the one-stage-ahead cost plus cost-togo problem, results in the improvement of the cost values. The improved prices were again fitted to the neural network, described above to obtain subsequent iterations of Belman function and so on ..., until they are converged. Cost is said to be "converged" if the sum of the absolute error is less than 5% of the maximum cost. The cost is converged in 7 iterations for our system.

The converged cost-to-go function from above was used for solving the one-stage-ahead problem.

The optimal value of feed flow rate before and after optimization is shown in Fig. 3. The L-lysine production before and after optimization is shown in Fig. 4. Fig. 4 shows the increase of the L-lysine after optimization by 39.41%. The results show that after 48th the process stands still and it continuing is economically disadvantageous. It becomes clear that after 48th the process goes into a steady state. Therefore, the fixed right end for 48 hours is appropriate.

In this optimization problem the time is discredited in six hours. It is assumed that this is a step of discretization of this process in terms of features and well-known computational difficulties.

IV.

6 CONCLUSION

An approach for the optimal control of fermentation processes for a L-lysine fed-batch fermentation is developed for searching an optimal feed rate strategy using Neuro-dynamic control. It is proposed as a method for alleviation of the "curse of dimensionally" of DP.

The conventional approach to solving an optimization problem with DP method involves gridding of the state space, solving the optimization for each grid point and performing the stagewise optimization until convergence. Exhaustive sampling of state space can be avoided by identifying relevant regions of the state space by simulation under judiciously chosen suboptimal policies, which is presented using NDP methods with the help of a neural network for functional approximator.

The results show that the L-lysine quantity is highly raised at the end of the process which is the desired criterion for process quality. The result shows that NDP is a convenient and easy to use application method for optimal control. The approach is particularly simple to implement and it should be used for on-line implementation, after necessary additional training of the relevant neural network is obtained. L-lysine before L-lysine after [1] [2] [3]

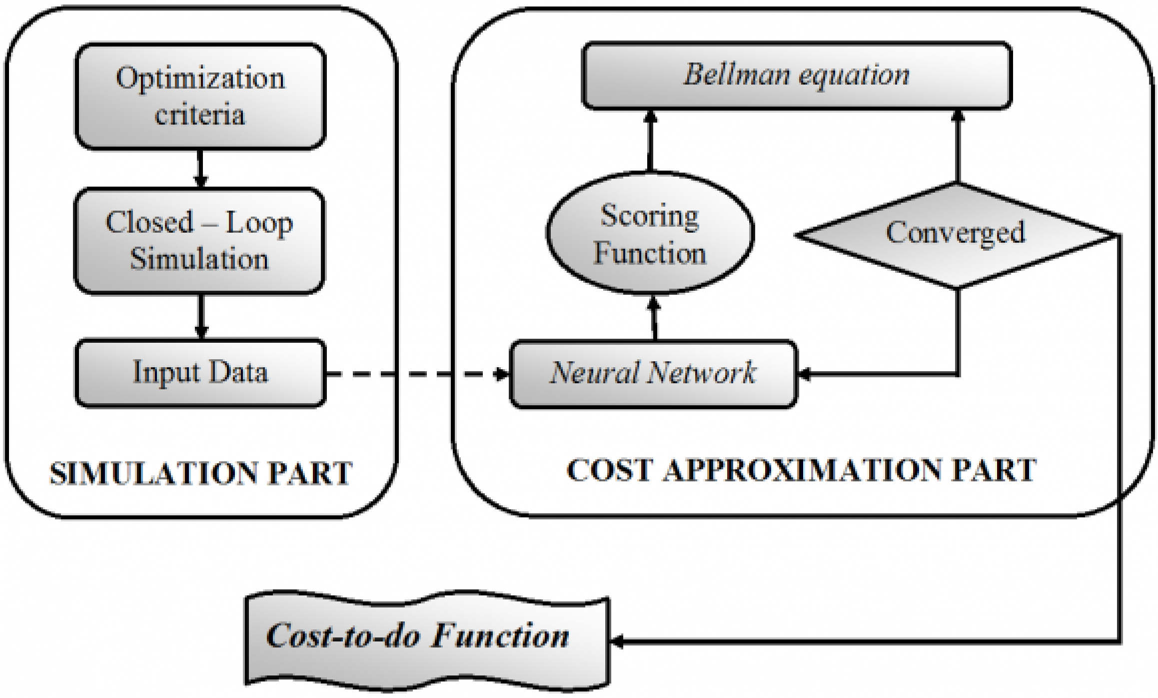
1

Figure 1: Fig. 1 :

[1]© 2011 Global Journals Inc. (US) L-lysine Neuro-Dynamic Optimal Control

[2]December

[3]DecemberVolume XI Issue IV Version I



1

Figure 2: . Step 1 ,

Figure 3:

[Soni ()] *A multi-scale approach to fed-batch bioreactor control*, A Soni . 2002. Pennsylvania: University of Pittsburgh Press.

[Lee et al. ()] 'Choice of approximator and design of penalty function for an approximate dynamic programming based control approach'. J M Lee , N S Kaisare , J H Lee . *Journal of Process Control* 2006. 16 p. .

[Ilkova and Petrov ()] *Dynamic and neurodynamic optimization of a fed-batch fermentation process. Lecture Notes in Artifical Intelligence*, T Ilkova , M Petrov . 2008. 5253 p. .

[Vlachos et al. ()] 'Hierarchical Multiscale model-based design of experiments, catalysts and reactors for fuel processing'. D G Vlachos , A B Mhadeshwar , N S Kaisare . *Computer & Chemical Engineering* 2006. 30 p. .

[Driessens and Dzeroski ()] 'Integrating guidance intro relational reinforcement learning'. K Driessens , S Dzeroski . *Machine Learning*, 2004. 57 p. .

[Anastassiadis ()] 'L-Lysine fermentation'. S Anastassiadis . *Recent patents on biotechnology* 2007. 1 p. .

[Sutton ()] 'Learning to predict by the methods of temporal differences'. R S Sutton . *Machine Learning*, 1988. 3 p. .

[Xiong and Zhang ()] 'Neural Network model-based on-line re-optimisation control of fedbatch processes using a modified iterative dynamic programming algorithm'. Zh Xiong , J Zhang . *Chemical Engineering and Processing* 2005. 44 p. .

[Bertsekas and Tsitsiklis ()] *Neuro-dynamic programming*, D Bertsekas , J Tsitsiklis . 1996. Massachusetts: Athena Scientific Belmont.

[Barto et al. ()] 'Real-time learning and control using asynchronous dynamic programming'. A G Barto , S J &bradtke , S P Singh . *Artificial Intelligence* 1995. 72 p. .

[Kaisare et al. ()] 'Simulation based strategy for nonlinear optimal control: application to a microbial cell reactor'. N S Kaisare , J M Lee , J H Lee . *International Journal of Robust and Nonlinear Control* 2003. 13 p. .