Applying Multi-Agent Techniques to Cancer Modeling

Matthew Brown[1], Emma Bowring[2], Shira Epstein[1], Mufaddal Jhaveri[1], Rajiv Maheswaran[1], Parag Mallick[3], Shannon Mumenthaler[3], Michelle Povinelli[4], and Milind Tambe[1]

[1]Computer Science Department, University of Southern California

[2]School of Engineering and Computer Science, University of the Pacific

[3]Center for Applied Molecular Medicine, University of Southern California

[4]Ming Hsieh Department of Electrical Engineering, University of Southern California

ABSTRACT

Each year, cancer is responsible for 13% of all deaths worldwide. In the United States, that percentage increases to 25%, translating to an estimated 569,490 deaths in 2010 [1]. Despite significant advances in the fight against cancer, these statistics make clear the need for additional research into new treatments. As such, there has been growing interest in the use of computer simulations as a tool to aid cancer researchers. We propose an innovative multi-agent approach in which healthy cells and cancerous cells are modeled as opposing teams of agents using a decentralized Markov decision process (DEC-MDP). We then describe changes made to traditional DEC-MDP algorithms in order to better handle the complexity and scale of our domain. We conclude by presenting and analyzing preliminary simulation results. This paper is intended to introduce the cancer modeling domain to the multi-agent community with the hope of fostering a discussion about the opportunities and challenges it presents. Given the complexity of the domain, we do not claim our approach to be a definitive solution but rather a first step toward the larger goal of creating realistic simulations of cancer.

Categories and Subject Descriptors

I.2.11 [**ARTIFICIAL INTELLIGENCE**]: Distributed Artificial Intelligence

General Terms

Algorithms

Keywords

Agent-based Simulations, Computational Cancer Modeling, Distributed MDPs

1. INTRODUCTION

The National Institutes of Health estimates the overall damage of cancer within the U.S. for 2010 at $263 billion, covering direct medical, indirect morbidity, and indirect mortality costs [2]. It is hoped that computer models of cancer, and the way in which it spreads, will allow for greater understanding of the complex cellular biology taking place as well as provide an efficient, low-cost method for evaluating the efficacy of various treatments. In order to model cancer convincingly, the challenges presented by the complexity as well as the scale of the domain must be addressed.

Among the previous approaches in this domain, which include evolutionary game-theoretic models [3] and cellular automata models [7], our work is most similar to the agent-based model found in [10]. In both cases, individual cells are modeled as agents. However, there is a fundamental difference in how the actions of agents are determined. In [10], the actions of agents are implicitly derived from a complex set of biophysically-inspired rules and differential equations. However, these rules and equations must be generated by hand and all computation must be done at execution-time.

In order to assess the potential of multi-agent techniques for the purposes of cancer modeling, we have framed the problem as a decentralized Markov decision process (DEC-MDP) [11]. Our pairing of domain and approach is innovative from the perspective of both the fields of cancer modeling and multi-agent systems. DEC-MDPs represent an entirely new approach to cancer modeling in which cells are viewed as autonomous agents working as a team. At the same time, cancer modeling is a new domain for DEC-MDPs which presents several research challenges. Our hypothesis is that by modeling healthy cells and cancerous cells as opposing teams, and having policies generated automatically rather than generated by hand as done in previous work, we may gain a more fundamental understanding of cell behavior. By adopting a well-established multi-agent formalism, we have access to algorithms that can provide teams of agents with policies even for environments that feature a high degree of uncertainty.

The complexity of our cancer modeling domain presents two main technical challenges that must be addressed. First, we must deal with the issue of scalability with respect to both planning and execution. The richness of our domain necessitates a detailed state space, while the scale of our domain necessitates the modeling of a large number of agents. The combination of these two factors creates a prohibitively high computational cost when employing traditional DEC-MDPs algorithms. Thus, we need to modify existing techniques in order to provide the scalability our domain demands. Second, our cancer modeling domain requires a new formulation of agent communication. Typical communication in DEC-MDPs is done point-to-point between agents. However, cells communicate through a complex system of biological and chemical signaling which relies on concepts such as diffusion and reception. Therefore, we must develop a new method of message and information passing that incorporates the unique aspects of cell signaling.

This paper is intended to introduce the readers to a useful domain for the techniques developed by the MSDM community as

The Sixth Annual Workshop on Multiagent Sequential Decision-Making in Uncertain Domains (*MSDM-2011*), held in conjunction with *AAMAS-2011* on May 3, 2011 in Taipei, Taiwan.

well as to suggest ways in which those techniques can be applied. By presenting our initial approach, we will highlight the issues and challenges related to this domain that future research must address. Given the complexity of this real-world challenge of modeling cancer cells, we do not claim to have all the answers or even most of the answers. However, we believe this paper will raise some key questions which we hope will initiate a discussion within the community. Indeed, cancer modeling is a significant challenge, which requires a significant team effort, on a large collaborative interdisciplinary scale. By assembling an interdisciplinary team, we have made a beginning, and would invite others to join in this challenge.

At present, we are evaluating our model by running a large number of simulations and performing statistical analysis in search of patterns and emerging trends. By manipulating various parameters, we have started to gain an understanding of how they interact and influence our model. In the near future, we hope to validate our results against experiments conducted by the Center for Applied Molecular Medicine at the University of Southern California.

2. DOMAIN

Recently, whole-cell and whole-organism analyses have been widely applied to study biological processes and disease states. These techniques have permitted the examination of cellular processes and their relationship to physiologic effects in a greater detail than previously possible, enabling better characterizations of pathologic states, such as cancer. The biological community has accordingly begun to emphasize the importance of studying the interactions among cells across the wide length scales of biology. In particular, the multi-scale nature of diseases, like cancer, have recently become highlighted. For example, it has been clearly demonstrated that subtle alterations in a single gene (e.g. Ras, EGFR, p53) can lead to a significant cellular disruption, leading ultimately to a cancer that has a physiologic impact on the whole organism. Consequently, organismic states, such as "healthy" and "diseased", are now hypothesized to arise from cellular defects that upset a cell-system's normal behavior through a combination of endogenous genetic modifications and exogenous environmental perturbations.

As our understanding of cellular processes has developed, so has our understanding that cancer, even in a single patient, is not one disease, but instead hundreds of heterogeneous diseases (sometimes even within the same tumor), unified by the single common gross phenotype of deregulation of cell growth.

Cell growth in normal human cells is typically a tightly regulated process wherein a cell's behavior is modulated through mechanical and chemical signals. Some of these signals are generated by the cell itself, others can come from neighboring cells. Cell growth is regulated when the signals each cell receives are sane and when a cell's reaction to those signals is sane. For example, it has been shown that a normal cell can increase or decrease its growth rate in response to sensing the presence of a chemical growth factor. If, for example, each cell secretes a constant amount of a particular chemical, an excess of that chemical may suggest an excess of cells in the neighborhood, thus signaling a cell to decrease its growth rate, or possibly even to commit suicide through a process called apoptosis. Cells may also alter their growth rate in response to contact with adjacent cells, or in response to changes in external pressure, as may arise from over-packing of cells. Despite the importance of understanding intercellular communication and how cells respond to receiving disparate messages, these interactions are only now being fully appreciated in the investigation of tumor biology.

Cancer cells are distinguished from normal cells by un-regulated, malignant growth. This aberrant growth may arise from unexpected environmental signals, such as the extended presence of inflammation or immune cells. Alternately, aberrant growth may arise from an incorrect response to environmental signals. For example, despite receiving signals suggesting overpopulation, cancer cells may continue to proliferate. The so-called "hallmarks" of cancer can be summarized into five cellular dysfunctions: 1) hyper-responsiveness to internally-generated growth signals, 2) insensitivity to external growth-inhibitory (antigrowth) signals, 3) evasion of programmed cell death (apoptosis), 4) limitless replicative potential, and 5) tissue invasion and metastasis [8].

A number of cell types contribute to a tumor including those that are considered "normal" such as inflammatory, stromal fibroblast, and vascular endothelial cells alongside "cancerous" epithelial cells. Tumorigenicity is controlled by the reciprocal interactions between these different populations of cells and gaining a better understanding of the consequences these interactions have on tumor progression and therapeutic response is critical to the field of tumor biology. For example, fibroblasts are responsible for the synthesis of the extracellular matrix as well as the production of soluble growth factors that can regulate cell proliferation and death and can have a functional role in promoting tumorigenesis. There is also in vitro and in vivo evidence showing that fibroblasts found in the vicinity of tumor cells can alter their production of soluble factors as well as display changes in phenotype that are characteristic of uncontrolled growth and disorganized patterns [12].

Many treatment strategies attack one or more of these dysfunctions to abrogate un-regulated growth or induce cell death. However, cancer cells with different underlying dysfunctions will typically respond to different therapies. As tumors are tremendously heterogeneous, containing multiple cell types in a range of environments, most therapeutic regimes are only effective in a small percentage of the population. Even those therapies that initially appear to be effective often lose effectiveness over time, presumably through changes in the composition of the tumor.

Extensive studies have attempted to enumerate factors that distinguish cancer cells from healthy cells. However, qualitative models describing the relationships between factors such as intercellular communication, cellular response to signals, therapeutic intervention and tumor demographics are still poorly elucidated. By developing quantitative models, it may become possible to both test biological hypotheses about cellular behavior as well as to ultimately develop more successful treatment regimes.

We hypothesize that accurate methods for simulating cancer may lead to dramatic improvements in cancer management. These simulations could be used to predict and characterize response and outcome, which in turn could "handicap the odds" of a therapy succeeding. By helping us ask and answer fundamental questions about the mechanisms, complexity, and evolution of cancer, computational models may enable a new paradigm in treatment. A small number of measurements could be taken from a patient, input as parameters into a model, and used to simulate the response of the patient to a particular therapy. Our understanding of the disease, its progression, and its response to therapy could be encapsulated in computational models which help determine the course of treatment that is most likely to produce a favorable outcome. We will be able to try out hundreds of therapeutic regimes virtually, before ever having to inflict likely unsuccessful treatments on a patient. In addition, we will be able to identify the signals (e.g. changes in a tumor) that indicate a patient is truly responding to therapy, leading to a radical improvement in the standard of care.

3. RELATED WORK

The previous work in this domain has consisted of three main approaches: evolutionary game-theoretic models [3, 6], lattice-based

models [7], and agent-based models [10]. In a game-theoretic model, the interaction between two cells is represented as a normal form game in which both players move simultaneously. A payoff matrix is constructed in which every pairing of player strategies is assigned a reward for each player. While classical game theory would analyze this payoff matrix for strategy equilibria, evolutionary game theory focuses more on the dynamics of strategy change.

In lattice-based models, cells exist in a lattice structure representing a discretized environment. At each time step, a cell transitions to a new state based upon its current state and the states of the neighboring cells, in accordance with physical and biological constraints. The most common lattice-based approach for cancer modeling is cellular automata (CA). In CA models, each location within the lattice has a uniform size and can be occupied by only a single cell. This simplified spatial arrangement makes CA models computationally efficient but imposes artificial limitations on the alignment and interaction of cells. Additionally, it is difficult for CA to model a high degree of heterogeneity in terms of the types of cells and the ways in which they interact.

Agent-based models attempt to address these problems by removing as many artificial constraints as possible. Cells are capable of moving freely through a continuous environment and arranging themselves in nonuniform alignments. Their actions are determined by a set of differential questions representing the biophysically inspired forces which act upon cells. By applying a free-body force diagram, a sophisticated model for cell behavior can be achieved. An agent-based model provides flexibility to design cells with an arbitrary amount of complexity which can be adjusted to fit the needs of a particular simulation. Agent-based techniques have been widely adopted for the purposes of modeling the behavior and interaction of cells [14]. For example, agents have been used to model brain tumors [17], breast cancer [10], and lung cancer [15].

However, the models used in current agent-based simulations are not like the fully-fledged BDI or MDP systems typically seen in the multi-agent community. Rather, the agents are modeled as a finite state machine (FSM), with transitions between states dictated by random variables and Poisson processes. These FSMs are coded by hand, where the goal-orientedness of the behaviors is implicit rather than explicit. Behaviors are not generated by the agents themselves but by a programmer. Additionally, the sophistication and flexibility of agent-based models can result in a steep computational cost which limits the number of cells which can be simulated.

4. APPROACH

We are complementing the previous approaches mentioned in Section 3 by adopting a new approach that allows agents themselves to generate rules of behavior from the rewards associated with performing different actions in different states. We are exploring to see if algorithms and techniques developed by the multi-agent systems community can potentially provide insights which will advance the state of the art in cancer modeling. One of the benefits of using DEC-MDPs is the ability to test two biological hypotheses: (i) that cells are not purely reactive or greedy but rather have evolved to perform actions beneficial to the organism even at their own expense; (ii) that cells are able to coordinate their actions and function collectively as teams.

The actions of a cell can be viewed as a series of sequential decisions. Cells do not exist in a vacuum and thus must consider the environment and the presence of other cells in their decision making. From a multi-agent perspective, sequential decision problems involving multiple agents are often modeled using DEC-MDPs. However, from a cancer modeling perspective, our choice of DEC-MDPs is, perhaps, unintuitive. Cells are not usually considered to be rational or capable of planning their actions. While cells do not explicitly plan, we view the planning phase of DEC-MDPs as reproducing the evolutionary process cells and tumors have undergone over millions of years. Evolution has trained cells to respond to both internal and external stimuli in the way that best ensures survival of the organism. Given this preference for actions that are globally "optimal", it can be argued that healthy cells are acting rationally. Additionally, cells have evolved to coordinate their actions with other cells to produce complex team behaviors. Cancerous cells pose a threat because mutations have caused them to deviate from this sense of rationality. The interests of cancerous cells are no longer inherently aligned with the interests of the organism.

An example of healthy cells making collective decisions occurs when a subpopulation is faced with resource limitation, in terms of either physical space or access to nutrients. The individual cells either adapt and/or acquire genetic changes that increase their fitness or they sacrifice their life to provide space and resources to their neighbors that are better suited to handle the current selective pressures being applied. Additionally, there is evidence of cooperation among subpopulations of cancerous cells as seen in a hypoxic environment when cells undergo anaerobic glycolysis, which leads to acidification as a byproduct. The acidity that is produced causes cell death among normal cells as well as cancerous cells that are not resistant to high acid content. Being able to control the microenvironment and prevent normal cells from co-habitating is a team goal. These cells made a choice to increase their rate of glucose uptake by increasing their glucose membrane transporters in order to proliferate in this low oxygen environment. This ultimately leads to toxicity and cell cycle arrest for cells not adapted to the low pH environment that was generated.

DEC-MDPs provide a unique framework for capturing the implicit coordination and teamwork exhibited by cells. While DEC-MDPs are an entirely new approach to cancer modeling, it can be viewed as a hybrid of lattice-based and agent-based approaches. For the time being, we have simplified our spatial representation by adopting a geometry similar to that found in lattice-based models. As one of the main limitations of lattice-based models is its rigid geometry, we will pursue a more flexible topology in the future while avoiding the fully continuous spatial representation of current agent-based models. Thus, DEC-MDPs may be able to provide a better balance between model fidelity and computational complexity than previous approaches. However, a significant amount of research must be done to improve DEC-MDP techniques before that point can be reached.

Traditionally, DEC-MDPs have featured a planning-centric framework, meaning all computation is performed during the planning phase to generate an optimal joint policy. During execution, agents enact their policies without the need of any additional reasoning. However, finding the optimal joint policy for general DEC-MDPs has been shown to be NEXP-complete [5]. Approaches have been developed for finding the optimal joint policy but are restricted to DEC-MDPs with loosely coupled agents [4, 16]. Reasoning about coordination during planning-time would require evaluating every interaction between every cell for every possible global state. Given that our domain features a large number of agents and a high level of coupling between them, planning a joint policy is infeasible using a planning-centric framework.

By using an execution-centric framework [9], we are able to reduce planning-time computation by shifting some of the burden to execution-time reasoning. To simplify planning, the DEC-MDP is reduced to a single-agent MDP with a modified reward function for the planning phase . This approach is similar to the TREMOR approach [13], in which the reward and transition functions are

shaped and then the DEC-POMDP problem is reduced to a single-agent POMDP. However, TREMOR does not have additional coordination at execution time which will eventually be a part of our work as discussed in Section 5. The result is a policy that has been generated for an individual agent but is team-oriented.

To avoid the worst-case computational complexities of DEC-MDPs, no explicit coordination reasoning occurs during the planning phase. Instead all coordination reasoning is done explicitly at execution time as in [9]. This simplification of the planning step increases the scalability of our system as agents reason about what and when to communicate in real time. In order to achieve this execution-time reasoning, the sequential decision making of each agent is modeled as a MDP defined by the tuple $\langle \mathbb{S}, \mathbb{A}, \mathbb{T}, \mathbb{R} \rangle$.

Another reason for adopting the execution-centric approach is that, unlike other DEC-MDP work, our goal is to simulate the execution of policies and understand the results of this execution. Just obtaining an expected value of a policy is insufficient; we wish to gain a fundamental understanding of the behavior and interactions of cells. We are seeking to exploit the execution-centric framework to its full potential. It is possible that, in the future, we will need to scale back and allow for some planning-centered coordination.

5. MODEL

In this section, we provide an in-depth description of our model as well as the reasoning behind some of our design decisions. We feel it is important to provide this detailed information so that it can serve as the starting point for future discussion and research.

Our model consists of an environment and two teams of agents. The environment is a 3-D lattice structure which represents a discretized section of tissue. Each location within the lattice has a nutrient level which remains static over time. This level combines the availability of various nutrients that cells need to be prosperous (oxygen, glucose, etc.) into a single integer value. Represented as an agent, each cell has an age, a team affiliation, and a location within the environment. The age of a cell indicates the number of time steps a cell has existed in the simulation. The team affiliation of a cell indicates whether a cell is a member of the team of healthy cells or the team of cancerous cells and what policy to execute.

In reality, a cell cannot directly sense the exact number of cells in the immediate vicinity. However, by communicating through chemical signaling a cell can get a noisy approximation of the number of neighboring cells. For the sake of simplicity, we have not yet implemented communication but rather two models for simulating communication. In the first model, cells are able to engage in perfect communication and thus the exact number of cohabitating cells is known perfectly. In the second model, communication is noisy which creates the possibility of a cell incorrectly perceiving the number of cohabitating cells. In Section 7.2, we examine the effect of noisy communication on the overall behavior of the system.

Using our single-agent MDP, we could generate a policy for a particular environment (a specific mapping of lattice locations to nutrient level). However, this policy would only be useful if our agents encountered that exact environment. Thus, we want to generate a policy that is general enough to be applicable to any environment. We can achieve this generalization but in the process we lose any guarantees on the optimality of our policies.

5.1 State Space

The initial representation of our state space, $\mathbb{S}$, was defined by the tuple, $\langle X, Y, Z, A, N, C \rangle$. X, Y, and Z are the spacial coordinates that comprise the physical location of the cell within the environment. A is the age of the cell represented as an integer value. N is the nutrient level of the location occupied by the cell within the environment. C is the number of other cells who are cohabitating in the same location in the environment. In the following sections, we will refer to the cell age, nutrient level, and number of cohabitating cells for state s as s_a, s_n, s_c.

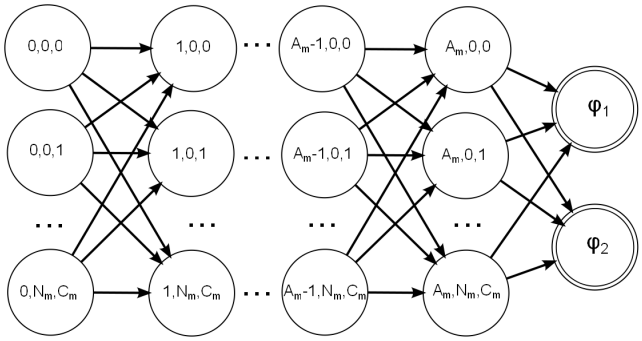


Figure 1: State Space Representation

To apply our MDP algorithms, we need a finite state space. Thus, we define the maximum value for each variable which establishes a range of possible values. In doing so, we create a state space of size $(X_{max}+1) \times (Y_{max}+1) \times (Z_{max}+1) \times (A_{max}+1) \times (N_{max}+1) \times (C_{max}+1)$. This representation suffers from two main drawbacks. First, the size of the state space scales so quickly that computing policies for any large environment would be impractical. Second, the policy generated for such a state space would contain a significant amount of redundant information. Our cells would be required to learn the same lessons over and over again for each location in the environment. We desired a more generalized model which reduces the scale and redundancy of our state space. Thus, we chose a more compact state representation, $\mathbb{S} = \langle A, N, C \rangle$, which reduces the size of the state space to $(A_{max}+1) \times (N_{max}+1) \times (C_{max}+1)$. This allows us to capture the most salient features of the environment while creating a state space whose size is independent of the physical dimensions of our environment.

Additionally, there are two terminal states ϕ_1 and ϕ_2 which represent, respectively, a cell dying in a controlled, beneficial manner and a cell dying in a uncontrolled, harmful manner. Transitioning into either ϕ_1 or ϕ_2 results in the cell being removed from the simulation during the following time step. An explanation as to how a cell transitions into a terminal state can be found in Section 5.3.

Figure 1 shows a visual representation of the state space in which A_m, N_m, and C_m are equivalent to the values A_{max}, N_{max}, and C_{max} described above. For each age level, there are $(N_{max}+1) \times (C_{max}+1)$ states. Each of these states can transition to any state in the next age level as well as the terminal states. This holds until the age of the cell reaches A_{max} at which point it can only transition to the terminal states. For the sake of clarity, we have removed all transition edges to the terminal states except for those emanating for the A_{max} age level. It should be noted that the age of a cell increases with each time step, so all transition edges are directed.

5.2 Action Space

At each time step, each agent selects an action to perform from the action space $\mathbb{A} = \{\mathcal{Q}, \mathcal{M}, \mathcal{P}, \mathcal{A}\}$. $\mathcal{Q}$ represents a cell becoming quiescent and remaining in its current location in the environment. $\mathcal{M}$ represents a cell becoming motile and migrating to a neighboring location in the environment. $\mathcal{P}$ represents a cell proliferating, a process by which a cell divides into two daughter cells. $\mathcal{A}$ represents a cell undergoing apoptosis, a process of programmed cell death in which a cell chooses to commit suicide in a controlled

manner in order to promote the overall health of the organism. With one exception, a cell can perform any action from any non-terminal state. This exception is when a cell reaches the age of A_{max} and must perform either $\mathcal{P}$ or $\mathcal{A}$.

5.3 Transition Function

Creating a sophisticated transition function is one way in which the complexity and uncertainty of our domain can be encoded into our model. The transition function, $\mathbb{T}$, is defined as:

- $\mathbb{T}(\phi_2|s,a) = \epsilon$
- $\mathbb{T}(s'|s,\mathcal{Q}) = 1-\epsilon, \{s'|s'_a = s_a + 1, s'_n = s_n, s'_c = s_c\}$
- $\mathbb{T}(s'|s,\mathcal{M}) = \frac{1-\epsilon}{(N_{max}+1)\times(C_{max}+1)}, \{s'|s'_a = s_a + 1\}$
- $\mathbb{T}(\phi_1|s,\mathcal{P}) = 1-\epsilon$
- $\mathbb{T}(\phi_1|s,\mathcal{A}) = 1-\epsilon$

$\mathbb{T}(s'|s,a)$, defines the probability of an agent transitioning to state s' after performing action a in state s. While performing any action, it is possible for a cell to become necrotic (transitioning to state ϕ_2) with probability ϵ, resulting in an uncontrolled, premature death due to factors external to the cell. Quiescence ($\mathcal{Q}$) causes a cell to transition, with probability $1-\epsilon$, to a state similar to its previous state except with an increased age. Both proliferation ($\mathcal{P}$) and apoptosis ($\mathcal{A}$) result in a cell transitioning to terminal state ϕ_1 with probability $1-\epsilon$. A generalized environment causes one significant complication in the transition function. It is no longer possible to definitively know which states can be transitioned into from a given state when a cell is migrating ($\mathcal{M}$). We have no choice but to assume that s can transition, with uniform probability, into any state s' in which $s'_a = s_a + 1$.

5.4 Reward Function

By decomposing a team reward function into carefully constructed individual reward functions, we can recreate the propensity cells have for working together as a team. The reward function, $\mathbb{R}$, is defined as:

- $\mathbb{R}(\phi_2|s,a) = -20$
- $\mathbb{R}(s'|s,\mathcal{Q}) = 0$
- $\mathbb{R}(s'|s,\mathcal{M}) = \alpha_1(s'_n - n_0) + \alpha_2(c_0 - s'_c)$
- $\mathbb{R}(\phi_1|s,\mathcal{P}) = \alpha_3(s_n - n_0) + \alpha_4(c_0 - (s_c + 1))$
- $\mathbb{R}(\phi_1|s,\mathcal{A}) = \alpha_5(s_a - a_0) + \alpha_6(n_0 - s_n) + \alpha_7(s_c - c_0)$

$\mathbb{R}(s'|s,a)$ defines the reward an agent receives by transitioning from state s to state s' by performing action a. a_0, n_0, and c_0 are the nominal values for cell age, nutrient level, and number of cohabitating cells. These values must be specified and represent the normal or expected values of these variables. These nominal values can then be compared to the actual values to determine, in part, the reward received for the tuple $\langle s,a,s'\rangle$. The reward function has been generalized with a parameterized vector $\alpha = \{\alpha_1 ... \alpha_7\}$. This general form allows for the same reward function structure to be used by both healthy cells and cancerous cells as well as the fine tuning of cell behavior.

We constructed the reward function to reflect basic biologicial principles in a such way that, for healthy cells, all α parameters should be positive. The method of comparison between the nominal and actual values is thus of great importance. For $(actual - nominal)$, it is beneficial to be above the nominal value, whereas with $(nominal - actual)$, it is harmful to be above the nominal value. This is useful for establishing the preferences of our cells.

Quiescence ($\mathcal{Q}$) is the default action for cells which is why it yields a reward of zero. A cell will only perform another action if it results in a higher reward than $\mathcal{Q}$. A cell is rewarded for migrating ($\mathcal{M}$) to or proliferating ($\mathcal{P}$) in a state with a high nutrient level and a low number of cohabitating cells. Conversely, a cell is rewarded for inducing apoptosis ($\mathcal{A}$) when it has a high age and is in a state with a low nutrient level and a high number of cohabitating cells. Despite being removed from the simulation in the process, performing $\mathcal{P}$ and $\mathcal{A}$ can yield a positive reward because they are vital biological processes which benefit the organism. In contrast, a cell receives a constant reward of -20 for becoming necrotic because it can result in the formation of microcalcifications which are detrimental to surrounding cells.

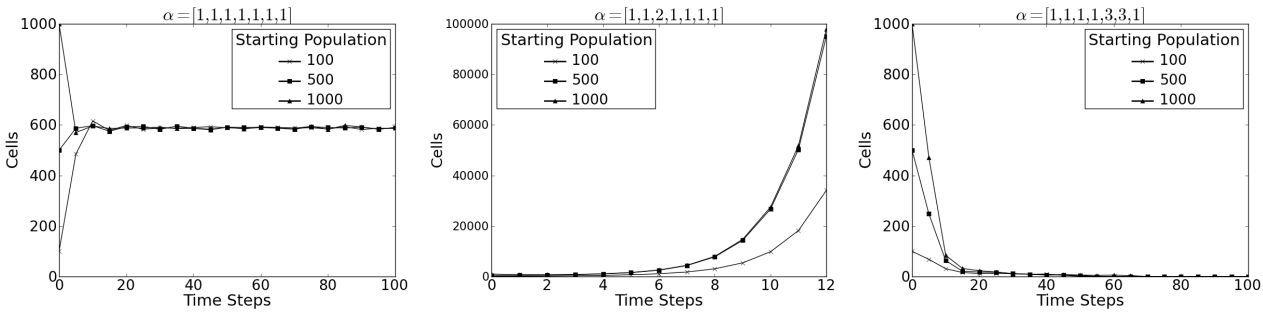
Manipulating the α vector allows us to generate a different policy for the team of cancerous cells. For example, cancerous cells have been shown to ignore the signals that cause healthy cells to regulate their proliferation. This phenomena can be modeled by decreasing the value of α_4, resulting in a smaller negative reward for proliferating in a location with a large number of cohabitating cells. Cancer cells are also less likely to respond to the signals that cause healthy cells to induce apoptosis. This behavior can be achieved by decreasing the values for α_5, α_6, and α_7.

5.5 Planning

Once our single-agent MDPs have been defined, we can use value iteration to automatically generate two policies, one for healthy cells and one for cancerous cells. Value iteration takes as input a discount factor, γ, where $0 < \gamma < 1$, which controls the time horizon over which the agent plans. When γ is small, the agent places greater value on immediate rewards rather than future rewards. As a result, the agent does little to no planning, instead always choosing the action which results in the highest reward. Whereas when γ is large, the agent values future rewards over immediate rewards. This provides incentive for the agent to maximize the expected reward over an entire sequence of actions rather than for each individual action. For example, consider a cell in a state with a low nutrient level and a high number of cohabitating cells. A greedy agent would perform action $\mathcal{A}$ and receive a positive reward. However, a planning agent would be able to determine if it was advantageous to instead perform $\mathcal{Q}$ for multiple time steps, risking necrosis in the process, until it's age reaches A_{max} at which point $\mathcal{A}$ would yield an even higher reward. In order to simulate the learning and adaptation resulting from evolution in our cells, we set $\gamma = .99$. The policies generated by value iteration are optimal only for the generalized model of the state space. Thus, when this policy is executed in a specific environment it is no longer guaranteed to be optimal.

6. SIMULATION AND VISUALIZATION

The interactive aspect of our system consists of our simulation and visualization environment. Given a policy for healthy and cancerous cells, our simulation environment can model the interactions of cells over time. That model is then delivered to our visualization environment where it is displayed to the user. This visualization environment features a graphical user interface which displays a 3-D wireframe lattice representing a section of tissue. The tissue is populated by both healthy and cancerous cells, whose type and age can be determined by color. Healthy cells appear red, cancerous cells appear blue, and as a cell ages, its color begins to darken. In order to allow for better viewing of our simulations, we added the ability to rotate the lattice as well as zoom in and out. Figure 2



(a) Steady Population (b) Exponential Proliferation (c) Exponential Die-Off

Figure 3: Homogeneous Cell Population Phenomena

shows a series of screenshots from our visualization environment.

Users are able to control the initial configuration of the system including the number of healthy cells and cancerous cells. Additionally, users will have the ability to select the policies executed by the cells from a list of policies that have been precomputed with different reward function parameter settings. Given that our model is discrete with respect to time, users step through a simulation by providing keyboard input. At each time step, the system will transition from one state to another and information about the actions of individual agents will be displayed. By trying out different configurations, users can see how changing the parameters of our model can affect the behavior of the overall system.

7. EVALUATION

We conducted preliminary testing to evaluate the capabilities of our model. We constructed a state space in which $A_{max} = 9$, $N_{max} = 9$, and $C_{max} = 9$, resulting in 1,002 states when ϕ_1 and ϕ_2 are included. In our reward function, we set $a_0 = 4.5$, $n_0 = 4.5$, and $c_0 = 4.5$. Our environment consists of a $5 \times 5 \times 5$ cube with randomly distributed nutrient levels. An initial population of both healthy and cancerous cells was specified and these cells were assigned random ages and locations within the environment. All the results shown in this section have been averaged over 30 independent simulations which ran for 100 time steps.

7.1 Homogeneous Agents

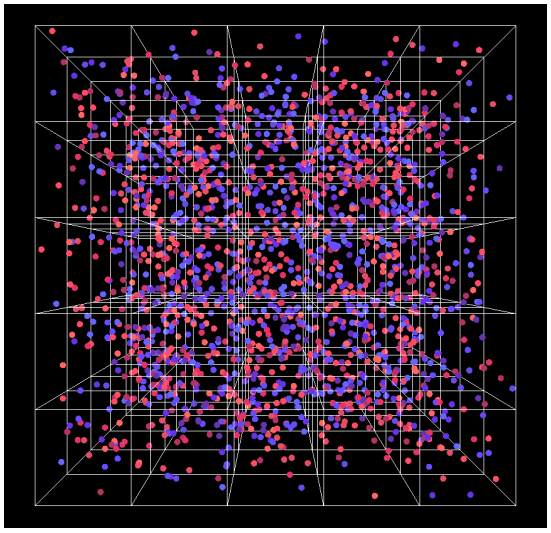
An initial round of simulations was conducted with homogeneous populations of cells. For these simulations, we were interested in observing how the size of the cell population varied over time. Additionally, we tested to see if the size of the starting population had any effect on the system. The baseline cell population was defined by $\alpha = [1, 1, 1, 1, 1, 1, 1]$, i.e. all components of the reward function are weighted equally. Figure 3(a) shows the results from our baseline simulations with starting populations of 100, 500, and 1,000 cells. Here the y-axis represents the total number of cells and the x-axis represents the number of simulation time steps that have elapsed. We can see that for each starting population, the total population converges to a steady population of approximately 600 cells. This suggests that the structure of our reward function naturally lends itself to a balance between the desire to proliferate and the desire to avoid overcrowding.

However, by parametrizing our reward function with the α vector we are able to adjust this balance as well as deviate from it completely. For example, in Figure 3(b), we observe an unstable population of cells defined by $\alpha = [1, 1, 2, 1, 1, 1, 1]$. For each starting population, the total population increases exponentially with each time step. After only 12 times steps, the simulations were stopped as the population had exceeded 100,000 cells. Setting $\alpha_3 = 2$ increases the importance of nutrient level in determining when to proliferate so much that cells are willing to ignore the overcrowding taking place. By increasing α_3, these cells will begin to prefer proliferation at lower nutrient levels than the baseline population. Thus, the number of states where proliferation represents the optimal action increases. Another example of an unstable population is shown in Figure 3(c). At each time step, the cell population defined by $\alpha = [1, 1, 1, 1, 3, 3, 1]$ decreases until eventually the entire population has died out. When compared to the baseline population, increasing α_5 encourages cells to induce apoptosis at younger ages, whereas α_6 prompts cells to become apoptotic at higher nutrient levels. The combination of these two factors results in the exponential die-off that we observed.

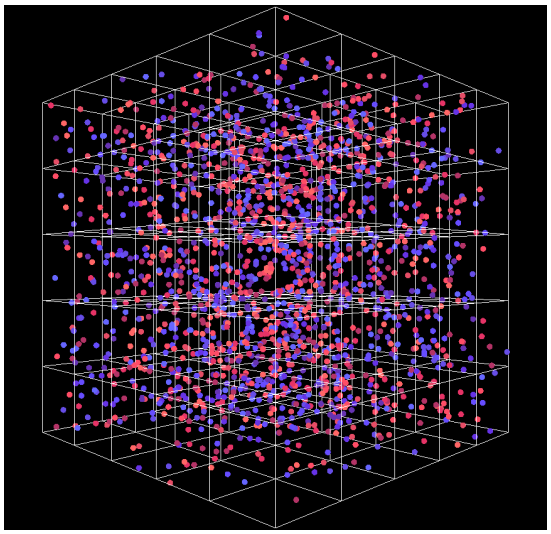
Some α vectors generate a steady population, while others lead to unstable populations that experience exponential proliferation or die-off. Fine tuning of the α vectors can influence the size of a steady population or the rate of proliferation and die-off. We have shown that the α vector is capable of capturing a variety of cell behaviors and has significant impact on the population dynamics of cells. The next step is to calibrate our α vector and reward function to match the real world phenomena observed in lab experiments to be conducted by the Center for Applied Molecular Medicine at the University of Southern California. These experiments will focus on analyzing the growth, death, and migratory capabilities of population admixtures under varying nutrient and cell confluency conditions.

7.2 Noise

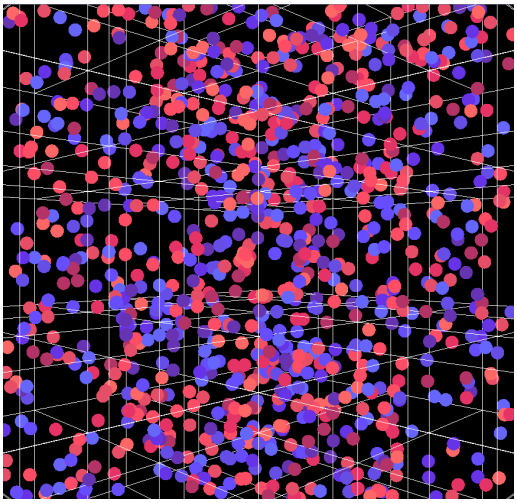
The results in Section 7.1 assume cells have perfect knowledge about the environment. This was meant to simulate the ability of cells to engage in perfect communication. In reality, communication between cells occurs through a series of chemical signals which are inherently noisy. Thus, we wanted to approximate this noise and investigate its effect on our simulations. We implemented noise in the following manner; with probability $1 - p$, a cell perceives the exact number of cohabitating cells and with probability p perceives a random number uniformly distributed between 0 and C_{max}. The effect of varying the amount of noise on a starting population of 1,000 cells is shown in Figure 4. We can see that when there is no noise, the cell population defined by $\alpha = [1, 1, 1, 1, 2.5, 2.5, 1]$ converges to a steady population of approximately 250 cells. As the level of noise is increased, we observe that the steady population decreases. It is also worthwhile to note that this α vector has similar properties to the vector presented



(a) Default View



(b) Rotated View



(c) Zoomed View

Figure 2: Visualization Environment

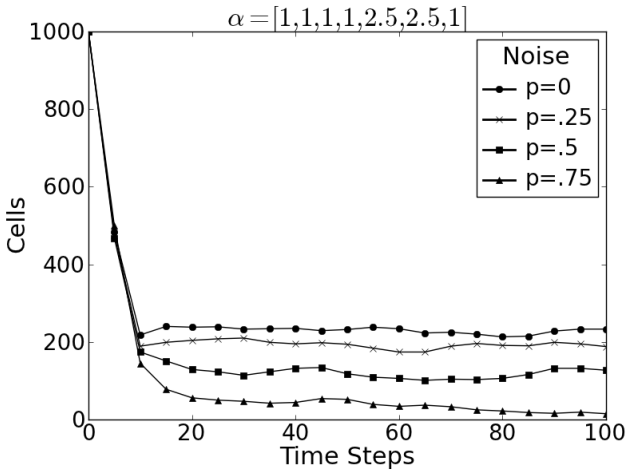


Figure 4: Effect of Noise on Homogeneous Cell Populations

in Section 7.1 which resulted in exponential die-off, but results in fundamentally different behavior.

In this case, the steady population is small, and thus the average number of cells in each location is low. Therefore, when a cell receives a noisy approximation of the number of cohabitating cells, there is a high probability it will be an overestimation. Falsely thinking it is in an overcrowded location, the cell induces apoptosis resulting in a lower steady population. These results show the impact noise can have on the behavior of cells and makes clear the need to implement realistic cell communication in the future.

7.3 Heterogeneous Agents

While simulating homogeneous cell populations is an important first step, in order to convincingly model cancer we need to able to model heterogeneous cell populations. We have started by modeling two groups of cells and observing how they interact. Group A consists of cells defined by $\alpha = [1, 1, 1, 1, 1, 1, 1]$, which were shown in Section 7.1 to converge to a steady population of approximately 600 cells. Group B consists of cells defined by $\alpha = [1, 1, 1, 1, 2.5, 2.5, 1]$, which were shown in Section 7.2 to converge to a steady population of approximately 250 cells.

Figure 5 shows the results of simulations in which each group had a starting population of 500 cells. With an initial total population of 1,000, both groups are above their steady populations and feel the need to induce apoptosis. This continues until the total population approaches 600, as cells in A are now content to be quiescent. B is still above its steady population and so its cell continue to be apoptotic. However, any decrease resulting from the apoptosis of B cells is soon matched by the proliferation of A cells trying to push the population back to 600. Still feeling overcrowded, B continues to apoptose until all B cells have died out.

In our current model, the most "stable" cell type is the one that proliferates the most. Forcing a cell type to remain above its steady population will always cause it to die out, as apoptosis is the only recourse to improve the situation. Thus, one cell type will dominate the other unless they converge to the same stable population. While this may be effective for modeling certain types of cancer, we would like to be able to model a wider variety of interactions. To do so, we will need to increase the sophistication of our environment. Currently, there is no punishment mechanism for greedy cell types that proliferate regardless of the number of cohabitating cells. There are two potential solutions that we will look at go-

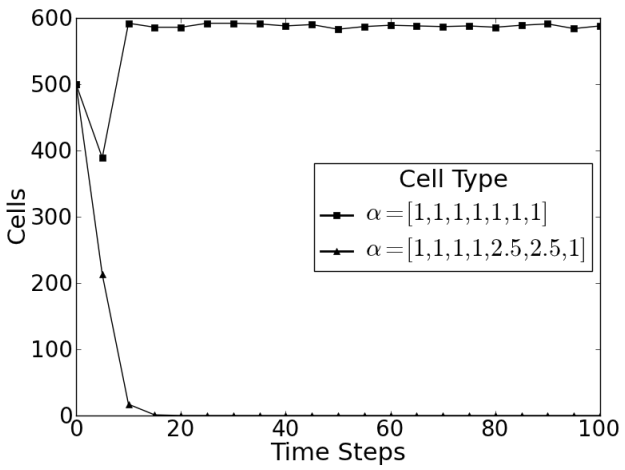


Figure 5: Heterogeneous Cell Populations

ing forward. First, implementing dynamic nutrient levels which are depleted or replenished based on the number cells in a location. Second, making the probability of experiencing of necrosis, ϵ, a function of the current state rather than being constant. Not only will these changes improve the fidelity of our environment but they will also increase the sophistication of our MDP model, forcing our agents to reason about a more complex and dynamic state space.

8. DISCUSSION AND CONCLUSION

Cancer places a significant social and economic burden on society. Annually claiming hundreds of thousands of lives, cancer poses one of the greatest health risks not only in the U.S., but around the world as well. Computer models have been suggested as a method for gaining a more fundamental understanding of cancer. It is hoped that these models could eventually help expedite new treatments that will improve the survival rate for cancer patients. Numerous models have been presented using game theoretic, latticed-based, agent-based approaches. We have set out to assess the potential of applying multi-agent techniques to model cancer. In doing so, we have proposed an innovative model using DEC-MDPs, in which healthy cells and cancerous cells are modeled as opposing teams of agents. As this is our initial effort in the domain, we have attempted to abstract away much of the inherent complexity while retaining the salient aspects. In the future, we will increase the sophistication and complexity of our model so that we will be able to more accurately simulate the processes and interactions that occur in nature. The goal is to develop an integrated framework that relies on both the model simulations as well as validation from the experimental data to make an accurate assessments of cancer progression. While there is still significant work to be done, we feel we have take a successful first step toward our larger goal. Our current focus is on DEC-MDPs, but future developments may lead us to explore other multi-agent approaches such as partial observable stochastic games (POSG). In this paper, we set out to introduce the cancer modeling domain to MSDM and the multi-agent techniques as a whole. We invite others to join us in applying DEC-MDPs and other multi-agent approaches to cancer modeling. Continued effort in this area has the potential to save lives and lead to innovations in both the multi-agent and biomedical communities.

9. REFERENCES

[1] S. Altekruse, C. Kosary, M. Krapcho, N. Neyman, R. Aminou, W. Waldron, J. Ruhl, N. Howlader, Z. Tatalovich, H. Cho, et al. SEER Cancer Statistics Review, 1975-2007, National Cancer Institute. Bethesda, MD.

[2] American Cancer Sociey. *Cancer Facts & Figures 2010*, Atlanta: American Cancer Socity; 2010.

[3] D. Basanta, M. Simon, H. Hatzikirou, and A. Deutsch. Evolutionary game theory elucidates the role of glycolysis in glioma progression and invasion. *Cell Proliferation*, 41(6):980–987, 2008.

[4] R. Becker, S. Zilberstein, V. Lesser, and C. Goldman. Transition-independent decentralized Markov decision processes. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 41–48. ACM, 2003.

[5] D. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of Markov decision processes. *Mathematics of operations research*, 27(4):819–840, 2002.

[6] D. Dingli, F. Santos, S. Van Segbroeck, and J. Pacheco. Cancer phenotype as the outcome of an evolutionary game between normal and malignant cells. *British journal of cancer*, 101(7):1130–1136, 2009.

[7] P. Gerlee and A. Anderson. An evolutionary hybrid cellular automaton model of solid tumour growth. *Journal of theoretical biology*, 246(4):583–603, 2007.

[8] D. Hanahan, R. Weinberg, et al. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.

[9] J. Kwak, R. Yang, Z. Yin, M. Taylor, and M. Tambe. Teamwork and Coordination under Model Uncertainty in DEC-POMDPs. In *AAAI Workshop on Interactive Decision Theory and Game Theory (IDTGT)*, 2010.

[10] P. MacKlin and M. E. Edgerton. Discrete cell modeling. In V. Cristini and J. Lowengrub, editors, *Multiscale Modeling of Cancer*. Cambridge University Press, 2010.

[11] D. Pynadath and M. Tambe. The communicative multiagent team decision problem: Analyzing teamwork theories and models. *Journal of Artificial Intelligence Research*, 16(1):389–423, 2002.

[12] A. Van den Hooff. Stromal involvement in malignant growth. *Advances in cancer research*, 50:159–196, 1988.

[13] P. Varakantham, J. Kwak, M. Taylor, J. Marecki, P. Scerri, and M. Tambe. Exploiting coordination locales in distributed POMDPs via social model shaping. In *Proc. of ICAPS*, 2009.

[14] D. Walker, J. Southgate, G. Hill, M. Holcombe, D. Hose, S. Wood, S. Mac Neil, and R. Smallwood. The epitheliome: agent-based modelling of the social behaviour of cells. *Biosystems*, 76(1-3):89–100, 2004.

[15] Z. Wang, L. Zhang, J. Sagotsky, and T. Deisboeck. Simulating non-small cell lung cancer with a multiscale agent-based model. *Theoretical Biology and Medical Modelling*, 4(1):50, 2007.

[16] S. Witwicki and E. Durfee. Influence-based policy abstraction for weakly-coupled Dec-POMDPs. In *Proceedings of the Twentieth International Conference on Automated Planning and Scheduling (Toronto, Canada, 2010)*, 2010.

[17] L. Zhang, Z. Wang, J. Sagotsky, and T. Deisboeck. Multiscale agent-based cancer modeling. *Journal of mathematical biology*, 58(4):545–559, 2009.