Ian Wood
5/3/13
I690, Prof. Flammini

**Final Project Summary: Exploring the Parameter Space of a T-Cell Cross-Regulation Machine Classifier**

For my final project, I wanted to investigate the conditions under which a set of parameters for a machine classifier inspired by a model of the immune system would result in good performance. My work is a continuation of Alaa Abi-Hadar's work adapting Jorge Carneiro's mathematical model of T-Cell Cross-Regulation to a machine classifier. The rest of this summary will be organized as follows: a description of the system and the problem, previous work on the BioCreative II.5 corpus, my approach, preliminary results, and a discussion of future directions.

**System Description**

When a substance enters the body it can be broken down into small peptide chains called antigens and presented to the immune system on antigen-presenting cells (APCs). T-cells aid in an immune response by proliferating and signaling other cells when they encounter antigens that can bind to their receptors on APCs. Jorge Carneiro developed a mathematical model of the interactions between two types of t-cells, effectors and regulators. Biological experiments have shown that regulator t-cells will not proliferate in the presence of APCs presenting antigens they recognize unless effector t-cells are also present, although effectors proliferate without the presence of regulators. If regulators and effectors are both present however, the proliferation of effectors is suppressed. Carneiro found that a few simple production rules can model the interactions of said t-cells, producing for a given set of parameters and initial populations of t-cells two stable equilibriums depending on the density of APCs, one in which effectors dominate the t-cell population and a shorter regime in which regulators dominate. The rules are: two t-cells can bind to an APC conjugation site, without binding to an APC site, t-cells die at some rate, if one alone or two effectors bind to an APC they duplicate, and if a regulator and an effector bind to an APC the regulator will duplicate (nothing happens if regulators bind to an APC without effectors). The two stable equilibriums allow the system to classify antigens as either non-self or self, where non-self represents potentially harmful intruders, while self represents endogenous antigens that should not trigger an immune response [1].

Alaa Abi-Hadar performed work to turn the above model into a machine classifier, and I have been working to re-implement and continue developing his system [2]. In the metaphor of the system, the features of a document (in the current implementation simply words), are antigens that are presented to the system. Documents are labeled self, non-self, or unknown, and for each new feature presented in a document, an initial population of effectors and regulators are produced that can bind to said feature. The features are presented in an array that represents an APC, and (non-overlapping) consecutive pairs represent conjugation sites. For each document, t-cells bind to the features they match, and proliferate and die according to the above rules. The populations of regulators and effectors that can bind to features on the APC can then be used to classify the document. For a document d, represented by a set of features $A_d$, and populations of regulators $R_f$ and effectors $E_f$ for each feature f, the cosine normalized scores for regulator population R(d) and effector population E(d) are calculated by:

$$R(d) = \sum_{f \in A_d} R_f / \sqrt{R_f^2 + E_f^2}$$

$$E(d) = \sum_{f \in A_d} E_f / \sqrt{R_f^2 + E_f^2}$$

**The Problem**

Some possible benefits of the system that I have been trying to investigate are the system's innate ability to respond to temporal dynamics in the presented documents, and its usefulness in classifying unbalanced sets of documents, since effectors have an advantage. However, there is a problem, since the agent-based system takes a significant amount of time to run, so the large parameter space is difficult to explore. Without including variations on the algorithm, the parameters include: $N_{slot}$, the number of antigens produced for each unique feature; $D_E$ and $D_R$, the death rates for unbound effectors and regulators respectively; $E_0^-$, $E_0^+$, $E_0^u$ the number of effectors initially produced for features first encountered in nonself, self, and unknown labeled documents respectively; and $R_0^-$, $R_0^+$, $R_0^u$ which are the same parameters for the initially produced regulators.

**Prior Work**

Al performed an exhaustive search over this parameter space in order to conduct experiments regarding variations on the algorithm and the ordering of the presentation of documents from the BioCreative II.5 corpus. He set only one $E_0$, so that $E_0^- = E_0^+ = E_0^u$, and considered non-self-labeled and unknown documents to be the same, so $R_0^- = R_0^u$. This makes some biological sense, since antigens do not come labeled, but without a model for the hyper-mutation and selection of t-cells in the thymus, these labeled initial biases are necessary for the performance of the machine classifier. He found the performance of this classifier is comparable to Naïve Bayes and the other submissions to the BioCreative II.5 challenge. I have been trying to replicate some of his work on the BioCreative II.5 corpus regarding the ordered presentation of documents to the classifier, as described in [2]. However, I have not limited the features presented to the system by the ranked product of their TF.IDF and separation scores, and I have not limited the parameter configurations in the same way Al did.

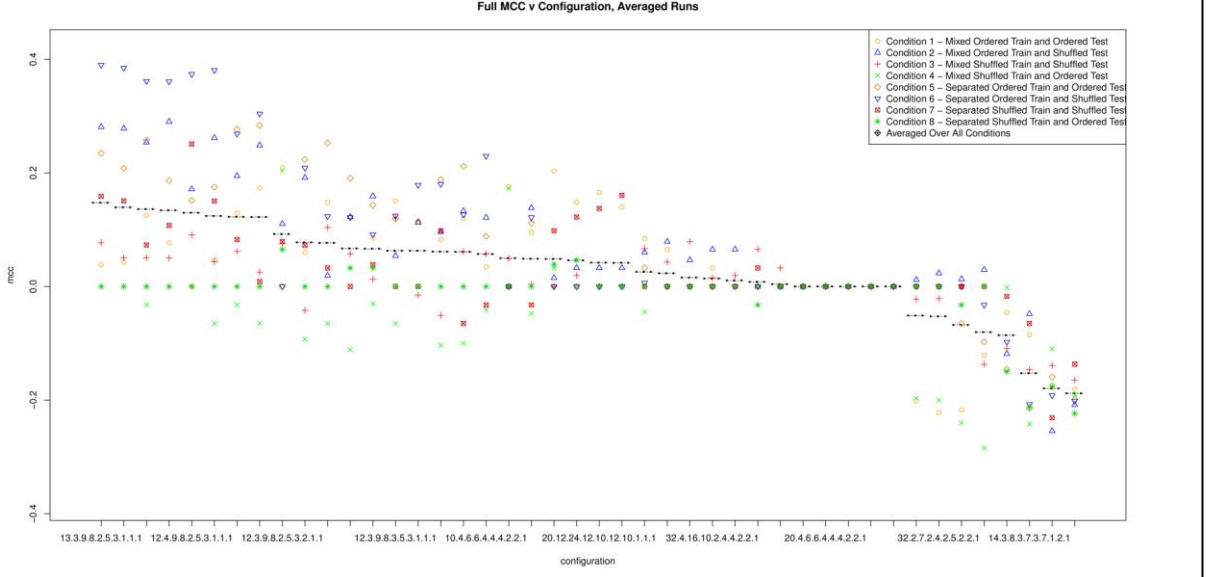


**Fig. 1.** MCC performance of configurations over multiple document orderings and selections

My results so far are displayed in Fig. 1, but these are preliminary and involve a haphazard search over different configurations of parameters. Without going into too much detail, Fig. 1. plots parameter configurations along the x-axis and the Matthew's correlation coefficient on the y-axis, measuring the classifier's performance on variously ordered presentations of the BioCreative II.5 corpus (classification tries to determine whether an article from the biomedical literature involves protein-protein interaction or not). The differently colored conditions correspond to different orderings of the documents by their publication date and class (training and testing) as well as different random

selections of which negative documents would go into the balanced training and testing sets. The clustering of performance across conditions for each configuration (i.e. a given configuration tends to perform well or poorly across conditions) suggests that there are some relations between the parameters that can determine performance without respect to the features of the documents being classified. For this project I started to investigate what those relations might be.

**Approach**

My approach is in three parts. I first looked at how the distributions of T-Cell population R(d) and E(d) scores produced by different parameter configurations corresponded to system parameters. Second, I created an artificial data set, to look at performance under idealized conditions. Third, I attempted a mathematical analysis of the system under ideal conditions. The first two approaches have produced some suggestive preliminary results, but my last approach was largely a failure.

**Preliminary Results**

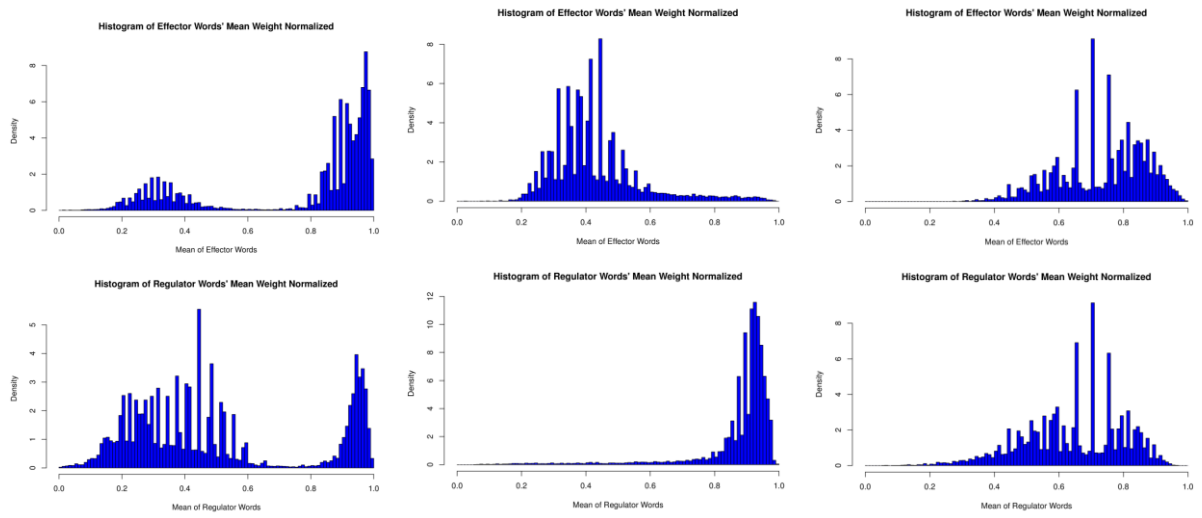


**Fig. 2.** Distribution of mean E(d) and R(d) scores across features. Mean was calculated over all documents after the feature's initial appearance. Top: distributions for effector scores, Bottom: distributions for regulator scores. Left: distributions for row 1 of Table 1, Middle: distributions for row 2 of Table 1, Right: distributions for row 3 of Table 1.

**Table 1**

| $N_{slot}$ | $E_0^+$ | $R_0^+$ | $E_0^-$ | $R_0^-$ | $E_0^u$ | $R_0^u$ | $D_E$ | $D_R$ | precision | accuracy | recall | mcc | f1 | tpos | tneg | fpos | fneg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 3 | 9 | 8 | 2 | 6 | 3 | .1 | .1 | 0.74 | **0.8** | 0.93 | **0.62** | **0.82** | 28 | 20 | 10 | 2 |
| 14 | 3 | 8 | 3 | 7 | 3 | 7 | .1 | .2 | 0 | **0.38** | 0 | **-0.36** | **-1** | 0 | 23 | 7 | 30 |
| 20 | 4 | 6 | 6 | 4 | 4 | 4 | .1 | .1 | 0 | **0.5** | 0 | **0** | **0** | 0 | 30 | 0 | 30 |

Fig. 2. presents the distribution of mean E(d) and R(d) scores for features of the t-cell population produced by individual runs of the classifier over all documents. The parameters and performances listed in Table 1 top to bottom correspond to the distributions left to right. There are three peaks in each distribution, corresponding to the cosine normalized scores of the initial populations for each type of label. From initial observations so far, there seems to be no clear effect from death rate or document ordering on the qualitative aspects of these distributions. Parameter configurations that tend to perform well tend to have separated peaks for both effector and regulator distributions and areas of the distributions that don't overlap between classes. Parameter configurations that tend to do poorly tend

to have clustered peaks, and distributions that overlap between the classes. This makes sense, since initial distributions of t-cells should be clearly distinguishable between self and non-self documents. By the production rules of the system, regulators are at a disadvantage to effectors, so asymmetries in the distributions of effectors and regulators would also seem required.

For my second approach, I generated an artificial corpus to see how a single common feature's E(d) and R(d) scores could change over time. The artificial corpus consists of 10 documents of 100 words each. Each document has 99 unique random words, and one common word, arbitrarily chosen to be "lambda". Since the system should be able to correct for errors in incorrect initial classifications of document features, I investigated how the score of "lambda" could change after an initial appearance in a document of the incorrect class. In set 1, the first document is labeled Self and the rest Non-self, and in set 2, the first document is labeled Non-self and the rest Self. I explored over the parameter space presented in Table 2 (with a step size of one for each range), using the simplifications of the space that Al used, resulting in 2,000 parameter configurations for each set.

**Table 2**

| Parameter | Values |
|---|---|
| $N_{slot}$ | [10, 13] |
| $D_E$ | 0.1 |
| $D_R$ | 0.1 |
| $E_0$ | [5, 14] |
| $R_0^-$ | [1, 6] |
| $R_0^+$ | [6, 16] |

To select the best performing parameters, I set a threshold of 0.2 or greater of correct score change over the 10 documents. Fig. 3 displays those parameter configurations that achieved appropriate behavior in set 1, set 2, and both. It is clear from the graphs that for this parameter space, it is much easier for a feature to transfer from the self to non-self class than it is to change from the non-self to the self class, which is expected, given the regulators' disadvantage in the interaction rules. Across all 2,000 parameter configurations, only 12 configurations were able to achieve appropriate corrective behavior in both set 1 and set 2, presented in Table 3.

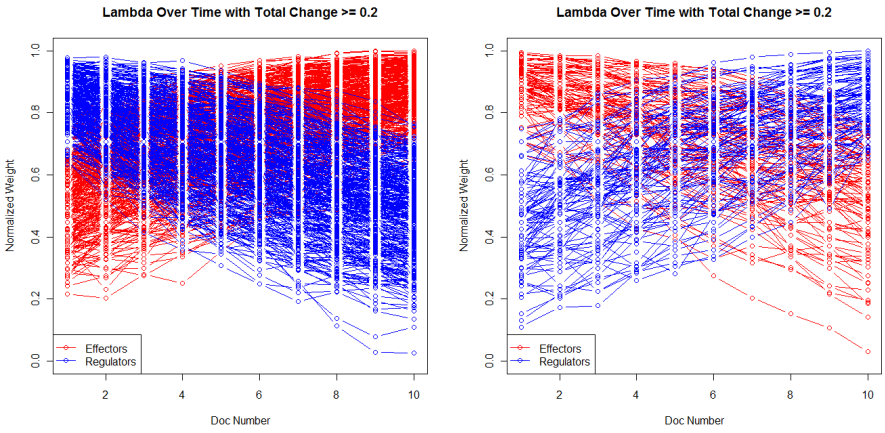


**Fig. 3.** Correct "lambda" feature score changes of >= .2 over 10 documents. Left: Set 1, Right: Set 2

| Table 3 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $N_{slot}$ | $E_0^+$ | $R_0^+$ | $E_0^-$ | $R_0^-$ | $E_0^u$ | $R_0^u$ | $D_E$ | $D_R$ |
| 10 | 5 | 12 | 5 | 3 | 5 | 3 | .1 | .1 |
| 10 | 6 | 12 | 6 | 2 | 6 | 2 | .1 | .1 |
| 10 | 7 | 13 | 7 | 1 | 7 | 1 | .1 | .1 |
| 10 | 8 | 10 | 8 | 5 | 8 | 5 | .1 | .1 |
| 11 | 11 | 11 | 11 | 2 | 11 | 2 | .1 | .1 |
| 12 | 6 | 12 | 6 | 3 | 6 | 3 | .1 | .1 |
| 12 | 7 | 10 | 7 | 4 | 7 | 4 | .1 | .1 |
| 12 | 9 | 11 | 9 | 3 | 9 | 3 | .1 | .1 |
| 13 | 5 | 14 | 5 | 4 | 5 | 4 | .1 | .1 |
| 13 | 6 | 15 | 6 | 3 | 6 | 3 | .1 | .1 |

To understand what makes the parameter configurations in Table 3 perform better than others, I attempted a mathematical analysis of the system under these artificial conditions. The equations for the expected change in feature t-cell populations between documents (or discrete time intervals) were fairly easy to write. However, in trying to investigate under what conditions the parameters would cause an expected correct change in "lambda"'s score between the first and second document, I made a number of simplifying assumptions that were incorrect. First, I assumed that there would be no empty slots on the APCs, which holds only for some of the above configurations. I also assumed that the populations of "lambda" at the beginning of second step would have changed negligibly from its initial population. Finally I assumed that the populations of "lambda" would have a negligible impact on the overall regulator and effector populations. While this simplified the mathematics, the resulting conditions have no bearing on the above runs of the classifier, so I won't include further details here.

**Discussion of Future Directions**

I will emphasize here that I recognize that my work so far can draw no clear conclusions. First, my qualitative descriptions of the distributions of mean E(d) and R(d) scores is not the result of systematic analysis across a complete picture of the parameter space, or even the subspace that I have explored. Words that appear only once are included in the distributions, over-emphasizing the peaks around the initial biases. It will be necessary to go back and normalize a feature's contribution to the distribution by its appearances in the documents. To draw any accurate observations from these distributions I will need to systematically compare them across runs. I am not sure yet how to do this and welcome advice, but I am planning to simply compare the differences in the regulator and effector distributions (calculated by subtracting bins) to the performance of the classifier, or looking directly at the separations in the initial population scores for each label. Similarly, only a small parameter subspace was explored over the artificial data, and I mean to run more expansive searches to see if regulators' disadvantage can be characterized. If I am successful in this, I will also need to investigate the effects of co-occurrences, since more than one word will be shared between real documents. I also need to put in the time to perform the mathematical analysis without over-simplifying assumptions. Through this I hope to find a measure of the system's sensitivity to the appearance of a common feature in documents of opposite labels. As a final note, to defend this work (or perhaps change my own mind about whether it is worth doing), I need to develop a better understanding of the Naïve Bayes classifier, and how this system differs.

**Works Cited**

[1] J. Carneiro, et al., “When three is not a crowd: a Crossregulation model of the dynamics and repertoire selection of regulatory CD4+ T cells.,” *Immunological Reviews*, vol. 216, pp. 48–68, 2007.

[2] A. Abi-Haidar and L. M. Rocha, “Collective Classification of Textual Documents by Guided Self-Organization in T-Cell Cross-Regulation Dynamics,” Evolutionary Intelligence, p. In press, 2011.