Evolving Hypernetworks for Language Modeling

AI Course Material
Oct. 12, 2009

Byoung-Tak Zhang

Biointelligence Laboratory
School of Computer Science and Engineering
Brain Science, Cognitive Science, Bioinformatics Programs
Seoul National University
Seoul 151-742, Korea

btzhang@cse.snu.ac.kr
http://bi.snu.ac.kr/
Outline

- Problem: Language Modeling
  - Data
  - Model
- Model: Hypernetwork
  - Individuals
  - Population
- Method: Evolving Hypernetworks
  - Variation
  - Selection
  - Amplification
  - Fitness Evaluation
- Experimental Results

Problem: Language Modeling
A Language Game

We ? ? a lot ? gifts.
⇒ We don't have a lot of gifts.

? still ? believe ? did this.
⇒ I still can't believe you did this.
Evolutionary Hypernets for Linguistic Memory

I'm gonna go upstairs and take a shower

? have ? visit the ? room
I have to visit the ladies' room

? still ? believe ? did this
I still can't believe you did this

? ?? decision
to make a decision

? appreciate it if ? call her by ? ?
I appreciate it if you call her by the way

Would you ? to meet ? ? Tuesday ?
Would you nice to meet you in Tuesday and

Why ? you ? come ? down ?
Why are you go come on down here


Data and Model

- Data \( D = \{ S_i \mid S_i \text{ are sentences} \} \)
  - Eg. Dialogue sentences from “Friends”
- Model \( M = \{ R_j \mid R_j \text{ are grammar rules} \} \)

- Generator \( g: S \times M \rightarrow S \)
- Learner \( L: D \rightarrow M \)

- Goal: to learn the grammar by evolution
1. Sentences

<table>
<thead>
<tr>
<th>1</th>
<th>x1</th>
<th>x4</th>
<th>x10</th>
<th>y=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>x1</td>
<td>x4</td>
<td>x12</td>
<td>y=1</td>
</tr>
<tr>
<td>3</td>
<td>x4</td>
<td>x10</td>
<td>x12</td>
<td>y=1</td>
</tr>
</tbody>
</table>

2. Many Micro Grammar Rules

<table>
<thead>
<tr>
<th>1</th>
<th>x2</th>
<th>x3</th>
<th>x9</th>
<th>y=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>x2</td>
<td>x3</td>
<td>x14</td>
<td>y=0</td>
</tr>
<tr>
<td>3</td>
<td>x3</td>
<td>x9</td>
<td>x14</td>
<td>y=0</td>
</tr>
</tbody>
</table>

3. Hyperedges

<table>
<thead>
<tr>
<th>1</th>
<th>x3</th>
<th>x6</th>
<th>x8</th>
<th>y=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>x3</td>
<td>x6</td>
<td>x13</td>
<td>y=1</td>
</tr>
<tr>
<td>3</td>
<td>x6</td>
<td>x8</td>
<td>x13</td>
<td>y=1</td>
</tr>
<tr>
<td>4</td>
<td>x8</td>
<td>x11</td>
<td>x15</td>
<td>y=0</td>
</tr>
</tbody>
</table>

4. Hypernet = Grammar (Population)

Hypernetwork Memory of Language

She was my best friend
this is my friend Rachel
so your best friend would want to do her
because I wanna be a good friend
I see two Monicas the one that was my friend

... my best friend 21
... is my friend 10
... have a boyfriend 9
your best friend 9
... a good friend as a friend 7
... have a friend 5
my friend rachel 5
... friend of 5
... my good friend 4
... was my friend 4

Filtering by keyword ‘friend’

(c) 2009 SNU Biointelligence Laboratory, http://bi.snu.ac.kr/
Evolving Hypernetworks
Initial Library $L_0$

\[(x_1=0, x_2=1, x_3=1, y=1)\]

\[
\text{AAAAATGCATGCCG}
\]

\[(x_1=0, x_2=0, x_3=1, y=0)\]

\[
\text{AAAAACCAATCCCAAGGCATGCCG}
\]

\[(x_2=1, x_3=1, y=1)\]

\[
\text{AATTCGCCATGCCG}
\]

\[(x_2=1, y=0)\]

\[
\text{AATTCGCCATGCCG}
\]

\[(x_2=1, x_3=0, y=0)\]

\[
\text{AATTCAGGCCATGCCG}
\]

\[(x_1=0, x_2=0, y=0)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_1=0, x_2=0, x_3=1, y=1)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_1=0, x_2=1, y=1)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_2=1, x_3=0, y=0)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_1=0, x_2=1, y=0)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_2=0, x_3=1, y=1)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_1=0, x_2=0, x_3=0, y=0)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_1=0, x_2=0, x_3=0, y=1)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_2=0, x_3=1, y=0)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_1=0, x_2=0, x_3=1, y=0)\]

\[
\text{AAAACCATGCGG}
\]

\[(x_1=0, x_2=1, y=0)\]

\[
\text{AAAACCATGCGG}
\]

where:

\[
\begin{align*}
\text{AAAA} & \quad x_1 \\
\text{AATT} & \quad x_2 \\
\text{CC} & \quad x_3 \\
\text{ATGC} & \quad y
\end{align*}
\]
Hybridization

Example 1

(x₁=0, x₂=1, x₃=1, y=1)

AAACCAATTCGATTGATGCCTG

(x₁=0, x₂=0, x₃=1, y=0)

AAACCAATTCGAAATCCCGATGCCTG

(x₂=1, x₃=1, y=1)

AATTGGGCTTTTGGATGCCTG

(x₂=1, x₃=0, y=0)

AATTCGAAATCCCGATGCCTG

(x₂=1, y=0)

AATTCGATGCCTG

(x₁=0, x₂=1, x₃=0, y=0)

TTTGG GTTACG TTTCCG TACCGC

(x₁=0, x₂=0, x₃=1, y=0)

AAACCAATTCGATTGATGCCTG

(x₂=1, x₃=1, y=1)

AATTGGGCTTTTGGATGCCTG

(x₂=1, x₃=0, y=0)

AATTCGAAATCCCGATGCCTG

(x₂=1, y=0)

AATTCGATGCCTG

Amplify

Updated Library $L_1$

where

\[ \begin{array}{ccc}
\text{AAAA} & x_1 & \text{ATGC} \\
\text{AATT} & x_2 & \text{CC} \\
\text{AAGG} & x_3 & \text{GG}
\end{array} \]

\[ y \]

\[ \begin{array}{c}
0 \\
1
\end{array} \]
Example 2

\[(x_1=0, x_2=1, x_3=1, y=1)\]

Library

\[
\begin{align*}
(x_1=0, x_2=0, x_3=1, y=0) & \quad \text{AAAACCAATTGGAATTGGATGCGG} \\
(x_2=1, x_3=1, y=0) & \quad \text{AAAATTTGGAATTGGATGCGG} \\
(x_2=1, x_3=0, y=0) & \quad \text{TTTTGG} \\
(x_2=1, y=0) & \quad \text{TTAACC}
\end{align*}
\]

Example 2

\[
\begin{align*}
(x_1=0, x_2=1, x_3=1, y=1) & \quad \text{AAAACCAATTGGAATTGGATGCGG} \\
(x_1=0, x_2=0, x_3=1, y=0) & \quad \text{AAAACCAATTGGAATTGGATGCGG} \\
(x_1=0, x_2=0, x_3=1, y=0) & \quad \text{AAAACCAATTGGAATTGGATGCGG} \\
(x_2=1, x_3=1, y=1) & \quad \text{AAAACCAATTGGAATTGGATGCGG}
\end{align*}
\]

Amplify

\[
\begin{align*}
(x_1=0, x_2=1, x_3=1, y=1) & \quad \text{AATTGGCCTTGGATGCGG} \\
(x_1=0, x_2=1, x_3=1, y=1) & \quad \text{AATTGGCCTTGGATGCGG} \\
(x_1=0, x_2=0, x_3=1, y=0) & \quad \text{AATTGGCCTTGGATGCGG} \\
(x_1=0, x_2=0, x_3=1, y=0) & \quad \text{AATTGGCCTTGGATGCGG}
\end{align*}
\]

\[
\begin{align*}
(x_2=1, x_3=1, y=1) & \quad \text{AATTGGCCTTGGATGCGG} \\
(x_2=1, x_3=0, y=0) & \quad \text{AATTGGCCTTGGATGCGG} \\
(x_2=1, x_3=0, y=0) & \quad \text{AATTGGCCTTGGATGCGG} \\
(x_2=1, y=0) & \quad \text{AATTGGCCTTGGATGCGG}
\end{align*}
\]

\[
\begin{align*}
(x_2=1, x_3=1, y=1) & \quad \text{AAAATTTGGAATTGGATGCGG} \\
(x_2=1, x_3=0, y=0) & \quad \text{AAAATTTGGAATTGGATGCGG} \\
(x_2=1, x_3=0, y=0) & \quad \text{AAAATTTGGAATTGGATGCGG} \\
(x_2=1, y=0) & \quad \text{AAAATTTGGAATTGGATGCGG}
\end{align*}
\]

\[
\begin{align*}
(x_2=1, x_3=1, y=1) & \quad \text{TTTTGG} \\
(x_2=1, x_3=0, y=0) & \quad \text{TTTTGG} \\
(x_2=1, x_3=0, y=0) & \quad \text{TTTTGG}
\end{align*}
\]

\[
\begin{align*}
(x_2=1, y=0) & \quad \text{TTAACC} \\
(x_2=1, y=0) & \quad \text{TTAACC}
\end{align*}
\]
Updated Library $L_2$

\[ (x_1=0, x_2=1, x_3=1, y=1) \]

\[ (x_1=0, x_2=0, x_3=1, y=0) \]

\[ (x_2=1, x_3=0, y=0) \]

where

\[ \begin{align*}
A & \quad x_1 \\
A & \quad ATGC \\
A & \quad x_2 \\
A & \quad AAGG \\
A & \quad x_3 \\
A & \quad CC \\
A & \quad y
\end{align*} \]

Hybridization

Library Query

Predict the class

Majority voting

Molecular Programming (MP): The Evolutionary Learning Algorithm

1. Let the library \( L \) represent the current distribution \( P(X,Y) \).
2. Get a training example \((x,y)\).
3. Classify \( x \) using \( L \) as follows
   3.1 Extract all molecules matching \( x \) into \( M \).
   3.2 From \( M \) separate the molecules into classes:
      Extract the molecules with label \( Y=0 \) into \( M^0 \)
      Extract the molecules with label \( Y=1 \) into \( M^1 \)
   3.3 Compute \( y^* = \arg\max_{Y \in \{0,1\}} |M^Y| / |M| \)
4. Update \( L \)
   If \( y^* = y \), then \( L_n \leftarrow L_{n-1} + \{\Delta c(u, v)\} \) for \( u=x \) and \( v=y \) for \((u, v) \in L_{n-1}\),
   If \( y^* \neq y \), then \( L_n \leftarrow L_{n-1} - \{\Delta c(u, v)\} \) for \( u=x \) and \( v \neq y \) for \((u, v) \in L_{n-1}\)
5. Goto step 2 if not terminated.

[Zhang, GECCO-2005]
Learning the Hypernetwork (by Evolutionary Self-assembly)

- Next generation
- Library of combinatorial molecules
- Select the library elements matching the example
- Hybridize

Amplify the matched library elements by PCR

[Zhang, DNA11]

Animation for Molecular Evolutionary Learning

- MP4.avi
Experimental Results
The Language Game Platform

I don't know what happened. Take a look at this.


289,468 Sentences (Training Data)

700 Sentences with Blanks (Test Data)
Hypernetwork Memory of Language

... She was my best friend
this is my friend Rachel
so your best friend would want to do her
because I wanna be a good friend
I see two Monicas the one that was my friend ...

(c) 2009 SNU Biointelligence Laboratory, http://bi.snu.ac.kr/
Evolutionary Hypernets for Linguistic Memory

I'm gonna go upstairs and take a shower

? have ? visit the ? room
I have to visit the ladies' room

? still ? believe ? did this
I still can't believe you did this

? ? ? decision
to make a decision

? appreciate it if ? call her by ? ?
I appreciate it if you call her by the way

Would you ? to meet ? ? Tuesday ?
Would you nice to meet you in Tuesday and

Why ? you ? come ? down ?
Why are you go come on down here


## Memories for *Friends* and *Prison Break*

<table>
<thead>
<tr>
<th>Corpus: <em>Friends</em></th>
<th>Corpus: <em>Prison Break</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Keyword:</strong> “mother”</td>
<td><strong>Keyword:</strong> “mother”</td>
</tr>
<tr>
<td>you're mother killed herself</td>
<td>tells his mother and his family</td>
</tr>
<tr>
<td>it's my mother was shot by a woman at eight</td>
<td>she's the mother of my eyes</td>
</tr>
<tr>
<td>we're just gonna go to your mother that I love it</td>
<td>speak to your mother used to be</td>
</tr>
<tr>
<td>feeling that something's wrong with my mother and father</td>
<td>tells his mother made it pretty clear on the floor has</td>
</tr>
<tr>
<td>she's the single mother</td>
<td>speak to your mother never had life insurance</td>
</tr>
<tr>
<td>i put this on my friend's mother</td>
<td>she's the mother of lincoln's child</td>
</tr>
<tr>
<td>apparently phoebe's mother killed herself</td>
<td>she's the mother of my own crap to deal with you</td>
</tr>
<tr>
<td>thanks for pleasing my mother killed herself</td>
<td>just lost his mother is fine</td>
</tr>
<tr>
<td>i'm your mother told you this</td>
<td>just lost his mother and his god</td>
</tr>
<tr>
<td>is an incredible mother</td>
<td>tells his mother and his stepfather</td>
</tr>
<tr>
<td>that's not his mother or his hunger strike</td>
<td>she's the mother of my time</td>
</tr>
<tr>
<td>holy mother of god woman</td>
<td>his mother made it clear you couldn't deliver fibonacci</td>
</tr>
<tr>
<td>i like your mother and father on their honeymoon suite</td>
<td>she's the mother of my brother is facing the electric chair</td>
</tr>
<tr>
<td>with her and never called your mother really did like us</td>
<td>same guy who was it your mother before you do it</td>
</tr>
<tr>
<td>is my mother was shot by a drug dealer</td>
<td>they gunned my mother down</td>
</tr>
</tbody>
</table>

(c) 2009 SNU Biointelligence Laboratory, [http://bi.snu.ac.kr/](http://bi.snu.ac.kr/)
Learning Languages from Kids Video

Goal: (1) Natural language generation at sentence level based on the probabilistic graphical model, and (2) Natural language processing without the explicit grammar rules.

<table>
<thead>
<tr>
<th>Training data</th>
<th>Kids video scripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Script sequence</td>
<td>Generated sentence</td>
</tr>
<tr>
<td>Timothy</td>
<td>I like it too nora.</td>
</tr>
<tr>
<td>Hello kitty</td>
<td>I like it too mom.</td>
</tr>
<tr>
<td>Looney toons</td>
<td>I like it too this time you’re a diving act today.</td>
</tr>
<tr>
<td>Dora Dora</td>
<td>I like it too this time you’re a hug.</td>
</tr>
</tbody>
</table>

Sentence structure
Converting sentences into graph structure

Application
Sentence completion and generation

© 2009 SNU CSE Biointelligence Lab
Generated Sentences and Evolved Grammar

- **Generated sentences**
  - (Good) On my first day of school
  - (Good) Yes timothy it is time to go to school
  - (Good) Thomas and Percy enjoy working in the spotlight
  - (Good) Well it is morning
  - (Bad) He couldn’t way to go outside and shoot
  - (Bad) the gas house gorillas are a lot of fun players

- **Grammar rules analyzed from the generated sentences**
  - G1: $S = NP + VP$
  - G2: $NP = PRP$
  - G3: $S = VP$
  - G4: $PP = IN + NP$
  - G5: $NP = NN$
  - G6: $NP = DP + NN$
  - G7: $ADVP = RB$
  - G8: $NP = NP + PP$
  - G9: $SBAR = S$
Sentence Generation Accuracy

- Corpus: scripts from kids video (Miffy, Looney, Caillou, Dora Dora, Macdonald, Thoams & Friends, Timothy, Pooh)
- Corpus: Video scripts (kids video + sitcom Friends, 120K sentences)
- In each phase, corpus size is incremented by addition of a video script.
- Learning: building a language model based on a hypernetwork.
- Task: Sentence completion from a partial sentence.

Sentence generation result

- Rate of grammatically plausible sentences

- $D^1 = $ Miffy, $D^2 = D^1 + $ Looney,
- $D^3 = D^2 + $ caillou, $D^4 = D^3 + $ Dora Dora
- $D^5 = D^4 + $ Macdoland, $D^6 = D^5 + $ Thomas,
- $D^7 = D^6 + $ Timothy, $D^8 = D^7 + $ Pooh,
- $D^9 = D^8 + $ Friends

© 2009 SNU CSE Biointelligence Lab
Evolution of Grammar Rules

KL divergence between the distribution of training corpus ($P$) and the generated sentences ($Q$).

$$D_{KL}(P||Q) = \sum_{i} P(i) \log \frac{P(i)}{Q(i)}$$

• The right curve shows occurrence number of grammar rules are increasing as training progresses.
DNA Computing and DNA Nanotechnology: An Introduction
Hypercenetworks: More Details
Hypergraphs

- A hypergraph is a (undirected) graph $G$ whose edges connect a non-null number of vertices, i.e. $G = (V, E)$, where
  $$V = \{v_1, v_2, \ldots, v_n\},$$
  $$E = \{E_1, E_2, \ldots, E_n\},$$
  and $E_i = \{v_{i1}, v_{i2}, \ldots, v_{im}\}$

- An $m$-hypergraph consists of a set $V$ of vertices and a subset $E$ of $V^m$, i.e. $G = (V, V^m)$ where $V^m$ is a set of subsets of $V$ whose elements have precisely $m$ members.

- A hypergraph $G$ is said to be $k$-uniform if every edge $E_i$ in $E$ has cardinality $k$.

- A hypergraph $G$ is $k$-regular if every vertex has degree $k$.

- Rem.: An ordinary graph is a 2-uniform hypergraph.
An Example Hypergraph

\[ G = (V, E) \]
\[ V = \{v_1, v_2, v_3, \ldots, v_7\} \]
\[ E = \{E_1, E_2, E_3, E_4, E_5\} \]

\[ E_1 = \{v_1, v_3, v_4\} \]
\[ E_2 = \{v_1, v_4\} \]
\[ E_3 = \{v_2, v_3, v_6\} \]
\[ E_4 = \{v_3, v_4, v_6, v_7\} \]
\[ E_5 = \{v_4, v_5, v_7\} \]
A hypernetwork is a hypergraph of weighted edges. It is defined as a triple $H = (V, E, W)$, where

$V = \{v_1, v_2, \ldots, v_n\}$,

$E = \{E_1, E_2, \ldots, E_n\}$,

and $W = \{w_1, w_2, \ldots, w_n\}$.

An $m$-hypernetwork consists of a set $V$ of vertices and a subset $E$ of $V^m$, i.e. $H = (V, V^m, W)$ where $V^m$ is a set of subsets of $V$ whose elements have precisely $m$ members and $W$ is the set of weights associated with the hyperedges.

A hypernetwork $H$ is said to be $k$-uniform if every edge $E_i$ in $E$ has cardinality $k$.

A hypernetwork $H$ is $k$-regular if every vertex has degree $k$.

Rem.: An ordinary graph is a 2-uniform hypergraph with $w_i=1$. 

[Zhang, 2006, in preparation]
A Hypernetwork
The Hypernetwork Model of Learning

The hypernetwork is defined as

\[ H = (X, S, W) \]
\[ X = (x_1, x_2, ..., x_J) \]
\[ S = \sum_i S_i, \quad S_i \subseteq X, \quad k = |S_i| \]
\[ W = (W^{(2)}, W^{(3)}, ..., W^{(K)}) \]

Training set:

\[ D = \{ x^{(n)}_{i1} \}^N_{i=1} \]

The energy of the hypernetwork

\[ E(x^{(n)}; W) = -\frac{1}{2} \sum_{i_1,i_2} W^{(2)}_{i_1,i_2} x^{(n)}_{i_1} x^{(n)}_{i_2} - \frac{1}{6} \sum_{i_1,i_2,i_3} W^{(2)}_{i_1,i_2,i_3} x^{(n)}_{i_1} x^{(n)}_{i_2} x^{(n)}_{i_3} - ... \]

The probability distribution

\[ P(x^{(n)} | W) = \frac{1}{Z(W)} \exp\left[ -\beta E(x^{(n)}; W) \right] \]

\[ = \frac{1}{Z(W)} \exp \left[ \frac{1}{2} \sum_{i_1,i_2} W^{(2)}_{i_1,i_2} x^{(n)}_{i_1} x^{(n)}_{i_2} + \frac{1}{6} \sum_{i_1,i_2,i_3} W^{(2)}_{i_1,i_2,i_3} x^{(n)}_{i_1} x^{(n)}_{i_2} x^{(n)}_{i_3} + ... \right] \]

\[ = \frac{1}{Z(W)} \exp \left[ \sum_{k=2}^K \frac{1}{c(k)} \sum_{i_1,i_2,...,i_k} W^{(k)}_{i_1,i_2,...,i_k} x^{(n)}_{i_1} x^{(n)}_{i_2} ... x^{(n)}_{i_k} \right], \]

where the partition function is

\[ Z(W) = \sum_{x^{(n)}} \exp \left[ \sum_{k=2}^K \frac{1}{c(k)} \sum_{i_1,i_2,...,i_k} W^{(k)}_{i_1,i_2,...,i_k} x^{(m)}_{i_1} x^{(m)}_{i_2} ... x^{(m)}_{i_k} \right] \]

[Zhang, 2006, in preparation]
Deriving the Learning Rule

\[
P(\{x^{(n)}\}_{n=1}^{N} | W) = \prod_{n=1}^{N} P(x^{(n)} | W)
\]

\[
\ln P(\{x^{(n)}\}_{n=1}^{N} | W) = \ln \prod_{n=1}^{N} P(x^{(n)} | W^{(2)}, W^{(3)}, ..., W^{(K)})
\]

\[
= \sum_{n=1}^{N} \left\{ \exp \left[ \sum_{k=2}^{K} \frac{1}{c(k)} \sum_{i_1, i_2, ..., i_k} w^{(k)}_{i_1i_2...i_k} x^{(n)}_{i_1} x^{(n)}_{i_2} ... x^{(n)}_{i_k} \right] - \ln Z(W) \right\}
\]

\[
\frac{\partial}{\partial w^{(s)}_{i_1i_2...i_k}} \ln P(\{x^{(n)}\}_{n=1}^{N} | W)
\]

Derivation of the Learning Rule

\[
\frac{\partial}{\partial w^{(s)}_{h_i l_2 \ldots l_s}} \ln P \left( \{ x^{(n)} \} \right) = \sum_{n=1}^{N} \left\{ \exp \left[ \sum_{k=2}^{K} \frac{1}{c(k)} \sum_{i_1, i_2, \ldots, i_k} w^{(k)}_{i_1 i_2 \ldots i_k} x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_k} \right] - \ln Z(W) \right\} \\
= \sum_{n=1}^{N} \left\{ \frac{\partial}{\partial w^{(s)}_{h_i l_2 \ldots l_s}} \left[ \sum_{k=2}^{K} \frac{1}{c(k)} \sum_{i_1, i_2, \ldots, i_k} w^{(k)}_{i_1 i_2 \ldots i_k} x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_k} \right] - \frac{\partial}{\partial w^{(s)}_{h_i l_2 \ldots l_s}} \ln Z(W) \right\} \\
= \sum_{n=1}^{N} \left\{ x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_s} - \left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{P(x|W)} \right\} \\
= N \left\{ \left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{Data} - \left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{P(x|W)} \right\}
\]

where

\[
\left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{Data} = \frac{1}{N} \sum_{n=1}^{N} \left[ x^{(n)}_{i_1} x^{(n)}_{i_2} \ldots x^{(n)}_{i_s} \right] \\
\left\langle x_{i_1} x_{i_2} \ldots x_{i_s} \right\rangle_{P(x|W)} = \sum_{x} \left[ x_{i_1} x_{i_2} \ldots x_{i_s} \right] P(x \mid W)
\]
Molecular Self-Assembly of Hypernetworks

Hypernetwork Representation

Molecular Encoding

Encoding Hyperedges with DNA

a) \[ z_1 : (x_1=0, x_2=1, x_3=0, y=1) \]
\[ z_2 : (x_1=0, x_2=0, x_3=1, x_4=0, x_5=0, y=0) \]
\[ z_3 : (x_2=1, x_4=1, y=1) \]
\[ z_4 : (x_2=1, x_3=0, x_4=1, y=0) \]

Collection of (labeled) hyperedges

b) \[ z_1 : \text{AAAAACCAATTGGAAGGCCATGCGG} \]
\[ z_2 : \text{AAAAACCAATTGGAAGGCCCTTCCCAACCATGCCC} \]
\[ z_3 : \text{AATTGCGCTTGGATGCGG} \]
\[ z_4 : \text{AATTGGAAGGCCTTGGATGCTC} \]

Library of DNA molecules corresponding to (a)

where

\[
\begin{array}{cccc}
\text{AAAA} & x_1 & \text{CCTT} & x_4 & \text{CC} & 0 \\
\text{AATT} & x_2 & \text{CCAA} & x_5 & \text{GG} & 1 \\
\text{AAGG} & x_3 & \text{ATGC} & y & & \\
\end{array}
\]
The Theory of Bayesian Evolution

- Evolution as a Bayesian inference process
- Evolutionary computation (EC) is viewed as an iterative process of generating the individuals of ever higher posterior probabilities from the priors and the observed data.

\[ P(A|D) \quad \text{generation 0} \]

\[ P(A|D) \quad \text{generation } g \]

\[ P_0(A_i) \]

\[ \Rightarrow \quad \ldots \quad \Rightarrow \]

\[ P_g(A_i) \]

\[ P_g(A_i|D) \]

[Zhang, CEC-99]
Unconventional Computing

- Quantum Computing
  - Atoms
  - Superposition, quantum entanglements
- Chemical Computing
  - Chemicals
  - Reaction-diffusion computing
- Molecular Computing
  - Molecules
  - “Evolutionary Hypernetworks”
## Molecular Computers vs. Silicon Computers

<table>
<thead>
<tr>
<th></th>
<th>Molecular Computers</th>
<th>Silicon Computers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Processing</strong></td>
<td>Ballistic</td>
<td>Hardwired</td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td>Liquid (wet) or Gaseous (dry)</td>
<td>Solid (dry)</td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>3D collision</td>
<td>2D switching</td>
</tr>
<tr>
<td><strong>Configuration</strong></td>
<td>Amorphous (asynchronous)</td>
<td>Fixed (synchronous)</td>
</tr>
<tr>
<td><strong>Parallelism</strong></td>
<td>Massively parallel</td>
<td>Sequential</td>
</tr>
<tr>
<td><strong>Speed</strong></td>
<td>Fast (millisecond)</td>
<td>Ultra-fast (nanosecond)</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Density</strong></td>
<td>Ultra-high</td>
<td>Very high</td>
</tr>
<tr>
<td><strong>Devices</strong></td>
<td>Unreliable</td>
<td>Reliable</td>
</tr>
</tbody>
</table>

DNA as “Programmable” Nanomatter

- **Information Density:**
  - $10^6$ Gbits per cm$^2$ (1 bit per nm$^3$)
  - Semiconductor: 1 Gbits per cm$^2$

- **Massive Parallelism:**
  - $10^{26}$ reactions per 1 mmol of DNA
  - Desktop: $10^9$ operations / sec
  - Supercomputer: $10^{12}$ operations / sec

- **Energy Consumption:**
  - $10^{19}$ operations per Joule
  - Semiconductor: $10^9$ operations per Joule
Properties of DNA Molecules

Repeat

↓ Heat

↓ Cool

↓ Polymer

Self-replication

Self-assembly

Molecular recognition