

## توضیح پروژه

در این پروژه قصد داریم رمزگشایی یک متن رمزگذاری شده را با الگوریتم ژنتیک انجام دهیم.

همانطور که در صورت پروژه گفته شد، رمزنگاری بدین صورت انجام می شود که هر حرف الفبا را به یک حرف دیگر مپ می کند و برای decode کردن می بایست کلید درست متن رمز شده را پیدا کنیم تا بتوانیم با استفاده از آن هر حرف متن رمزنگاری شده را

با معادل درست آن جایگذاری کرده و به متن درست برسیم.

روند کلی به این صورت است که ابتدا پارامترهای مساله شامل کروموزوم و ژن را تعریف کرده و سپس به تشریح روش های به کار برده شده برای cross، mutation و ... می پردازیم.

## تعریف کروموزوم

گام اول برای حل یک مسئله جست و جو به روش ژنتیک تعریف کروموزوم های مسئله می باشد. کروموزوم یک راه حل برای مسئله می باشد و از تعدادی ژن تشکیل شده است که بیانگر خصوصیات راه حل یا کروموزوم می باشند.

کروموزوم در این مسئله نیز یک راه حل برای آن می باشد که این راه حل همان کلید متن است و با ساختمان داده دیکشنری طراحی شده و از 26 ژن تشکیل شده است که هر ژن کاراکتری از متن را به یک حرف الفبای انگلیسی مپ کرده است.

به عنوان مثال اگر یک ژن که بخشی از کروموزوم است با a:o نشان داده شده است به این معنا می باشد که حرف a در متن رمزنگاری شده با o جایگذاری می شود تا متن decode شود.

## ساخت جمعیت اولیه

با خطا و آزمون این گونه به نظر می رسد که جمعیت اولیه انتخابی با 96 کروموزوم مناسب است.

این 96 کروموزوم را با رندم گیری ایجاد کرده و همانطور که در تعریف کروموزوم گفته شد، هر کاراکتر را به یک کاراکتر حروف الفبا مپ می کنیم (توجه شود که هیچ دو key و همین طور هیچ دو value در دیکشنری کروموزوم نباید یکسان باشند چرا که هر کاراکتر متن باید به یک حرف الفبا و همینطور هر حرف الفبا فقط به یک کاراکتر متن نسبت داده شود.)

پس از ساخت جمعیت اولیه روند ایجاد جمعیت ها تا زمان رسیدن به کروموزوم انتخابی شروع می شود.

---

---

گام اول الگوریتم ژنتیک روی هر نسل، اختصاص نمره **fitness** به هر کروموزوم جمعیت می باشد.

### تابع **fitness**

این تابع به هر کروموزوم جمعیت یک نمره نسبت می دهد. سپس کروموزوم ها بر اساس این نمره برای حضور در نسل بعد ارزیابی می شوند.

نحوه محاسبه نمره در این تابع:

به ازای هر کلمه موجود در متن رمزنگاری شده، کاراکتر های آن را با کاراکتر هایی که دیکشنری کروموزوم دستور به جایگزینی می کند، جایگزین می کنیم (به عنوان مثال همانطور که گفته شد **a:o** در دیکشنری به معنای این است که کاراکتر **a** در متن رمزنگاری شده باید با **a** تعویض شود).

امتیاز برابر با تعداد کلمات پیدا شده در لغت نامه (تعداد کلمات درست پیدا شده) می باشد.

پس از اختصاص نمره به کروموزوم ها باید جمعی از آن ها برابر با جمعیت اولیه را برای حضور در نسل بعد انتخاب کنیم.

### تابع **parent selection**

روش های زیادی برای انتخاب والد آزمون و خطا کردیم. روشی را که در زمان معقول تری الگوریتم را هدایت می کرد توضیح خواهیم داد.

در هر نسل پس از اختصاص دهی نمره به کروموزوم ها آن ها را بر اساس نمره **fitness** مرتب می کنیم و  $\frac{1}{4}$  برتر را از میان آن ها انتخاب می کنیم و با چندین مرحله تولید مثل (**cross**) نسل آن ها را گسترش می دهیم تا به تعداد جمعیت نسل مورد نظر (همان تعداد جمعیت نسل اولیه) برسیم.

پس از انتخاب والد ها بدو به دو آن ها را به احتمال **pc** کراس کنیم.

### **Cross over**

برای کراس کردن دو پرنس و تولید دو بچه از آن ها، والد اول را در فرزند اول و والد دوم را در فرزند دوم کپی می کنیم.

حال از آنجا که فرزند اول تمامی ژن های خود را از پرنس اول دارد و فرزند دوم نیز تمام ژن های خود را از پرنس دوم، می بایست سیاستی طراحی کنیم که ترکیب ژن ها در دو فرزند تعدیل شود و هر دو فرزند هم از والد اول و هم از والد دوم ژن داشته باشند.

برای این منظور، 7 عدد رندم در بازه 0 تا 25 برای انتخاب حرف الفبا انتخاب می کنیم (عدد 0 به معنای حرف 'a'، عدد 1 به معنای حرف 'b' و...).  

---

حروف مربوط به این 7 عدد را در فرزند اول که تمامی ژن های خود را از والد اول داشت، با حروف متناظر در والد دوم جایگزین می کنیم (به عنوان مثال در فرزند اول a:o از والد اول رسیده و اگر عدد رندم 0 انتخاب شود و a:z در والد دوم باشد، فرزند باید آن را از والد 2 بگیرد و جایگزین کند. بدیهی است که در صورت برداشتن ژن a:z، فرزند اول باید z:? را با o:? جایگزین کند). توجه شود که با هر کدام از 7 عدد رندوم حداکثر 14 جایگزینی در ژن ها رخ می دهد و هم چنان 12 ژن از والد اصلی (والدی که نخست همه ی ژن هایش را گرفت) در فرزند باقی می ماند.

همین عملاً متناظر در مورد فرزند دوم که ژن های والد دوم را داراست اتفاق می افتد.

بدین ترتیب ترکیب ژن های دو فرزند تعدیل شده و دو فرزند ترکیبی از ژن های دو والد را به ارث خواهند برد.

حال دو فرزند حاصل کراس را در چند مرحله با احتمال mutate pm می کنیم.

## mutation

برای mutate کردن فرزندان تولید شده در مرحله قبل، در 4 مرحله، 4 جفت ژن انتخاب می کنیم تا مقادیر آن ها را تعویض کنیم (مثلاً دو جفت ژن s:t و r:y در یک فرزند که پس از اعمال mutation با احتمال pm دو جفت s:y و r:t را خواهیم داشت).

پس از اعمال mutation فرزندان تولید شده که تعدادشان برابر با تعداد نسل اولیه است به نسل کنونی منتقل شده و همه مراحل دوباره روی آن ها بررسی می شود.

در هر زمان که کروموزوم با fitness برابر با سایز کلمات متن رمز شده پیدا شود، decode خاتمه یافته و کلید درست پیدا می شود.

## تمیز کردن داده ها

برای تمیز کردن لغت نامه مراحل زیر را انجام می دهیم:

- 1) متن داده ها را از punctuation ها شامل =, و... خالی می کنیم (آن ها را با white space جایگزین می کنیم).
- 2) White space های پشت سر هم چندتایی را یکی می کنیم.
- 3) داده ها را lowercase می کنیم.
- 4) داده ها را با space جدا (split) می کنیم تا کلمه کلمه شوند و در یک لیست از کلمات نگه می داریم.
- 5) کلمات تکراری را از لیست حذف می کنیم.
- 6) Stop word ها را که شامل کلماتی می شوند که معمولاً در یک متن مکرر ظاهر می شوند را حذف می کنیم (در زبان انگلیسی واژگانی مانند the و ...)

در روش بکار گرفته شده برای تابع fitness این الگوریتم، از آن جا که fitness برابر با تعداد کلمات متن رمزگذاری شده حاضر در لغت نامه می باشد و ممکن است با حذف stop word های لغت نامه هیچ گاه یک کلمه رمزگذاری شده را در لغت نامه پیدا نکنیم، از حذف stop word ها خودداری کرده ایم.

**اگر تعداد جمعیتی را که در هر دوره نگه میداریم، افزایش دهیم، چه تاثیری بر روی سرعت و دقت میگذارد؟**

اگر در هر مرحله تعداد اعضای نسل را بیشتر کنیم، مدت زمان مصرف شده برای محاسبه توابع محاسبه fitness،

Cross و mutation افزایش می یابد و از طرفی باید مدام ضرایب و ثابت های اتخاذ شده را تغییر دهیم. از طرفی سیاست ما در واقع این بود که در هر مرحله جمعیت کنونی را بهتر کنیم و به مرحله بعد انتقال دهیم. حال اینکه با تغییر جمعیت ممکن است کروموزوم هایی را وارد رقابت کنیم که موثر نیستند و غیر قابل پیش بینی هستند و مدام بین نسل ها منتقل می شوند که زمان الگوریتم را طولانی و حافظه مصرفی را زیاد می کنند.

**به نظر شما تاثیر mutation چیست؟ اگه فقط از cross over استفاده بشود، چه مشکلی پیش میآید؟**

در mutation یک ژن به چیزی که از والد خود لزوما نگرفته است تغییر می کند. یعنی در هر بار عمل mutation یک جهش در یک یا چند ژن کروموزوم رخ می دهد که می تواند جهش مثبت از جهت نزدیک شدن به کروموزوم نهایی یا دور شدن از آن باشد.

در cross over فرزندان تولید شده همه ژن های خود را از والدین خود به ارث می برند و شانس یک ژن متفاوت با ژن های نسل های خود را ندارند و همواره مجبور به ارث بری از والدین خود دارند.

اما mutation می تواند یک جهش مثبت در فرزند به وجود بیاورد که او را به هدف نهایی نزدیک کند(ممکن است ژنی را به او بدهد که هیچ یک از اجداد او نداده اند). پس تجربه و پذیرش ریسک mutation با احتمال کم می تواند موثر واقع شود.

**Mutation موثر تر است یا cross over؟ کدام باعث سریعتر بالا رفتن دقت میشود؟**

Cross over از آنجا که در تلاش است تا ژن های خوب والدین را به فرزندان بدهد معمولاً بهتر عمل می کند و دقت بالاتری دارد(کروموزوم را سریع تر به هدف نهایی نزدیک تر می کند). یعنی cross over با ترکیب والدین فرزند را می سازد و احتمال یافته شدن ژن های خوب در فرزند را بالا می برد اما وجود آن به تنهایی کافی نیست و گاهی می بایست برای نزدیک شدن به هدف ریسک جهش در ژن ها را که توسط mutation انجام می شود و به ژن های والدین ربطی ندارد، تجربه کرد.

ممکن است اجداد یک فرزند از ابتدا ژن مناسب نداشته یا کم داشته باشند و هر چه cross over انجام دهیم به هدف خیلی نزدیک نشویم، پس mutation در این مواقع سودمند است.

پس به طور کل و با تقریب خوبی cross over موثر تر است و احتمال بالایی می گیرد(pc بالا) اما برای mutation هم شانس کمی قائل هستیم(pm).

---

با استفاده از این روش، باز ممکن است کروموزومهایتان بعد از چند مرحله دیگر تغییر نکنند. چرا این اتفاق رخ میدهد؟  
این سوگیری کروموزومها چه مشکلی پیش میآورد؟ برای حل آن چه راهحلی پیشنهاد میدهید؟

حتی با وجود انتخاب والدین با fitness بالا، ممکن است یک ژن در بین آن ها وجود نداشته باشد که در کروموزوم جواب حاضر است. همین طور ممکن است بعد از تعداد زیادی cross ژن های مشابه در کروموزوم ها دیده شود و کروموزوم های تقریباً مشابه به وجود بیایند. یعنی روند تولید کروموزوم ها در نسل ها به سمت کروموزوم های مشابه برود که امتیاز والد بالایی داشته اند و این یعنی کروموزوم های با نمره پایین از دور رقابت کنار گذاشته می شوند و ممکن است در کروموزوم های حاضر نقصی باشد که یک کروموزوم بد ح آن نقص را نداشت و این موضوع خلل ایجاد می کند.

پس دلیل سوگیری می تواند تابع fitness بدی باشد که بسیاری از کروموزوم ها را بی ارزش نشان دهد و از سوی دیگر تابع انتخاب والدی باشد که تنها به کروموزوم های با نمره بالا شانس شرکت در cross دهد.

راه حل هایی برای رفع این مشکل اتخاذ شده اند.

- 1) به عنوان مثال شانس mutation همان طور که گفته شد بدهیم تا بافت کروموزوم های جدید مشاهده کنیم
- 2) ترتیب لیست پدر های انتخاب شده را بهم بریزیم تا والد های از نسل های گوناگون و با ژن های متفاوت کنار هم قرار گیرند تا بتوانند cross over کنند.
- 3) تغییرات cross over و mutation را همواره روی تعدادی ژن خاص اعمال نکنیم و از انتخاب تصادفی و رندم برای انتخاب ژن های متاثر بهره بگیریم. مثلاً در یک کروموزوم ژن های 1 5 7 و در دیگری 18 4 9 را تغییر بدهیم تا ترکیب ژن ها و مشابهت آن ها در کروموزوم ها چه در cross over و چه در mutation رو به یکسان شدن نرود.

```
Python 3.7.2 Shell
File Edit Shell Debug Options Window Help

>That you don't like what God told people to do says nothing about God
>or God's commands, it says only that there was an electrical event in your
>nervous system that created an emotional state that your mind coupled
>with a pre-existing thought-set to form that reaction.

I think you've been reading the wrong sort of comic books, but in
prying through the gobbledegook I basically agree with what you're
saying. I do believe that my mental reactions to stimuli such as "God
commanded the genocide of the Canaanites" is mechanistic, but of
course I think that's true of you as well. My reaction has little to
do with whether God exists or even with whether I think he does, but
if a god existed who commanded genocide, I could not consider him
good, which is supposedly an attribute of God.

>All of this being so, you have excluded
>yourself from any discussion of values, right, wrong, good, evil,
>etc. and cannot participate. Your opinion about the Bible can have no
>weight whatsoever.

Hmm. Yes, I think some heavy FAQ-reading would do you some good. I
have as much place discussing values etc. as any other person. In
fact, I can actually accomplish something in such a discussion, by
framing the questions in terms of reason: for instance, it is clear
that in an environment where neighboring tribes periodically attempt
to wipe each other out based on imagined divine commands, then the
quality of life will be generally poor, so a system that fosters
coexistence is superior, if quality of life is an agreed goal. An
absolutist, on the other hand, can only thump those portions of a
Bible they happen to agree with, and say "this is good", even if the
act in question is unequivocally bad by the standards of everyone in
the discussion. The attempt to define someone or a group of people as
"excluded from discussion", such that they "cannot participate", and
their opinions given "no weight whatsoever" is the lowest form of
reasoning (ad hominem/poisoning the well), and presumably the resort
of someone who can't rationally defend their own ideas of right,
wrong, and the Bible.

--
Jim Perry perry@dsinc.com Decision Support, Inc., Matthews NC
These are my opinions. For a nominal fee, they can be yours.

--
Jim Perry perry@dsinc.com Decision Support, Inc., Matthews NC
These are my opinions. For a nominal fee, they can be yours.

0:00:24.251268
>>> |
```