به نام خدا

پروژه دوم - ژنتیک
هوش مصنوعی

طراحان: علی الهی، امیرمحمد رنجبرپازکی، بهزاد شایق
مهلت ارسال: نیمه شب ۱۶ فروردین

دانشکده مهندسی برق و کامپیوتر دانشگاه تهران
زمستان ۹۸

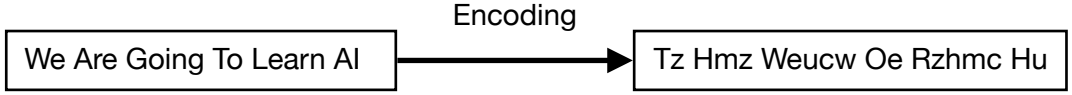
رمزگشایی رمزنگاری جایگزینی

هدف از رمزنگاری، تبدیل متن خام(پیام) به متن رمزشده است تا هیچکس جز مقصد پیام آن را نفهمد. هر چه بازگرداندن متن رمزشده به پیام اصلی از نظر زمانی، پیچیده تر باشد و بیشتر طول بکشد، رمزنگاری ارزشمندتر است.

یکی از روش‌های مرسوم رمزنگاری، رمزنگاری به روش جایگزینی است. در این روش، هر حرف به یک حرف دیگر نگاشت داده می‌شود و در پیام اصلی جایگزین آن قرار داده می‌شود تا رمز موردنظر به‌دست‌آید. به این نگاشت ۲۶ حرفی کلید گفته می‌شود. جدول زیر نمونه‌ای از این کلید است.

Alphabet	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
Key	h	i	j	a	z	d	w	v	u	p	q	r	b	c	e	g	f	m	n	o	k	l	t	s	x	y

هنگام رمزنگاری هر حرف ردیف بالا با حرف متناظر از ردیف پایین جایگزین می‌شود و در فرآیند رمزگشایی، برعکس این اتفاق می‌افتد. به عنوان مثال، عملیات رمزنگاری با استفاده از کلید بالا به صورت زیر است.



در فرآیند رمزنگاری حروف بزرگ به حروف بزرگ متناظر خود می‌روند.

یکی از روش‌های برگرداندن پیام اصلی امتحان کردن تمام نگاشت‌های ممکن است. اما این کار $26!$ حالت دارد و انجام آن به روش آزمون و خطا با کامپیوتر شخصی امکان پذیر نیست. در مسائلی به این شکل که فضای حالت بسیار بزرگی وجود دارد، استفاده از الگوریتم ژنتیک بسیار موثر است..
با کمی دقت، می‌توان دید که این الگوریتم رمزنگاری آسیب پذیر است چراکه با داشتن لغت‌نامه‌ای از حروف می‌توان میزان مفید بودن یک کلید را سنجید و این کلید را آنقدر عوض کرد که به متن اصلی رسید.
هنگامی که می‌خواهیم بررسی کنیم که رمزگشایی تا چه حد مفید بوده‌است، به دنبال کلمات آشنا و معنادار در آن می‌گردیم. هر چقدر طول یک کلمه معنادار بلندتر باشد، رمزگشایی ما مفیدتر بوده‌است. این یکی از روش‌های سنجش میزان مفیدبودن رمزگشایی است و روش‌های دیگر نیز می‌تواند مفید باشد.
در این پروژه یک پیام رمزشده به شما داده می‌شود و شما باید با استفاده از الگوریتم ژنتیک، کلید موردنظر را پیدا کنید و پیام رمزگشایی‌شده را بازگردانید.

برای این منظور، ابتدا باید مفهوم کروموزوم را در پروژه تعریف کنید. سپس، جمعیت[1] اولیه‌ای از کروموزوم‌های خود تعریف کنید. سپس، میزان مفید بودن هر کروموزوم را بسنجید. در این جا نیاز به تابع تناسب[2] دارید که بتواند به هر کروموزوم یک عدد به عنوان میزان مناسب بودن نسبت دهد.

برای محاسبه امتیاز تناسب کروموزوم، نیاز به یک لغت‌نامه[3] داریم. یکی از کارهای اساسی و اولیه در پروژه‌های هوش مصنوعی فراهم کردن مجموعه‌ی داده[4] است. لغت‌نامه در این پروژه مجموعه داده‌ی ماست. **برای درست کردن آن متنی را که در پیوست آمده‌است، باید پردازش کنید.** تضمین می‌شود که تمامی لغات پیام‌ها در متن پیوست آمده‌اند. (در پروژه‌های واقعی خودتان باید لغت‌نامه خود را ایجاد و گسترش دهید تا پوشش مناسب را ایجاد کنید.)

یکی از مراحل جمع‌آوری مجموعه داده **تمیز کردن داده‌هاست. تمیزکردن در قالب این پروژه چگونه باید انجام شود؟ روند تولید لغت‌نامه از متن را توضیح دهید. (به کاراکترهای غیر کارآمد و ایست واژه‌ها[5] فکر کنید.)**

پس از انتخاب کروموزوم‌های برتر، عملیات cross over و mutation انجام می‌شود تا جمعیت جدید به دست آید. روند ایجاد جمعیت جدید ادامه پیدا می‌کند تا با بهبود مرحله به مرحله کلید،کلید اصلی و متن خام اولیه حاصل شود.

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

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

Mutation موثرتر است یا cross over ؟ کدام باعث سریع‌تر بالا رفتن دقت می‌شود؟

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؟

راه‌حل‌های مختلف خود را بیان کنید و آن‌ها را امتحان کنید. بهترین آن‌ها را بر روی پروژه خود پیاده‌سازی کنید. (به میزان استفاده از انتخاب تصادفی فکر کنید.)

---

[1] Population

[2] Fitting Function

[3] Dictionary

[4] Dataset

[5] Stop Words

نکات پایانی:

- تمام مراحل کار خود (تعریف کروموزوم، تولید جمعیت اولیه، تابع تناسب، عملیات cross over، mutation و جلوگیری از سوگیری کروموزوم‌ها، پردازش داده‌اولیه و تولید لغت‌نامه) را در گزارش کار خود بیاورید.
- برخی از پرسش‌های مطرح شده در متن از جمله مشکلاتی هستند که باید راهکاری برای آنها ارائه دهید. این دست از مشکلات و راه‌حل پیشنهادی باید در گزارش کار تفسیر شوند. برخی دیگر سوالاتی هستند که فکر کردن به آن‌ها در روند یادگیری و کیفیت پروژه شما تاثیرگذار است. بر روی آن‌ها فکر کنید و مبنی بر پیاده‌سازی خود به آن‌ها پاسخ دهید.
- Decoder خود را در قالب یک کلاس پیاده‌سازی کنید. این کلاس باید تمام ویژگی‌های لازم برای رمزگشایی را داشته باشد تا بتوان به راحتی آن را تست کرد. سازنده این کلاس باید صرفا یک ورودی متن رمزشده را بگیرد. همچنین این کلاس، باید یک تابع decode داشته باشد که هیچ ورودی ندارد و خروجی آن متن رمزگشایی‌شده است. تست‌ها به صورت اتوماتیک اجرا خواهند شد.
  کد شما باید بتواند به صورت زیر تست شود.

```
encoded_text = open("encoded_text.txt").read()
from code import Decoder
d = Decoder(encoded_text)
decoded_text = d.decode()
print(decoded_text)
```

- تمامی نتایج باید در یک فایل فشرده با عنوان CA۲_<SID>.zip تحویل داده شود. این فایل باید شامل موارد زیر باشد:
  - یک فایل code.py که شامل تمام کدهای شماست.
  - گزارش پروژه با فرمت pdf یا HTML و شامل شرح تمامی کارهای انجام شده، نتایج به دست آمده و تحلیل‌ها و بررسی‌های خواسته شده در صورت پروژه.
  - می‌توانید از Jupyter notebook برای نگارش گزارش خود استفاده کنید. (به شدت توصیه می‌شود.)
- توجه داشته باشید که علاوه بر ارسال فایل‌های پروژه، این پروژه به صورت حضوری نیز تحویل گرفته خواهد شد.
- هیچگونه شباهتی در انجام این پروژه بین افراد مختلف پذیرفته نمی‌شود. در صورت کشف هرگونه تقلب برای همه افراد متقلب نمره ۱۰۰- در نظر گرفته می‌شود.
- در صورتی که سوالی در مورد پروژه داشتید بهتر است در فروم درس مطرح کنید تا بقیه دوستانتان نیز از آن استفاده کنند. در غیر این صورت به یکی از طراحان پروژه ایمیل بزنید.

iamirranjbar@gmai.com
behzad.shayegh.b@gmail.com

ae.۵۶۱۹۹۹@gmail.com

موفق باشید