

به نام خدا



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

دانشکده مهندسی کامپیوتر

گزارش پروژه‌ی چهارم مبانی هوش مصنوعی

استاد: دکتر روشن فکر

دانشجویان: طه‌ورا مجلسی و بهار کاویانی

(شماره دانشجویی: ۹۷۳۱۰۵۹ و ۹۷۳۱۰۵۱)

نیم‌سال دوم ۹۹-۰۰

گزارش انجام پروژه

تاثیر حذف کلمات پرتکرار و کم تکرار در دقت بدست آمده

با حذف کلمات پرتکرار احتمال هر کدام از کلمات در قسمت unary بالاتر رفت. و همچنین باعث شد که احتمال کلمه های قسمت binary هم بالاتر شود. با حذف کردن کلمات کم تکرار هم باعث شد که تعداد کل کلمات ما کم تر شود و سرعت محاسبه ی احتمال جملات بالاتر برود و همچنین احتمال کلمات متفاوت تر و شاخص تر شد.

تاثیر مقدار λ و ϵ در دقت بدست آمده

هر چقدر که مقدار لاندای ضرب شده در احتمال باینری ها بیشتر باشد، پاسخ با دقت بیشتری به واقعیت نزدیک است. اما اگر λ یا ϵ زیاد باشند، در آن صورت پاسخ خروجی ممکن است خطاهایی داشته باشد.

علت تقریباً واضح است. هر چه تاثیر احتمال های محاسبه شده بیشتر باشد، در نتیجه آن محاسبات در خروجی نقش پر رنگ تری خواهند داشت و جواب عقلانی تر می شود. اما اگر بخش انتهایی که تنها یک محاسبه ی رندوم به وجود می آورد، بیشتر باشد؛ ممکن است پاسخ اشتباه باشد.

بهترین دقت دست یافته و تحلیل تاثیر پارامترها در آن

دقت	بخشی از دیتاست آموزش داده نشده است	λ_1	λ_2	λ_3	ϵ
0.765	22 جمله از هر مجموعه	0.001	0.9	0.098	0.001
0.72	22 جمله از هر مجموعه	0.0001	0.9	0.0998	0.001
0.765	22 جمله از هر مجموعه	0.001	0.4	0.559	0.001
0.723	22 جمله از هر مجموعه	0.001	0.8	0.198	0.1

هر چقدر مقدار پارامتر λ_1 بیشتر شود، دقت نهایی کمتر می شود. تاثیر ϵ نیز همانند λ_1 می باشد. اما متغیرهای λ_2 و λ_3 هر چه بیشتر باشند، دقت نهایی بیشتر خواهد شد.

آدرس رپازیتوری پروژه : <https://github.com/Baharkaviani/NLP>