

علی فرجی ۹۶۳۴۰۲۴

برای ۴ حالت در کد اجرا شد. به خاطر تکرار بیشتر کلمات فارسی در اشعار شاعران و کم بودن کلمات نا آشنا هرچه ضریب تقریب بایگرام در مدل بک آف بیشتر باشد بهتر است.

اگر Main کد را اجرا کنید برای ۴ حالت گفت شده خروجی را محاسبه می کند.

طبق آزمایش های انجام شده هرچه اپسیلون کوچکتر و هرچه ضریب لاندا در بایگرام بیشتر باشد بهتر است.

همانطور که گفته شد در برخی موارد مثلاً یک کلمه است که کاملاً نا آشنا است و با گذاشتن اپسیلون به مقدار زیاد باعث گمراهی الگوریتم می شود و رفتار درستی نشان نمی دهد. هرچه لاندا بایگرام را زیاد کنیم به خاطر حساسیت بیشتر روی ترتیب کلمات، دقت بهتر می شود. ولی یک حالت بهینه دارد و صرفاً افزایش محض لاندا ۱ تأثیر منفی خواهد گذاشت چون باعث کم شدن انعطاف در مقابل کلمات جدید می شود و دچار مشکل می شویم. (عملاً دیگر از حالت بک آف خارج می شود)

چهار حالت اجرا شده:

lamda 1	lamda 2	lamda3	epsilon	precision
0.9	0.08	0.02	0.1	0.793241
0.9	0.08	0.02	0.9	0.747820
0.3	0.3	0.4	0.1	0.248910
0.3	0.3	0.4	0.9	0.654797

طبق نتایج موجود در کد این مقدار بهینه به دست آمد:

لاندا یک = ۰.۹
لاندا دو = ۰.۰۸
لاندا سه = ۰.۰۲
اپسیلون = ۰.۱