

(تحلیل تشخیص خطا با استفاده از CRC)

۱-۱ مقدمه

این پژوهه با هدف بررسی توانایی CRC در تشخیص خطا، چندین نسخه از CRC شامل CRC-8 ، CRC-16 و CRC-32 و CRC-64 بر روی پنج دیتاست استاندارد اعمال شد. این دیتاستها از نظر تعداد ویژگی، نوع داده و پیچیدگی تفاوت‌های قابل توجهی دارند که امکان تحلیل CRC را در شرایط مختلف فراهم می‌سازد. پس از محاسبه مقدار CRC برای هر نمونه، سه مدل مختلف خطا شامل وارونگی تصادفی بیت‌ها (Random Bit Flip)، خطای پشت‌سرهم (Burst Error) و کanal با احتمال خطا (Binary Symmetric Channel) بر روی داده‌ها اعمال شد تا میزان توانایی CRC در تشخیص اختلاف بین نسخه اصلی و نسخه خراب شده ارزیابی گردد.

نتایج نشان می‌دهد که CRC در تشخیص کوچک‌ترین تغییرات داده عملکرد بسیار قوی دارد و در اغلب سناریوهای آزمایشی نرخ تشخیص خطا نزدیک به ۱۰۰٪ بوده است. این رفتار نشان‌دهنده کارآمدی CRC در کاربردهای عددی و داده‌های با ابعاد بالا است؛ جایی که کوچک‌ترین تغییر بیتی تفاوت قابل توجهی در مقدار CRC ایجاد می‌کند. تحلیل برخوردهای CRC و توزیع خروجی نیز نشان می‌دهد که نسخه‌های ۱۶ بیت به بالا تقریباً بدون برخورد عمل کرده و برای کاربردهای حساس توصیه می‌شوند.

۲-۱ انواع CRC مورد استفاده :

CRC-8-ATM (۱)

CRC-8-Dallas (۲)

CRC-16-CCITT (۳)

CRC-16-MODBUS (۴)

CRC-32-IEEE (۵)

CRC-32C (۶)

CRC-64-ECMA (۷)

۳-۱ دیتاست مورد استفاده

در این پژوهه ۵ دیتاست مختلف مورد استفاده قرار گرفتند تا عملکرد CRC روی انواع متفاوت داده ارزیابی شود. این دیتاست‌ها از مجموعه استاندارد کتابخانه Scikit-Learn انتخاب شده‌اند و شامل داده‌های طبقه‌بندی با بعد و پیچیدگی گوناگون هستند.

الف) دیتاست Iris

توصیف:

این دیتاست کلاسیک مربوط به گل‌های زنبق است و یکی از پرکاربردترین دیتاست‌ها در یادگیری ماشین است.

مشخصات:

- تعداد نمونه‌ها ۱۵۰
- تعداد ویژگی‌ها ۴
- طول و عرض کاسبرگ و گلبرگ
- تعداد کلاس‌ها (Setosa, Versicolor, Virginica)
- نوع داده: کاملاً عددی (float)

** دلیل انتخاب :

به دلیل سادگی داده و نبود نویز، برای تست پایه‌ای توزیع CRC بسیار مناسب است.

ب) دیتاست Digits

توصیف:

این دیتاست تصاویر اعداد دستنویس (۰ تا ۹) را در قالب بردارهای عددی ارائه می‌دهد.

مشخصات:

تعداد نمونه‌ها ۱۷۹۷

تعداد ویژگی‌ها ۶۴

هر تصویر 8×8 پیکسل است

تعداد کلاس‌ها ۱۰

** دلیل انتخاب :

به دلیل ابعاد بالاتر داده، احتمال برخورد CRC کمتر شده و رفتار آن بهتر قابل مشاهده است.

ج) دیتاست Wine

توصیف:

این دیتاست شامل آنالیز شیمیایی سه نوع نوشیدنی شراب است.

مشخصات:

• تعداد نمونه‌ها ۱۷۸

• تعداد ویژگی‌ها ۱۳

مثل مقدار الكل، اسیدها، منیزیم، فنول‌ها و ...

• نوع داده float

• تعداد کلاس‌ها ۳

** دلیل انتخاب :

ویژگی‌ها پیوسته و متنوع هستند، بنابراین توزیع بایت‌ها در تبدیل به نوع float متفاوت و برای آزمون CRC مناسب است.

د) دیتاست Breast Cancer

توصیف:

این دیتاست مربوط به اندازه‌گیری ویژگی‌های سلول‌های تومور پستان است و یکی از دیتاست‌های مهم در حوزه پزشکی و طبقه‌بندی دودویی است.

مشخصات:

- تعداد نمونه‌ها ۵۶۹
- تعداد ویژگی‌ها ۳۰
- نوع داده‌های float
- کلاس‌ها (Malignant / Benign)

** دلیل انتخاب :

به علت تعداد ویژگی‌های زیاد، اندازه بایت زیادی تولید می‌شود و تست رفتار CRC در برابر خطاهای متعدد مناسب است.

ه) دیتاست Synthetic Sاختگی

این دیتاست با استفاده از تابع make classification ساخته شده است.

مشخصات:

- تعداد نمونه‌ها ۳۰۰
- تعداد ویژگی‌ها ۱۲
- نوع داده‌های float

• کلاس‌ها ۲

** دلیل انتخاب :

این دیتاست کاملاً کنترل شده است و برای بررسی رفتار CRC روی داده‌هایی با توزیع مصنوعی و تصادفی بسیار مفید است.

جدول یک) دیتاهای مورد استفاده برای تست CRC

دیتاست	تعداد ویژگی	تعداد داده	نوع داده	پیچیدگی	کاربرد در تحلیل CRC
Iris	کم	float	ساده	CRC	تست پایه‌ای توزیع
Digits	متوسط	integer	متوسط	CRC	ارزیابی برخورد CRC در داده‌های پیکسلی
Wine	متوسط	float	متوسط	CRC	تحلیل داده‌های علمی/آزمایشگاهی
Breast Cancer	زیاد	float	پیچیده	CRC	بررسی رفتار CRC در داده‌های بعد بالا
Synthetic	متوسط	float	قابل کنترل	CRC	تست خاص مدل خطأ و حساسیت

۱-۴ نتیجه گیری

در این پژوهه به بررسی عملکرد چندین الگوریتم CRC شامل CRC-8 ، CRC-16 ، CRC-32 و CRC-64 در تشخیص خطأ بر روی مجموعه‌ای متنوع از دیتاست‌ها پرداخته شد. نتایج به دست آمده نشان داد که CRC به عنوان یکی از ساده‌ترین و در عین حال مؤثرترین روش‌های تشخیص خطأ دیجیتال، توانایی بسیار بالایی در شناسایی کوچک‌ترین تغییرات در داده‌ها دارد. اعمال سه مدل خطأ شامل وارونگی تصادفی بیت‌ها، خطأ پشت‌سرهم و خطاهای کانالی بر روی داده‌ها نشان داد که خروجی CRC در برابر هرگونه تغییر حتی بسیار کوچک به شدت حساس بوده و در اغلب موارد مقدار CRC به شکلی کامل تغییر می‌کند. مقایسه نسخه‌های مختلف CRC نشان داد که هرچه تعداد بیت‌های CRC افزایش یابد، احتمال برخورد (Collision) کاهش یافته و دقت تشخیص خطأ افزایش می‌یابد. در این میان، CRC-16 عملکرد بسیار قابل قبولی داشته و CRC-32 و CRC-64 تقریباً بدون برخورد عمل کردند و در تمام دیتاست‌ها نرخ تشخیص نزدیک به ۱۰۰٪ را نشان دادند.

همچنین لازم به ذکر است جدول خروجی اجرای کد (نتایج) به شرح زیر آورده شده است :

[crc_results_full.xlsx](#)