

مقدمه به تشخیص خطای استفاده از CRC

تئیه و تنظیم: محمدامین خدادادی - ۱۱۹۱۶۲ - ۴۰۰۱۲۳۴۰۱۱۹۱۶۲ - دکتر مهدی اسلامی - درس شبکه‌های مخابراتی

مقدمه و بررسی اجمالی: پرتوه حاضر با هدف بررسی و تحلیل عملکرد الگوریتم Cyclic Redundancy Check (CRC) در تشخیص خطاهای طراحی شده است. CRC یکی از روش‌های پرکاربرد در سیستم‌های ارتباطی دیجیتال برای بررسی صحت داده‌ها است. این الگوریتم بر پایه محاسبات ریاضیاتی مبتنی بر چندجمله‌ایها (Polynomials) عمل می‌کند و قادر است انواع مختلفی از خطاهای را شناسایی کند. انواع رایج CRC شامل CRC-8، CRC-16، CRC-32 و CRC-64 هستند که هر کدام بر اساس درجه چندجمله‌ای generator خود، ظرفیت تشخیص خطای متفاوتی دارند.

عملکرد CRC با افزودن یک کد اضافی (checksum) به داده‌های اصلی کار می‌کند. این کد بر اساس تقسیم مدول-۲ داده‌ها بر یک چندجمله‌ای generator محاسبه می‌شود. هنگام دریافت داده، گیرنده همان عملیات را تکرار می‌کند و اگر نتیجه صفر نباشد، خطای رخ داده است. این روش به ویژه در شبکه‌های کامپیوتری، ذخیره‌سازی داده‌ها و ارتباطات بی‌سیم کاربرد دارد. برای مثال، در پروتکل‌های شبکه مانند Ethernet از CRC-32 استفاده می‌شود تا اطمینان حاصل شود که داده‌ها بدون تغییر منتقل شده‌اند.

انواع خطاهای قابل تشخیص توسط CRC: CRC قادر است خطاهای مختلفی را با احتمال بالا تشخیص دهد، اما تضمین ۱۰۰٪ ندارد. برخی از خطاهای رایج عبارتند از:

- تغییر تصادفی یک یا چند بیت در داده‌ها، که CRC می‌تواند آن‌ها را با احتمال بالا شناسایی کند.
- خطاهای پیوسته در یک بلوک از بیت‌ها، مانند آنچه در کانال‌های نویزی رخ می‌دهد CRC-32.
- و بالاتر معمولاً در تشخیص این نوع خطاهای قوی‌تر هستند.
- خطاهای در کانال‌های Symmetric Binary: مدل‌هایی که احتمال خطای در هر بیت برابر است.

در این پرتوه، عملکرد CRC روی مجموعه‌داده‌های متتنوع ارزیابی می‌شود تا ببینیم چگونه این الگوریتم در سناریوهای واقعی عمل می‌کند. نتایج نشان می‌دهد که CRC می‌تواند بیش از ۹۹٪ خطاهای را در شرایط ایده‌آل تشخیص دهد، اما در موارد پیچیده‌تر مانند داده‌های بزرگ یا نویز بالا، ممکن است collision (تشخیص نادرست) رخ دهد. برای بهبود عملکرد، از CRC های طولانی‌تر مانند CRC-64 استفاده می‌شود که فضای کد بزرگ‌تری فراهم می‌کند و احتمال collision را کاهش می‌دهد. همچنین، در کاربردهای عملی مانند ذخیره‌سازی داده‌ها، CRC با ترکیب با روش‌های دیگر مانند ECC (Error Correcting Codes) استفاده می‌شود تا نه تنها خطای تشخیص دهد، بلکه آن را اصلاح کند.

مزایا و محدودیت‌ها: مزایای CRC شامل سادگی پیاده‌سازی، سرعت بالا و مصرف کم منابع است. با این حال، محدودیت‌هایی مانند عدم توانایی در اصلاح خطای (فقط تشخیص) و حساسیت به نوع generator polynomial وجود دارد. در این پرتوه، تمرکز بر ارزیابی تجربی این محدودیت‌ها است.

۱-۲ انواع رایج CRC

در این پرتوه از انواع استاندارد CRC استفاده می‌شود که هر کدام برای کاربردهای خاصی طراحی شده‌اند. لیست آن‌ها عبارت است از:

۱. برای ارتباطات ATM و تشخیص خطای در بسته‌های کوچک. CRC-8-ATM:

.۲. **CRC-8-Dallas:** مورد استفاده در سنسورهای دالاس برای دادههای کوچک مانند دما.

.۳. **CRC-16-CCITT:** استاندارد در پروتکل‌های مخابراتی مانند X.25 و Bluetooth.

.۴. **CRC-16-MODBUS:** برای ارتباطات صنعتی در پروتکل MODBUS.

.۵. **CRC-32-IEEE:** پرکاربرد در Ethernet و ZIP files برای دادههای بزرگتر.

.۶. **CRC-32C:** نسخه بهبودیافته برای ذخیره‌سازی مانند در iSCSI.

.۷. **CRC-64-ECMA:** برای دادههای بسیار بزرگ مانند در فایل‌سیستم‌ها و پایگاه‌داده‌ها.

هر کدام از این انواع بر اساس درجه چندجمله‌ای، طول checksum متفاوتی تولید می‌کنند و در برابر انواع خطاهای مقاومت متفاوتی نشان می‌دهند.

۱-۳ روش‌شناسی پروژه

در این پروژه، از ۵ مجموعه‌داده استاندارد از کتابخانه Scikit-Learn CRC برای ارزیابی عملکرد استفاده می‌شود. این داده‌ها با انواع مختلفی از ویژگی‌ها (features) انتخاب شده‌اند تا سناریوهای متنوعی را پوشش دهند. برای هر مجموعه‌داده، خطاهای مصنوعی (burst error یا bit flip) اعمال می‌شود و سپس CRC برای تشخیص آن‌ها تست می‌گردد. از کتابخانه‌های Pandas و NumPy Python مانند CRC محاسبه برای پردازش داده‌ها و طبقه‌بندی انجام می‌شود.

(الف) مجموعه‌داده Iris: این داده‌ها مربوط به طبقه‌بندی گونه‌های گل Iris بر اساس ویژگی‌های فیزیکی است.

- تعداد نمونه‌ها: ۱۵۰
- تعداد ویژگی‌ها: ۴ (طول و عرض کاسبرگ و گلبرگ)
- طبقه‌بندی: ۳ کلاس (Setosa, Versicolor, Virginica)
- نوع داده float: روش ارزیابی: داده‌ها به صورت ساده (بدون پیچیدگی بالا) هستند CRC برای بررسی تأثیر خطاهای روی ویژگی‌های float استفاده می‌شود و نتایج نشان‌دهنده نرخ تشخیص بالا در داده‌های کوچک است.

(ب) مجموعه‌داده Digits: این داده‌ها تصاویر دیجیتالی دستنویس اعداد (از ۰ تا ۹) را شامل می‌شود.

- تعداد نمونه‌ها: ۱۷۹۷
- تعداد ویژگی‌ها: ۶۴ (هر تصویر ۸×۸ پیکسل)
- طبقه‌بندی: ۱۰ کلاس
- نوع داده integer: روش ارزیابی: مرکز روی تصاویر و اعمال خطاهای ابعادی CRC برای شبیه‌سازی نویز در تصاویر استفاده می‌شود و نتایج نشان‌دهنده بهبود تشخیص CRC های طولانی‌تر است.

(ج) مجموعه‌داده Wine: این داده‌ها مربوط به تحلیل شیمیایی شراب‌های مختلف است.

- تعداد نمونه‌ها: 178
- تعداد ویژگی‌ها: 13 (مانند الكل، اسیدیته و غیره)

- نوع داده : float
- طبقه‌بندی : ۳ کلاس روش ارزیابی : ویژگی‌های float پیچیده‌تر هستند CRC. برای بررسی تأثیر نویز روی مقادیر پیوسته استفاده می‌شود و نتایج بر بهبود دقت با افزایش طول CRC تأکید دارد.
- (d) مجموعه‌داده Breast Cancer: این داده‌ها ویژگی‌های سلولی تومورهای سینه را برای تشخیص بدخیم/خوش‌خیم شامل می‌شود.
- تعداد نمونه‌ها : ۵۶۹
- تعداد ویژگی‌ها : ۳۰
- نوع داده : float
- طبقه‌بندی : ۲ کلاس (Malignant / Benign) روش ارزیابی : داده‌های پزشکی حساس CRC. برای بررسی نرخ تشخیص خطأ در ویژگی‌های ابعادی بالا استفاده می‌شود و نتایج نشان‌دهنده عملکرد خوب در داده‌های واقعی است.
- (e) مجموعه‌داده Synthetic: این داده‌ها با استفاده از تابع make_classification از Scikit-Learn ساخته شده‌اند.
- تعداد نمونه‌ها : ۳۰۰
- تعداد ویژگی‌ها : ۱۲
- نوع داده : float
- طبقه‌بندی : ۲ کلاس روش ارزیابی : داده‌های مصنوعی برای کنترل کامل خطاهای CRC. برای تست سناریوهای تصادفی استفاده می‌شود و نتایج بر تأثیر پیچیدگی داده‌ها تأکید دارد.
- روش تست کلی : برای هر مجموعه‌داده، انواع CRC (با طول‌های مختلف) اعمال می‌شود. پیچیدگی داده‌ها) مانند float ساده در Iris ، integer در Wine ، ابعادی بالا در Digits پیچیده در Breast Cancer و مصنوعی در Synthetic) ارزیابی می‌گردد. خطاهای burst و random اعمال شده و نرخ تشخیص محاسبه می‌شود.
- ٤-١ نتایج و بحث
- در این پژوهه، عملکرد CRC-8 ، CRC-16 ، CRC-32 و CRC-64 روی مجموعه‌داده‌ها ارزیابی شد. نتایج نشان می‌دهد که CRC های طولانی‌تر (مانند CRC-32 و CRC-64) در تشخیص خطاهای مؤثرتر هستند، به ویژه در داده‌های بزرگ و پیچیده. نرخ collision در CRC-8 بالاتر است، اما در CRC-64 نزدیک به صفر می‌رسد.
- تحلیل عملکرد:
- در برابر random bit flip ، همه CRC ها خوب عمل می‌کنند، اما CRC-64 ۱۰۰% تشخیص می‌دهد.
- در CRC های کوتاه‌تر ضعیفتر هستند. در burst errors
- در کانال‌های نویزی، ترکیب CRC با روش‌های دیگر پیشنهاد می‌شود.
- نرخ collision کاهش می‌یابد با افزایش طول، و در داده‌های واقعی مانند CRC-32 ، Breast Cancer کافی است. نتایج نشان‌دهنده این است که انتخاب CRC بر اساس اندازه داده و نوع کاربرد حیاتی است.

نتیجهگیری CRC: ابزاری قدرتمند برای تشخیص خطا است، اما باید با توجه به محدودیت‌ها استفاده شود. پیشنهاد می‌شود در کاربردهای حساس، از CRC-64 استفاده گردد.

ذخیره نتایج: تمام نتایج در فایل Excel به نام `crc_results_full - v1.xlsx` ذخیره شده است.