

تشخیص کپچا*

رضوان کرجی، امیرعلی کابلی، محمدحسین حاجیان

دانشگاه صنعتی امیرکبیر
دانشکده ریاضی و علوم کامپیوتر

۴ تیر ۱۴۰۰

چکیده

هدف این پروژه، بررسی عملکرد الگوریتم‌های رایج یادگیری ماشین در تشخیص کپچا است. بخش اصلی این تحقیق به مقایسه‌ی مدل‌های درخت تصمیم^۱، کی-نزدیک‌ترین همسایه^۲، ماشین بردار پشتیبانی^۳ و شبکه‌های عصبی می‌پردازد.

مقدمه

ناشناس بودن خدمات وب این امکان را فراهم می‌کند تا برنامه‌های کامپیوتری خود را به جای کاربران انسانی جایزنند. سرویس‌های خودکار می‌توانند تعداد بسیار زیادی ایمیل ارسال کنند، در پایگاه‌داده‌های عمومی جستجو انجام دهند و یا نظرسنجی‌های برخط را تحت شعاع قرار دهند. در دودهه‌ی اخیر پژوهش‌های بسیاری برای تشخیص این برنامه‌های خودکار از کاربران واقعی انجام شده‌است. اولین بار کپچا در سال ۲۰۰۰ در دانشگاه کارنگی ملون^۴ توسط جان لنگفورد، نیکلاس جی هوپر و لوئیس فون آن اختراع شد. [۲] کپچا به معنای تست تورینگ عمومی کاملاً خودکار به منظور جدا کردن رایانه‌ها و انسان‌ها^۵ است. [۳] کپچا از یک تست تورینگ معکوس پیروی می‌کند که در آن برنامه‌ی کپچا نقش قاضی و شرکت‌کننده مانند یک کاربر عمل می‌کند. کپچا یک سیستم دفاعی است که به عنوان ابزاری برای جلوگیری از سوءاستفاده‌ی ربات‌های شبکه از خدمات برخط اینترنت استفاده می‌شود. کپچا به دو صورت متنی و تصویری (شکل ۱) قابل استفاده است.

Captcha Recognition*

Decision Tree^۱

K-Nearest Neighbour^۲

Support Vector Machine^۳

Carnegie Mellon University^۴

Completely Automated Public Turing Test to tell Computers and Humans Apart^۵



[۴] Google's reCAPTCHA, image ((ب))



[۱] Google's reCAPTCHA, text ((آ))

شکل ۱: آ: متنی، شامل یک کلمه ی تغییرشکل یافته و یک کلمه ی اسکن شده از یک کتاب است.
ب: تصویری، کاربر باید بگوید چه چیزی در تصویر هست، شباهت ها را پیدا کند و یا به شیء خاصی اشاره کند.

داده

برای این پروژه از مجموعه داده ی تصاویر کیچا استفاده می کنیم. این مجموعه داده از خزش کیچاهای واقعی به دست آمده است. این مجموعه شامل ۹۹۹۵ جفت کیچا می باشد که هر کدام شامل ۴ کاراکتر هستند.

مراجع

- [1] Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. recaptcha: Human-based character recognition via web security measures. *Science (New York, N.Y.)* 321 (09 2008), 1465–8.
- [2] M. Blum, L. v. A., and Langford, J. The captcha project (completely automated public turing test to tell computers and humans apart). pp. 458–461.
- [3] Saini, B. s. A review of bot protection using captcha for web security. *IOSR Journal of Computer Engineering* 8 (01 2013), 36–42.
- [4] Sivakorn, S., Polakis, I., and Keromytis, A. I am robot: (deep) learning to break semantic image captchas. pp. 388–403.