

عنوان مقاله:

تصحیح خودکار داده ها مبتنی بر وابستگی تابعی و سیستم یادگیری مرکب

محل انتشار:

فصلنامه مهندسی برق دانشگاه تبریز، دوره 48، شماره 2 (سال: 1397)

تعداد صفحات اصل مقاله: 18

نویسندگان:

مهديه عطایان - تهران - دانشگاه تربیت دبیر شهید رجایی - دانشکده مهندسی کامپیوتر

نگین دانشپور - تهران - دانشگاه تربیت دبیر شهید رجایی - دانشکده مهندسی کامپیوتر

خلاصه مقاله:

صحت داده ها یکی از مهم ترین ابعاد کیفیت داده ها به شمار می رود. با توجه به حجم بالای منابع داده ای نیاز به روش هایی خودکار وجود دارد. در این مقاله راهکاری خودکار برای تصحیح داده هایی با انواع داده ای متفاوت ارائه شده است. در این راهکار در ابتدا رکوردهایی که احتمالا حاوی ویژگی خطا است با استفاده از وابستگی تابعی شناسایی می گردد، بدین صورت که رکوردی که به ازای یک وابستگی تابعی با بیش از از رکوردها در تناقض باشد، مشکوک به خطا است. سپس به ازای هر ویژگی از منبع داده مورد بررسی، سیستم یادگیری مرکب ساخته می شود. سیستم یادگیری مرکب از سه طبقه بند بیز، درخت تصمیم و شبکه عصبی MLP تشکیل شده است و دارای استراتژی ترکیب رای اکثریت است. سیستم یادگیری مرکب به وسیله رکوردهای صحیح شناسایی شده مورد آموزش قرار داده می شود. پس از آموزش طبقه بندها، هر ویژگی غلط به عنوان کلاس هدف سیستم یادگیری مرکب قرار می گیرد و مقداری برای آن پیش بینی می گردد. روش پیشنهادی قادر است چندین خطا در یک رکورد را شناسایی نماید. آزمایش ها نشان می دهد که true negative rate الگوریتم پیشنهادی در بخش تشخیص خطا به طور متوسط 93.7% و در بخش تصحیح خطا به طور متوسط 90.6% است. هم چنین آزمایش ها نشان می دهد که میزان پارامترهای ارزیابی در الگوریتم پیشنهادی در مقایسه با دو الگوریتم مشابه مبتنی بر وابستگی تابعی بهبود داشته است.

کلمات کلیدی:

تصحیح داده، تشخیص خطا، وابستگی تابعی، سیستم یادگیری مرکب

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/890168>

