

عنوان مقاله:

بهبود مدل فلگی-سانتر در اتصال رکوردی با استفاده از مدل لگ خطی و اصلاح وزن

محل انتشار:

مجله علوم آماری، دوره 17، شماره 1 (سال: 1402)

تعداد صفحات اصل مقاله: 23

نویسندگان:

علیرضا موقفی اردستانی - دانشگاه تهران

زهرا رضائی قهرودی - دانشگاه تهران

خلاصه مقاله:

امروزه با دسترسی روزافزون به پایگاه های داده اداری و حجم بالای داده های ثبت شده در سازمان ها، روش های سنتی گردآوری و تحلیل داده ها به دلیل بار پاسخ گویی بالا کارایی لازم را ندارند. بر این اساس، گذار از روش های گردآوری سنتی به روش های مدرن گردآوری و تحلیل داده ها با رویکرد آمارهای ثبتی مبنا بیش از پیش مورد توجه تحلیلگران داده ها قرار گرفته است. در روش های ثبتی مبنا، ایجاد یک پایگاه داده یکپارچه از طریق اتصال رکوردهای پایگاه های داده دستگاه های مختلف اهمیت ویژه ای دارد. بسیاری از الگوریتم های اتصال رکوردی بر پایه مدل فلگی و سانتر توسعه یافته است. یکی از نقص های مدل فلگی-سانتر این است که به درون اطلاعات موجود در مقادیر متغیرها نفوذ نمی کند و مقادیر متغیرهای رشته ای (رایج بودن یا نادر بودن مقدار ویژگی موردنظر) در آن اهمیت ندارد. در این مقاله به معرفی روشی پرداخته می شود که بتواند با اصلاح وزن های جوسازی مدل فلگی-سانتر، این تفاوت ها را در مقادیر یک متغیر رشته ای در مدل فلگی-سانتر القا کند. از طرف دیگر، مدلی که فلگی و سانتر پیشنهاد داده اند و روشی که برای تعدیل وزن های جوسازی در اتصال فراوانی مبنای رکوردها معرفی می شود، بر اساس فرض استقلال شرطی بنا شده اند. در برخی مسائل اتصال رکوردی، در تطابق و عدم تطابق میان متغیرهای مشترک مورد استفاده در جوسازی، فرض استقلال شرطی برقرار نیست. یک راهکار مورد استفاده در چنین حالتی، استفاده از مدل لگ-خطی است که امکان وجود اثرات متقابل میان متغیرهای جوسازی در مدل را فراهم می کند. در این مقاله به دو روش تعمیم مدل فلگی-سانتر، یکی با رویکرد اصلاح وزن های جوسازی و دیگری با رویکرد مدل لگ خطی با حضور اثرات متقابل میان متغیرهای اتصال دهنده در شرایطی که فرض استقلال شرطی برقرار نباشد، پرداخته می شود. روش های معرفی شده برای اتصال رکوردی در این مقاله، روی مجموعه داده های نیروی کار مرکز آمار ایران با استفاده از نرم افزار R پیاده سازی شده اند.

کلمات کلیدی:

Fellegi-Sunter model, Frequency-based matching, Adjusting weights, Conditional independence, Log-Linear model
مدل فلگی-سانتر، جوسازی فراوانی مبنا، اصلاح وزن ها، استقلال شرطی، مدل لگ خطی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1702129>

