



بیست و یکمین کنفرانس هشتای ایران

۷ و ۶ اسفندماه ۱۳۹۳ دانشگاه اصفهان

استفاده از روش خوشه بندی برای جداسازی دیجیتال تپهای نوترون و پرتوی گاما

سجاد آقابزرگی صحاف، ناصر وثوقی، ابوالفضل حسینی

دانشگاه صنعتی شریف، دانشکده مهندسی انرژی

چکیده:

جداسازی طیف نوترون-گاما در علوم هسته ای از اهمیت بسزایی برخوردار است. امروزه با گسترش صنایع دیجیتال و ساخت مبدل های آنالوگ به دیجیتال، روش های جدیدی برای جداسازی دیجیتال طیف نوترون-گاما ارائه شده است. در جداسازی دیجیتال علاوه بر افزایش دقت و سرعت طیف نگاری نوترون، امکان حذف نوفه و همچنین شناسایی و تحلیل یا حذف تپ های دارای همپوشانی نیز امکان پذیر است. در این مقاله از روش خوشه بندی برای جداسازی طیف نوترون-گاما استفاده شده است. در این روش از دو الگوریتم فازی $(FCM)cmeans$ و $Kmedoids$ برای جداسازی استفاده شده است. این روش را می توان برای جداسازی تپ های خروجی آشکارساز های مختلف و نیز تپ های حاصل از ذرات دیگری چون آلفا و بتا نیز مورد استفاده قرار داد. دقت دو الگوریتم به شیوه RI باهم مقایسه شده اند. بر اساس نتایج به دست آمده، $Kmedoids$ دقت بهتری در جداسازی دارد.

کلید واژه: تحلیل دیجیتال شکل تپ، تشخیص ذره، خوشه بندی، $Fuzzy C-means$ ، $K-medoids$



بیست و یکمین کنفرانس هسته‌ای ایران

۶ و ۷ اسفند ماه ۱۳۹۳ دانشگاه اصفهان

۱- مقدمه

جداسازی طیف ذرات با استفاده از تحلیل شکل تپ ابزاری مهم در طیف نگاری هسته‌ای است. تا قبل از استفاده از تجهیزات دیجیتال پیشرفته، جداسازی نوترون-گاما تنها از طریق آنالیز تپ خروجی از آشکارساز به وسیله مدارهای آنالوگ امکان پذیر بود. امروزه با ساخت مبدل های دیجیتال به آنالوگ و تحلیلگرهای دیجیتال پیشرفته، روش های جداسازی دیجیتال بسیاری برای جایگزینی و یا تکمیل روش های آنالوگ ارائه شده اند. از این روش ها می توان به روش تحلیل شیب تپ در حوزه زمان [۱] و فرکانس [۲]، روش مقایسه بار [۳]، استفاده از تبدیل موجک [۴]، روش گذر از صفر [۵]، شبکه های عصبی [۶] و استفاده از منطق فازی [۷] اشاره نمود. ویژگی مشترک بین روش های فوق، وابستگی آن ها به شکل پالس است. در واقع روش های فوق یک روش مبتنی بر ناظر^۱ هستند. همچنین این روش ها به معیاری مثل شیب تپ، مساحت زیر تپ یا محل عبور از صفر برای جداسازی شکل تپ ها نیاز دارند. در روش شبکه های عصبی نیز می بایست ابتدا تعدادی داده ی برچسب خورده به الگوریتم معرفی شوند و سپس الگوریتم اقدام به جداسازی خواهد کرد [۶]. روش دیگری که برای جداسازی تپ نوترون-گاما پیشنهاد شده است، استفاده از خوشه بندی با بکارگیری الگوریتم Fuzzy C-means است [۸]. مزیت این روش نسبت به سایر روش های جداسازی، بدون ناظر بودن این روش می باشد. این روش وابسته به نوع آشکار ساز نمی باشد. همچنین برای جداسازی سایر ذرات مثل آلفا و بتا نیز قابل استفاده است. ساوران برای اولین بار از FCM برای جداسازی نوترون-گاما استفاده کرده است. ساوران برای پیش پردازش داده ها از فیلتر FIR استفاده نموده و برای محاسبه شباهت داده ها از فاصله اقلیدسی آن استفاده نموده است. فیلتر FIR به دلیل ماهیت هموارکنندگی خود موجب حذف مقداری از اطلاعات موجود در قله تپ ها می شود. چنانچه بجای FCM از Kmedoids استفاده کنیم می توان جابجایی های زمانی را نیز در محاسبه شباهت در نظر گرفت که موجب افزایش دقت جداسازی می شود. الگوریتم Kmedoids برای اولین بار در این مقاله برای جداسازی تپ نوترون-گاما استفاده شده است.

۲- خوشه بندی

خوشه بندی به طور کلی به فرآیندی گفته می شود که داده های بدون برچسب موجود در یک بانک اطلاعاتی را گرفته و آن ها را در دسته هایی قرار می دهد که داده های موجود در یک دسته به لحاظ کمی یا کیفی بیشترین شباهت را نسبت به سایر دسته ها داشته باشند. به دسته هایی که دارای چنین ویژگی ای باشند خوشه می گویند. دو

^۱Supervised method



بیست و یکمین کنفرانس هسته‌ای ایران

۷ و ۸ اسفند ماه ۱۳۹۳ دانشگاه اصفهان

الگوریتم متداول در خوشه بندی FCM و Kmedoids هستند که در این مقاله از آن ها برای جداسازی تپ استفاده شده‌اند.

الگوریتم Fuzzy C-means یا به اختصار FCM یکی از پر کاربردترین الگوریتم های خوشه بندی است که اولین بار توسط بزدک [۹] پیشنهاد شده است. الگوریتم FCM به این ترتیب کار می کند که یک مجموعه از n بردار را به C گروه فازی بخش بندی می کند و یک مرکز خوشه بندی برای هر گروه به گونه ای مشخص می شود که تابع شباهت آن با سایر اعضای آن گروه کمینه شود. ساوران و همکاران از الگوریتم FCM برای جداسازی تپ نوترون-گاما استفاده کرده اند [۸].

روش دیگر خوشه بندی داده ها، حل مساله K-medoids است. مساله K-medoids عبارت است از یافتن تعداد K میانه که فاصله داده ها از آن ها کمترین مقدار باشد. یکی از الگوریتم هایی که جواب مناسبی برای این مساله داده است الگوریتم PAM^۱ است [۱۰]. مزیت این روش نسبت به FCM امکان استفاده از روش DTW برای مقایسه داده ها است. روش های محاسبه شباهت تپ ها در بخش ۵ توضیح داده شده است.

۳- تولید داده ها

برای تعیین صحت و دقت الگوریتم بکار رفته در جداسازی تپ نوترون-گاما می بایست ابتدا الگوریتم با تعدادی تپ شناخته شده آزموده شود. برای اینکار معمولا از چشمه های گاما مثل ^{60}Co تپ مربوط به گاما و از روش زمان پرواز (TOF) تپ نوترون استخراج می شود [۹]. روش دیگر استفاده از شکل تپ شبیه سازی شده می باشد [۶]. مارونه و همکاران [۱۱] شکل تپ خروجی آشکارساز سوسوزن را با استفاده از رابطه ای شامل سه تابع نمایی و شش متغیر تعریف کردند؛ که شش متغیر مجهول از طریق برازش منحنی بدست می آید. در معادله (۱) این رابطه آمده است.

$$y(t) = A \left[\exp\left(-\frac{t-t_0}{\theta}\right) - \exp\left(-\frac{t-t_0}{\lambda_s}\right) + B \exp\left(-\frac{t-t_0}{\lambda_l}\right) \right] \quad (1)$$

در این رابطه A و B به ترتیب دامنه ی قسمت سریع و کند تپ در $t = 0$ ، λ_s ثابت زمان واپاشی جزء سریع، λ_l ثابت زمان واپاشی جزء کند، θ سومین ثابت واپاشی و t_0 مرجع زمانی برای شروع تپ می باشند. با استفاده از این رابطه می توان پالس های نمونه نوترون و گاما را برای آزمودن الگوریتم ها تولید نمود. تفاوت دو تپ نوترون و گاما در قسمت پایین رونده تپ ها است و از این قسمت برای یافتن معیاری جهت جداسازی استفاده می شود. لذا می توان با ضرب یک تابع پله به تپ معادله (۱)، قسمت پایین رونده و مشابه تپ ها را حذف کرد. پر واضح است که

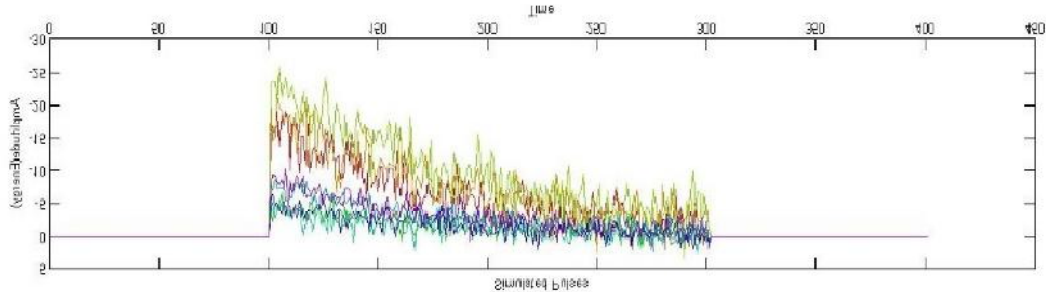
^۱Partitioning Around Medoids



بیست و یکمین کنفرانس هسته ای ایران

۶ و ۷ اسفند ماه ۱۳۹۳، دانشگاه اصفهان

این عمل هیچ تاثیری در امر جداسازی نخواهد داشت. با افزودن یک نوفه تپ شبیه سازی شده خروجی آشکارساز بدست می آید. شکل انمونه تپ های شبیه سازی شده را بعد از افزودن نوفه نشان می دهد.



شکل ۱- شکل تپ شبیه سازی شده نوترون و گاما بعد از افزودن نوفه سفید (۱۰dB)

۴- پیش پردازش

تپ های گرفته شده از آشکارساز (یا معادل آن تپ های شبیه سازی شده) را می بایست برای تحلیل توسط الگوریتم های خوشه بندی آماده نماییم. فرایند آماده سازی را برای داده کاوی پیش پردازش می گویند. مراحل پیش پردازش عبارتند از:

الف) کاهش طول پالس: حذف قسمت بالا رونده پالس و کاهش طول پالس.

ب) حذف نوفه: حذف نوفه به دو طریق انجام گرفته است و نتایج باهم مقایسه شده اند. اول با استفاده از فیلتر پاسخ ضربه محدود یا FIR [۹] و روش دوم با استفاده از تبدیل موجک نوفه های تپ ها حذف شده اند. فیلتر FIR به دلیل ماهیت هموار کنندگی یعنی حذف فرکانس های بالا، در قله تپ عملکرد مناسبی ندارند و مقدار آن را تغییر می دهند. این در حالی است که، در طیف نگاری مقدار دامنه قله تپ حائز اهمیت بوده و متناسب با انرژی ذره فرودی در آشکارساز است. ولی فیلتر های بر مبنای تبدیل موجک در قله های تیز عمل کرد بهتری دارند.

پ) بهنجارش: برای عدم وابستگی خوشه بندی تپ ها به ارتفاع قله تپ، همچنین چون در الگوریتم، فاصله اقلیدسی تپ ها محاسبه می شود، می بایست تپ های ورودی را بهنجار نماییم. به دو روش می توان این کار را انجام داد. ساده ترین روش آن بهنجار کردن تپ با تقسیم دامنه تپ بر ارتفاع دامنه قله تپ می باشد. روش دیگر بهنجارش زد یا Z-score است [۱۲]. فرض کنید $F_i = \{f_1, \dots, f_t, \dots, f_T\}$ یک تپ دیجیتال ورودی به طول T باشد. آنگاه بهنجارش زد به شکل زیر تعریف می شود.

$$Z_{Normalization}(F_i, M_i, sd) = \frac{f_t - M_i}{sd} (2)$$

که در آن M_i میانگین و sd انحراف معیار استاندارد تپ است. تپ های بهنجار شده دارای میانگین صفر و انحراف معیار یک خواهند بود.



بیست و یکمین کنفرانس هسته‌ای ایران

۷ و ۶ اسفند ماه ۱۳۹۳ دانشگاه اصفهان

ت) از مزایای استفاده از الگوریتم های داده کاوی امکان تشخیص و در صورت لزوم حذف داده های دارای همپوشانی^۳ است. برای تشخیص داده های پرت ابتدا میانگین کل داده ها گرفته می شود، سپس نمودار فراوانی فاصله داده های از میانگین کل داده ها مشخص می شود. داده هایی که بیشترین فاصله را از میانگین داده ها دارند به عنوان داده پرت شناخته می شوند.

در مراجع روش های دیگری مانند روش سلسله مراتبی نیز برای حذف داده های پرت ارائه شده اند ولی در این مساله به دلیل سادگی شکل تپ ها، روش نمودار فراوانی نتایج مطلوب جداسازی را فراهم می کند.

۵- محاسبه شباهت

محاسبه شباهت سری های زمانی در خوشه بندی، وابسته به محاسبه فاصله آن ها است. روش های ED^[۱۳] به علت بازدهی آن و DTW^[۱۴]، برای تاثیرگذاری خوب در محاسبه شباهت^۴ از معروف ترین و پرکاربرد ترین روش های محاسبه شباهت هستند. ED در واقع یک تطبیق یک-به-یک است که در اکثر مراجع (حدود ۸۰٪) از این روش استفاده شده است^[۱۵].

در مقابل روش محاسبه فاصله اقلیدسی یا ED که یک روش تطبیق یک-به-یک است، روش DTW^۵ یک روش یک-به-چند است. در واقع DTW جابجایی ها و پیچیدگی های یک تپ را نیز در نظر می گیرد و برای هر نقطه از سیگنال معادل نسبی آن را در سیگنال دیگر پیدا می کند و آن فاصله را محاسبه می کند^[۱۶].

۶- صحت سنجی

برای صحت سنجی الگوریتم های بکار رفته از شاخص RI استفاده شده است. شاخص RI^۱ یک معیار اندازه گیری متداول برای خوشه بندی سری های زمانی است^[۱۷]. RI توافق بین دو دسته را اندازه گیری می کند. به این معنا که چقدر نتایج خوشه بندی به Ground Truth نزدیک است. توافق بین دو دسته C و G را می توان با رابطه (۳) تخمین زد.

$$RI(C, G) = \sqrt{\frac{|TP|+|TN|}{|TP|+|TN|+|FP|+|FN|}} \quad (۳)$$

۷- مقایسه نتایج

^۱Pile up

^۴Similarity measurement

^۵Dynamic Time Warping

^۶Rand Index



بیست و یکمین کنفرانس هشتای ایران

۶ و ۷ اسفند ماه ۱۳۹۳ دانشگاه اصفهان

در جدول ۱ نتایج مربوط به شاخص RI برای دو روش اقلیدسی و DTW در الگوریتم Kmedoids و همچنین الگوریتم FCM آمده است. نوفه تپ های ورودی الگوریتم ها به دو روش فیلتر FIR و نیز استفاده از تبدیل موجک حذف شده اند که نتایج مربوط به هر شیوه در جدول ۱ نوشته شده است.

جدول ۱-مقایسه روش های FCM، Kmedoids-ED و Kmedoids-DTW (حذف نوفه با فیلتر FIR)

شاخص RI		الگوریتم خوشه بندی
حذف نوفه با تبدیل موجک	حذف نوفه با فیلتر FIR	
۹۵	۸۶	FCM
۵۱	۸۳	Kmedoids-ED
۵۶	۸۹	Kmedoids-DTW

همانطور که در جدول فوق ملاحظه می شود، محاسبه شباهت DTW به طور قابل توجهی نتایج را بهبود بخشیده است.

نتایج جدول فوق نشان می دهد که حذف نوفه به کمک تبدیل موجک می تواند تاثیر قابل توجهی در بهبود نتایج FCM داشته باشد. ولی در مورد الگوریتم های Kmedoids موجب بدتر شدن نتایج شده است.

۸- بحث و نتیجه گیری

روش FCM دارای سرعت بسیار خوب و دقت مناسبی برای جداسازی است [۹]. بنابراین با انتخاب روش مناسب پیش پردازش داده ها و همچنین گزینش مناسب صحت سنجی تپ های ورودی می توان از آن برای جداسازی تپ نوترون- گاما استفاده نمود. الگوریتم FCM به داده های پرت حساس و عدم حذف مناسب آن ها روی نتایج خوشه بندی تاثیر می گذارد. الگوریتم Kmedoids-ED دارای سرعت کمتری است ولی چنانچه پیش پردازش مناسبی روی داده های ورودی انجام بگیرد، نتایج آن در مقایسه با نتایج FCM قابل قبول است. اگر بجای محاسبه فاصله اقلیدسی در Kmedoids از DTW استفاده شود، دقت خوشه بندی به طور قابل ملاحظه ایی افزایش می یابد.

۹- مراجع

۱. Aspinall, M., et al., *Verification of the digital discrimination of neutrons and γ rays using pulse gradient analysis by digital measurement of time of flight*. Nuclear Instruments and Methods in Physics



بیست و یکمین کنفرانس هسته‌ای ایران

۶ و ۷ اسفند ماه ۱۳۹۳ دانشگاه اصفهان

- Detectors and Associated Equipment, 2007. **583**(2): ,Research Section A: Accelerators, Spectrometers
.p. 432-438
- Liu, G., et al., *A Digital Method for the Discrimination of Neutrons and Rays With Organic Scintillation Detectors Using Frequency Gradient Analysis*. Nuclear Science, IEEE Transactions on, .۲
.p. 1682-1691 : (۳) ۰۷ . ۰۱ ۰۲
- Kaschuck, Y. and B. Esposito, *Neutron/ γ -ray digital pulse shape discrimination with organic scintillators*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2005. **551**(2): p. 420-428 .۳
- Yousefi, S., L. Lucchese, and M.D. Aspinall, *Digital discrimination of neutrons and gamma-rays in liquid scintillators using wavelets*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2009. **598**(2): p. 551-555 .۴
- Nakhostin, M. and P. Walker, *Application of digital zero-crossing technique for neutron-gamma discrimination in liquid organic scintillation detectors*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2010. **621**(1):
.p. 498-501 .۵
- Liu, G., et al., *An investigation of the digital discrimination of neutrons and γ rays with organic scintillation detectors using an artificial neural network*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2009. **607**(3):
.p. 620-628 .۶
- Kumar, A., et al. *Digital Pulse Shape Discrimination of Detector Data using Fuzzy Clustering*. in .۷
Proceedings of the DAE Symp. on Nucl. Phys. 2011
- Bezadek, J., *fuzzy mathematics in pattern classification*. 1973, Cornell University: Applied .۸
Mathematics Center
- Savran, D., et al., *Pulse shape classification in liquid scintillators using the fuzzy c-means algorithm*. .۹
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2010. **624**(3): p. 675-683
- .Madhulatha, T.S., *An overview on clustering methods*. arXiv preprint arXiv:1205.1117, 2012 .۱۰
- Marrone, S., et al., *Pulse shape analysis of liquid scintillators for neutron studies*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2002. **490**(1): p. 299-307 .۱۱
- Han, J. and M. Kamber, *Pei. Data Mining Concepts and Techniques*. 2011, The Morgan Kaufmann .۱۲
Series in Data Management Systems, Morgan Kaufmann Publishers
- C. Faloutsos, M.R., Y. Manolopoulos, *Fast subsequence matching in time-series databases*. ACM .۱۳
SIGMOD Record, 1994. **23**: p. 419-429
- H. Sakoe, S.C., *Dynamic programming algorithm optimization for spoken word recognition*. IEEE .۱۴
Transactions on Acoustics, Speech and Signal Processing, 1978. **26**: p. 7
- F.K.P. Chan, A.W.C.F., *Haar wavelets for efficient similarity search of time-series: with and without time warping*. Ieee Transactions on Knowledge and Data Engineering, 2003. **15**: p. 10 .۱۵
- E. Keogh, C.R., *Exact indexing of dynamic time warping*. Knowledge and Information Systems, 2004. .۱۶
.7: p. 28
- Chiş, M., S. Banerjee, and A.E. Hassanien, *Clustering time series data: an evolutionary approach*, in .۱۷
Foundations of Computational, Intelligence Volume 6. 2009, Springer. p. 193-207