

سامانه‌ی تشخیص خودکار وزن عروضی اشعار فارسی

محمدجواد امیری*، مریم غفوریان**

لازم باشد به نوعیت قاعده‌های از پیش تعیین شده (وزن کلاسیک یا نو) اتکا کنیم [۷]. وجود وزن در شعر باعث می‌شود خواننده‌ی شعر، در آن احساس تمایزی بیشتری نسبت به کلام عادی بکند و به موازات این احساس، تسلیم و بی‌خویشتنی بیشتری در برابر شعر نشان دهد [۸]. مهمترین عاملی که در وزن اشعار مؤثر است تساوی هجاهای هر مصراع است. مبنای وزن در همه‌ی زبان‌ها یکسان نیست و در هر زبان شکلی خاص دارد. در بعضی زبان‌ها تعداد، در بعضی اختلاف شدت تلفظ، در بعضی زیر یا بم بودن و در بعضی کمیّت هجاها ملاک است. وزن شعر در فارسی و عربی بر مبنای کمیّت است و در انگلیسی بر مبنای شدت تلفظ [۱].

۱-۲ نقش و جایگاه وزن عروضی در ادبیات فارسی

وزن عروضی از دیرباز تا کنون همواره مورد توجه شاعران بوده است و تا ابتدای قرن جاری نمی‌توان اثری از شعر بدون وزن در آثار ادبی پیدا کرد و در میان شاعران بزرگ معاصر نیز کمتر شاعری را می‌توان یافت که سرودن شعر بدون وزن را مبنای کار خویش قرار داده باشد [۱].

وزن عروضی یکی از مشکل‌ترین و البته جذاب‌ترین رشته‌های ادبیات فارسی است که امروزه کاربرد وسیعی پیدا کرده‌است. زبان‌شناسان با استفاده از وزن عروضی به راحتی در مورد اشعار شعرا، مقتضیات زمانی شاعر، زندگی شاعر و ... نظر می‌دهند.

باید توجه داشت که عروض برای شناختن اجزاء شعر و تشخیص وزن صحیح از وزن شکسته و ناصحیح است و یادگیری آن هرگز برای به دست آوردن توانایی لازم برای سرودن اشعار نیست [۹].

۲- آشنایی با حرف و هجا

پیش از آشنایی با علم عروض نخست لازم است با حرف و هجا که مبنای زبان فارسی است آشنا شویم.

۱-۲ حرف

شاعر با واژه به سرودن شعر می‌پردازد و واژه خود از واحدهای کوچک‌تری به نام حرف تشکیل شده است. باید توجه داشت که در وزن شعر، صورت ملفوظ حروف مد نظر است نه شکل نوشته شده‌ی آنها. حرف بر دو گونه است: مصوت‌ها و صامت‌ها.

۱-۱-۲ مصوت‌ها

مصوت حرکتی است که به صامت داده می‌شود [۱]. زبان فارسی دارای سه حرف صدا دار کوتاه و سه حرف صدا دار بلند است. مصوت‌های کوتاه یا "حرکات" عبارتند از: $\bar{\text{ـ}}$ $\underline{\text{ـ}}$ $\overset{\sim}{\text{ـ}}$

چکیده - هدف این پژوهش طراحی سامانه‌ای جهت تشخیص وزن اشعار فارسی است. در این مقاله ابتدا با حرف و هجا آشنا می‌شویم و سپس به معرفی اوزان پرکاربرد شعر فارسی می‌پردازیم. بعد از آن نوبت به تبیین مراحل مختلف پردازش شعر می‌گردد. این مراحل شامل تصحیح، حرکت‌گذاری، تبدیل به رشته‌ی صامت-مصوت، اعمال اختیارات شاعری بر رشته صامت-مصوت‌ها، تبدیل به رشته‌ی «U» و «-»، اعمال اختیارات وزنی شاعری بر رشته‌ی تولید شده و انطباق با یکی از ۳۵ ساختار وزنی است. در این پژوهش از یک مجموعه کلمه‌ی استفاده شده است که هر کلمه‌ی حرکت‌گذاری شده و در صورت نیاز اختیارات شاعری مربوط به آن کلمه در کنار آن قرار گرفته شده است تا پس از ورود شعر کلمات آن حرکت‌گذاری شوند پس از طی مراحل و تبدیل به رشته‌ی «U» و «-» اصلاح شده با استفاده از الگوریتم پیشنهادی شعر ورودی را به یکی از ۳۵ وزن مشهور نسبت می‌دهیم. الگوریتم پیشنهادی یک الگوریتم ابداعی و شامل ۷ مرحله است.

کلید واژه- پایگاه داده‌ی کلمات، خط عروضی، رشته‌ی هجاهای کوتاه و بلند، وزن عروضی اشعار

۱- مقدمه

ما در طول تاریخ ادبیات کهن فارسی، بیش از بیست هزار شاعر با نام و نشان داشته‌ایم [۱]. همین گستردگی ادبیات فارسی زمینه‌ای مناسب برای پژوهش‌هایی تأثیرگذار را فراهم آورده است. اما عدم تعامل و همکاری مناسب میان زبان‌شناسان و ادیبان از سویی و اساتید علوم رایانه از سوی دیگر باعث گردیده است تا کمتر پژوهش مؤثری در این زمینه انجام شود. در این میان نکته‌ی دیگری نیز باعث عدم تمایل دانش‌پژوهان رایانه‌ای به پردازش شعر و ادبیات فارسی گردیده است و آن سختی این کار است. غنا و وسعت ادبیات فارسی به حدی زیاد است که پردازش آن در قالب توابع و الگوریتم‌های رایانه‌ای کاری سخت و زمان‌بر است و نیاز به دقت و حوصله‌ی کافی دارد.

علی‌رغم همه‌ی این مشکلات محققان در بسیاری از زبان‌ها موفق به سرآییدن شعر شده‌اند [۲-۵]. و امید آن است که با تعامل اساتید ادبیات و رایانه بتوان در ادبیات فارسی نیز به این مرحله رسید.

۱-۱-۱ وزن عروضی چیست

وزن شعر، حاصل نظم و تناسبی است که در صورت‌های ملفوظ ایجاد شده باشد [۶]. حقیقت وزن، همان چیزی است که ما از حضور آهنگین واژگان در تناسب با معنای آن‌ها در بافت شعر دریافت می‌کنیم؛ بی آن که

*محمدجواد امیری، ۸۶۵۲۱۰۵۳، گرایش مهندسی نرم‌افزار، دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت (email: mohamad.amiri61@gmail.com)
**مریم غفوریان، ۸۶۵۲۱۳۴۷، گرایش مهندسی نرم‌افزار، دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت (email: maryam.ghafourian@gmail.com)

به هجای بعد منتقل گردد، آن گاه در این هجای جدید، پس از حرف صدادار بلند قرار نمی‌گیرد و از این رو به شمار می‌آید (مثل جانا).
 ۳. "ا" در خط برابر است با همزه‌ی صدادار و حرف صدادار بلند "ا"، از این رو سه حرف به شمار می‌آید.

۳- اوزان شعر فارسی

شاعران فارسی بیش از حدود سیصد وزن را تجربه نکرده‌اند. از این میان، شاید کمتر از پنجاه وزن استفاده‌ی عام یافته و بقیه فقط در حد تفنن و طبع آزمایی به کار رفته است و بس [۱]. سامانه‌ی طراحی شده قادر به تشخیص ۳۵ وزن است که از [۱] استخراج شده است. لازم به ذکر است که شمیسا [۶] تعداد اوزان را پرکاربرد را ۳۱ عدد، الول [۱۱] تعداد آن‌ها را ۲۹ و فرزاد [۱۲] تعداد آن‌ها را ۳۳ عدد دانسته است. در زیر به معرفی ۳۵ وزن پرکاربرد شعر فارسی می‌پردازیم:

۱. مستفعّلن مستفعّلن مستفعّلن مستفعّلن
۲. فاعلاتن فاعلاتن فاعلاتن فاعلاتن
۳. مفاعیلن مفاعیلن مفاعیلن مفاعیلن
۴. مفاعیل مفاعیل مفاعیل مفاعیل
۵. فاعلاتن فاعلاتن فاعلاتن فاعلاتن
۶. مفاعّلن مفاعّلن مفاعّلن مفاعّلن
۷. مفاعلاتن مفاعلاتن مفاعلاتن مفاعلاتن
۸. متفاعّلن متفاعّلن متفاعّلن متفاعّلن
۹. فاعولن فاعولن فاعولن فاعولن
۱۰. مستفعّلن فاعلاتن مستفعّلن فاعلاتن
۱۱. مستفعّلن فع مستفعّلن فع
۱۲. مفتعلن فاعلات مفتعلن فاعلات
۱۳. مفتعلن مفاعّلن مفتعلن مفاعّلن
۱۴. مفاعّلن فاعلاتن مفاعّلن فاعلاتن
۱۵. فعلات فاعلاتن فعلات فاعلاتن
۱۶. مفعولن فاعلاتن مفعولن فاعلاتن
۱۷. مفعول مفاعیلن مفعول مفاعیلن
۱۸. فاعلن مفاعّلن فاعلن مفاعّلن
۱۹. فاعلن مفاعیلن فاعلن مفاعیلن
۲۰. فاعلن فاعلن فاعلن فاعلن
۲۱. فاعولن فاعولن فاعولن فاعولن
۲۲. مستفعّلن مستفعّلن مستفعّلن مستفعّلن
۲۳. فاعلاتن فاعلاتن فاعلاتن فاعلاتن
۲۴. فاعلاتن فاعلاتن فاعلاتن فاعلاتن
۲۵. مفاعیلن مفاعیلن مفاعیلن مفاعیلن
۲۶. مفعول مفاعیل مفعول مفاعیل
۲۷. مفعول مفاعّلن مفعول مفاعّلن
۲۸. مفعول مفاعّلن مفعول مفاعّلن

مصوت‌های صدادار بلند عبارتند از: "و"، "ا"، "ی" وقتی به صورت "او"، "آ" و "ای" عمل می‌کنند.
 ذکر نکات زیر در این باره ضروری به نظر می‌رسد:
 ۱. هر حرف صدادار بلند تقریباً دو برابر حرف صدادار کوتاه است. از این رو حرف صدادار بلند در وزن شعر فارسی دو حرف به شمار می‌آید.
 ۲. "و"، "ا"، "ی" زمانی حرف صدادار بلند و دو حرف به شمار می‌آید که دومین حرف هجا باشد.

۲-۱-۲-۲ صامت‌ها

صامت حرف بدون حرکت است [۱]. زبان فارسی دارای ۲۳ حرف بی‌صدا است:
 ء (=ع)، ب، پ، ت (=ط)، ج، چ، خ، د، ر، ز (=ذ، ظ، ض)، ژ، غ، س (=ث، ص)، ش، غ (=ق)، ف، ک، گ، ل، م، ن، و، ه (=ح)، ی.

۲-۲-۲ هجا

هجا یک واحد گفتار است که با هر ضربه‌ی هوای ریه به بیرون رانده می‌شود [۱۰]. در زبان فارسی هر هجا دارای یک حرف صدادار است که دومین حرف هجاست. از این رو در هر گفته به تعداد حرف‌های صدادار هجا وجود دارد.

۲-۲-۱ انواع هجا

هجاهای موجود در شعر فارسی را می‌توان به ترتیب زیر دسته‌بندی کرد [۱]:

۲-۲-۱-۱ هجای کوتاه

صامت + مصوت کوتاه (مثل بَ، شِبْ مْ)
 که با U نمایش داده می‌شود.

۲-۲-۱-۲ هجای بلند

صامت + مصوت کوتاه + صامت (مثل تَر، دِل، گُل)
 صامت + مصوت بلند (مثل پَا، بی، مو)
 که با - نمایش داده می‌شود.

۲-۲-۳ هجای کشیده

صامت + مصوت کوتاه + صامت + صامت (مثل سَنگ، عَشق، قفل)
 صامت + مصوت بلند + صامت (مثل رَاز، شَیر، هوش)
 صامت + مصوت باند + صامت + صامت (مثل کَارد، زَیست، گوشت)
 که با -U نمایش داده می‌شود.

۲-۲-۲ نکات ضروری

۱. در وزن شعر فارسی هر حرف صدادار بلند دو حرفی به شمار می‌آید.
 ۲. در وزن شعر، حرف "ن" پس از یک حرف صدادار بلند در یک هجا (یعنی نون ساکن)، به شمار نمی‌آید (مثل جان). ولی اگر حرف "ن" هجا

به تعداد حرف‌های صدادار هجا وجود دارد. هجای کشیده را نیز به یک هجای بلند (سه حرف نخست) و یک هجای کوتاه (یک یا دو حرف بعد) تقسیم می‌کنیم.

علامت هجای دو حرفی (کوتاه) "U" است.

علامت هجای سه حرفی (بلند) " – " است.

علامت هجاهای چهار یا پنج حرفی (کشیده) "U – " است.

۴-۳ تقطیع ارکان

غالب وزن‌های شعر فارسی از تکرار زنجیره‌وار یک یا دو خوشه‌ی چند هجایی ساخته شده‌اند. این نظم، به ما کمک می‌کند که هر وزن را با مجموعه هجاهای سازنده‌ی آن مشخص کنیم. به این مجموعه‌ها، «رکن» می‌گویند [۱]. منظور از تقطیع ارکان مشخص کردن مرز دسته‌ی هجاهای تکرار شونده است.

ارکان عروضی را بر اساس تعداد هجاهای آن دسته بندی می‌کنند.

جدول (۱) جدول ارکان [۶]

نوع	نوع هجا	رکن عروضی
یک هجایی	–	فَع
دو هجایی	U _	فَعْل فَعْلُن
سه هجایی	– _ – U _	فَعْلُن فَاعْلُن فَعولن مفعولن مفعولُ
چهار هجایی	– U _ _ – U _ U U U _ _ U _ U U U _ _ _ U _ _ U U _ U _ – _ U _ – _ U U – U U _	فاعلاتن فاعلات فعلاتن فعلات مفاعیلن مفاعیلُ مفاعیلن مستفعلن مستفعلُ مفتعلن
پنج هجایی	– _ U _ _ U U _ U _	مستفعلاتن مفاعیلن

۴-۴ اعمال اختیارات شاعری

اختیارات شاعری با این معناست که گاهی به شاعران اجازه داده شده است تغییراتی جزئی در حد کوتاه یا بلند کردن یک هجا و نظایر آن در

۲۹. مفعول فاعلات مفاعیل

۳۰. مفعول فاعلات مفاعیلن

۳۱. مفعول فاعلات مفاعیلن فع

۳۲. فاعلاتن فاعلاتن فاعلاتن فاعلاتن

۳۳. مفتعلن فاعلات مفتعلن فع

۳۴. مفعول فاعلات مفاعیل فاعلات

۳۵. مفاعیلن فاعلاتن مفاعیلن فاعلات

۴-۴ تعیین وزن عروضی اشعار

تعیین وزن عروضی اشعار از جمله مباحثی است که اگر چه یادگیری آن آسان است اما به کارگیری آن دشوار است و گاه بر سر آن اختلاف نظرهای اساسی بین صاحبان فن به وجود می‌آید. برای تعیین وزن عروضی اشعار فارسی روش‌های متفاوتی وجود دارد که در این جا ما از قواعد زیر که در [۶] بیان شده است استفاده خواهیم کرد.

۴-۱ درست خواندن و درست نوشتن شعر (خط عروضی)

وزن مسأله‌ای آوایی است و مبنای آن تلفظ واژه‌هاست، نه شکل نوشتاری آن‌ها [۱]. در تعیین وزن شعر باید خط را تا حد امکان به صورت ملفوظ شعر نزدیک کرد. این خط را "خط عروضی" می‌نامند.

در نوشتن خط عروضی نکات زیر را باید رعایت نمود:

۱. در خط عروضی باید حرکت حروف صدادار کوتاه را گذاشت.

۲. اگر در فصیح خوانی شعر، همزه‌ی آغاز هجا (وقتی پیش از آن حرف بی صدایی باشد) تلفظ نشود، در خط عروضی نیز همزه را باید حذف کرد.

۳. حروف تشدیددار اگر واقعا مشدد تلفظ شوند معادل دو حرف یکسان که اولی ساکن و دومی متحرک است عمل می‌کنند (مثل شدت = شد + دت).

۴. واو آخر کلمه گاهی به صورت مصوت بلند تلفظ می‌شود و گاهی به صورت مصوت کوتاه (مصوت بلند مثل او، مصوت کوتاه مثل تو).

۵. حروفی که در خط فارسی هست اما تلفظ نمی‌شود، در خط عروضی حذف می‌شود (مثل خواهر = خاهر).

پس از تعیین خط عروضی لازم است رشته صامت- مصوت‌ها تولید شود. برای این کار کافی است به جای صامت‌ها از C، به جای مصوت‌های کوتاه از V و به جای مصوت‌های بلند از VV استفاده کنیم.

۴-۲ تقطیع هجایی

تقطیع هجایی به معنی تجزیه‌ی شعر به هجاها است.

منظور از تقطیع هجایی مشخص کردن هجاهای شعر اعم از کوتاه، بلند و کشیده، سپس جدا کردن هجاها و نوشتن شعر به خط عروضی است.

وزن شعرشان اعمال کنند [۱]. برای تعیین وزن یک شعر، تقطیع یک مصراع از آن کافی است اما چون در شعرهای فارسی اغلب از اختیاریهای شاعری بهره گرفته می‌شود، با مقایسه‌ی دو مصراع یعنی از روی اختلاف هجاها، اختیاریهای شاعری را بهتر درمی‌یابیم.

اختیارات شاعری بر دوگونه زبانی و وزنی است و شامل موارد زیر می‌شود:

(الف) اگر قیل از الف حمزه‌ی آغاز هجا، حرف صامتی قرار بگیرد می‌توان الف حمزه را حذف کرد (مثال: هجای هفتم مصراع اول در شعر «آمد نوروز هم از بامداد، آمدنش فرخ و فرخنده باد»).

(ب) بلند تلفظ کردن مصوت‌های کوتاه پایان کلمه (مثال: هجای سوم مصراع دوم در شعر «نبینی باغبان چون گل بکار / چه مایه غم خورد تا گل برارد»).

(پ) بلند تلفظ کردن کسره‌ی اضافه (مثال: هجای سوم هر دو مصراع در بیت «به نام خداوند جان آفرین، حکیم سخن در زبان آفرین»).

(ت) بلند تلفظ کردن ضمه در واو عطف (مثال: هجای سوم در مصراع اول در بیت «ای چشم و چراغ اهل بینش / مقصود وجود آفرینش»).

(ث) بلند تلفظ کردن فتحه در کلمه‌ی "نه" (مثال: هجای اول مصراع اول در بیت «نه سبو پیدا در این حالت نه آب / خوش بین و لله اعلم بالصواب»).

(ج) بلند تلفظ کردن ضمه‌ی پایان کلمه (مثال: هجای اول مصراع اول «تو کجایی تا شوم من چاکرت / چارقت دوزم کنم شانه سرت»).

(چ) کوتاه تلفظ کردن مصوت بلند "او" (مثال: هجای نهم در مصراع «در چنان روز مرا آرزویی خواهد بود»).

(ح) کوتاه تلفظ کردن مصوت بلند "ای" (مثال: هجای سوم مصراع اول «بیا تا برآریم دستی ز دل / که نتوان برآورد فردا ز گل»).

(خ) شاعر می‌تواند در اول مصراع به جای فعلاتن از فعلاتن استفاده کند (مثال: نفس باد صبا مشک فشان خواهد شد، عالم پیر دگر باره جوان خواهد شد).

(د) شاعر می‌تواند گاهی به جای دو هجای کوتاه از یک هجای بلند استفاده کند (مثال: هجای دوم مصراع اول در شعر «آمد نوروز هم از بامداد، آمدنش فرخ و فرخنده باد»).

(ذ) شاعر می‌تواند بنا به ضرورت وزن گاهی به جای U - از U استفاده کند یا بالعکس. (مثال: کاشکی جز تو کسی داشتمی / یا به تو دسترسی داشتمی)

(ر) هجای آخر مصراع همواره بلند است اما شاعر می‌تواند در آخر مصراع از هجای کوتاه استفاده کند. (مثال: سرو را مانی ولیکن سرو را رفتار نه)

(ز) شاعر می‌تواند در پایان مصراع یک یا دو حرف صامت اضافه بر وزن بیاورد و به جای هجای بلند از هجای کشیده استفاده کند.

البته در تعداد این اختیارات نظرانی متفاوتی در بین ادیبان و زبان‌شناسان وجود دارد چنان‌که نجفی [۱۳] معتقد است: «در شعر فارسی

فقط دو اختیار شاعری داریم که عبارتند از اختیار فعلاتن به جای فعلاتن در آغاز مصراع و اختیار یک هجای بلند به جای دو هجای کوتاه متوالی در میان مصراع اما «قاعده‌ی قلب» را نمی‌توانیم اختیار سوم بدانیم، زیرا شمول ندارد و کاربرد آن تابع شرایطی بسیار محدود است».

۵- ایجاد پایگاه داده‌ی کلمات

برای درست خواندن و درست نوشتن اشعار فارسی که اولین مرحله از تعیین وزن عروضی اشعار است به قواعدی برای حرکت‌گذاری درست کلمات نیاز داریم. برای این کار ما از یک پایگاه داده کلمات به‌عنوان یکی از پیش‌نیازهای طرح استفاده می‌کنیم. در این پایگاه داده مجموعه‌ای از لغاتی را که در اشعار فارسی مورد استفاده قرار گرفته است استخراج کرده و همراه با حرکت‌گذاری صحیح آن و رشته صامت-مصوتش در پایگاه داده‌ی کلمات قرار می‌دهیم.

۵-۱ اختیارات شاعری غیر قابل تشخیص کلمات

ممکن است برای بعضی از کلمات اختیارات شاعری لحاظ شده باشد که در مراحل مختلف قابل تشخیص نباشد. در این صورت برای آن کلمه در پایگاه داده‌ی کلمات در ستونی مجزا رشته‌ی هجاهای کوتاه و بلند حاصل از اختیارات شاعری خاص آن کلمه را قرار می‌دهیم تا در صورتی که برای تعیین وزن به آن نیاز بود استفاده شود. برای مثال فتحه در کلمه‌ی «نه» ممکن است بلند تلفظ شود لذا هر دو صورت تلفظ ممکن را برای این کلمه در پایگاه داده وارد می‌کنیم. با این کار احتمال خطا را کاهش می‌دهیم.

۵-۲ پیشوندها و پسوندهای کلمات

هر کلمه در زبان و ادبیات فارسی دارای پیشوندها و پسوندهای متعددی است که اگر بخواهیم در پایگاه داده‌ی کلمات برای هر کدام از آن پیشوندها و پسوندها مکان مجزایی اختصاص دهیم پایگاه داده‌ی بزرگی خواهیم داشت. برای پیش‌گیری از این کار پیشوندها و پسوندهای هر کلمه را همراه با خود کلمه قرار می‌دهیم.

۵-۳ دسته‌بندی کلمات

فعل، اسم یا حرف بودن یک کلمه می‌تواند تاثیر بسزایی در اعمال اختیارات شاعری و همچنین ایجاد ارتباط بین کلمات مختلف داشته باشد. یکی از مواردی که تاثیر نوع کلمه را برای ما مشخص می‌کند حرکت‌گذاری حرف آخر کلمه است. برای مثال ما برای افعال تنها در صورتی باید به حرکت‌گذاری حرف آخر پردازیم که کلمه‌ی بعدی حرف ربط «و» باشد و در این صورت حرف آخر فعل مصوت «--» اختیار خواهد کرد. از این موضوع می‌توان نتیجه گرفت که حرف آخر افعال مصوت ندارد. همچنین برای حرف ربط «و» در صورتی که در ابتدای مصراع قرار داشته

باشد باید «و» تلفظ شود و در سایر موارد باید به صورت مصوت «و» به کلمه‌ی قبلی اضافه شود.

۶- راه حل پیشنهادی

سامانه‌ی مورد نظر می‌بایست یک بیت شعر فارسی را به عنوان ورودی دریافت کند و در انتها وزن شعر را به ما بدهد. بر اساس مرحله‌ی که در تشخیص وزن عروضی اشعار گفته شد راه حل پیشنهادی را می‌توان در مراحل زیر خلاصه کرد:

۱. تصحیح
۲. حرکت گذاری
۳. تبدیل به رشته‌ی صامت - مصوت
۴. اعمال اختیارات شاعری بر رشته صامت- مصوت
۵. تبدیل رشته مرحله‌ی قبل به رشته هجاهای بلند و کوتاه
۶. اعمال اختیارات شاعری بر رشته هجاهای بلند و کوتاه
۷. انطباق یکی از ۳۵ وزن موجود بر شعر

۶-۱ تصحیح

منظور از تصحیح در این‌جا حذف نویسه‌های اضافی است. نویسه‌های اضافی به آن دسته از نویسه‌هایی اطلاق می‌گردد که قواعد نوشتاری ملزم به رعایت آن‌هاست اما در پردازش متن از جمله در تشخیص وزن شعر کاربردی ندارند که از آن میان می‌توان به موارد زیر اشاره کرد:

- علامات از قبیل ؛ ، ؟ ، ، ، ...
- نویسه‌ی خط تیره که معمولاً برای تراز کردن آخر خط استفاده می‌گردد و ممکن است کاربر از آن‌ها استفاده کند.
- فاصله اضافی که شامل مواردی می‌گردد که کاربر بیش از یک نویسه‌ی فاصله در بین دو واژه استفاده می‌کند.

۶-۲ حرکت گذاری

پس از حذف نویسه‌های اضافی نوبت به حرکت گذاری شعر می‌گردد. برای این کار از پایگاه داده کلمات استفاده می‌کنیم.

در صورتی که واژه‌ای در پایگاه داده کلمات موجود نباشد با استفاده از قوانین حرکت گذاری این کار انجام می‌گیرد:

۱. اگر حرف «الف»، «و» یا «ی» باشد روی آن هیچ حرکتی قرار نمی‌گیرد.
 ۲. اگر حرف «الف» نباشد ولی حرف بعدی «الف» باشد باز هم حرکتی روی آن قرار نمی‌گیرد.
 ۳. روی آخرین حرف کلمه نیز حرکتی قرار نمی‌گیرد.
 ۴. در موارد دیگر روی حرف، مصوت «-» قرار می‌گیرد.
- نکته‌ی مهم در این میان عدم تشخیص حرکت حرف آخر کلمه توسط سامانه است که دقت آن را پایین می‌آورد. همچنین ذکر این نکته ضروری است که نوع مصوت کوتاه با تقریب خوبی تأثیری در وزن شعر ندارد.

۶-۳ تبدیل به رشته‌ی صامت - مصوت

همانطور که گفته شد برای سادگی و تسریع کار در پایگاه داده کلمات ستونی شامل رشته صامت مصوت‌ها را نیز گنجاندیم. به دست آوردن رشته‌ی صامت مصوت برای کلمات پایگاه داده به سادگی قابل انجام است. حال در صورتی که کلمه‌ای در پایگاه داده نباشد با استفاده از قواعد زیر آن را به رشته صامت مصوت‌ها تبدیل می‌کنیم.

۱. به جای صامت‌ها C را قرار می‌دهیم.
۲. به جای هر یک از مصوت‌های کوتاه V قرار می‌دهیم.
۳. صامت‌های بلند را با VV نمایش می‌دهیم.

۶-۴ اعمال اختیارات شاعری بر رشته صامت - مصوت

با استفاده از اختیارات زبانی شاعری رشته صامت مصوت تولید شده را اصلاح می‌کنیم.

از آن جمله می‌توان به اختیار شاعری «ز» که در بالا ذکر شد اشاره کرد. بر اساس این اختیار شاعر قادر است در پایان مصراع یک یا دو حرف صامت اضافه بر وزن بیاورد و به جای هجای بلند از هجای کشیده استفاده کند لذا برای اعمال این اختیار شاعری اگر در انتهای رشته چند C موجود بود آن‌ها را حذف می‌کنیم.

۶-۵ تبدیل رشته مرحله‌ی قبل به رشته‌ی هجاهای بلند و کوتاه

برای این کار باید از اول رشته‌ی تولید شده شروع به تبدیل رشته کنیم. عمل تبدیل با استفاده از دو قانون زیر انجام می‌پذیرد:

۱. هجای کوتاه به دو صورت C یا CV دیده می‌شود و باید به U تبدیل شود.
۲. هجای بلند به صورت CVV یا CVC دیده می‌شود و باید به - تبدیل شود.

۶-۶ اعمال اختیارات شاعری بر رشته هجاهای بلند و کوتاه

در این مرحله دسته‌ی دیگری از اختیارات شاعری بر روی رشته تولید شده اعمال می‌شود. این اختیارات شامل آن دسته از اختیاراتی می‌گردند که در رشته صامت- مصوت قابل تشخیص نبوده‌اند و پس از رسیدن به رشته هجاهای بلند و کوتاه می‌توان آن‌ها را اعمال کرد. از جمله‌ی این اختیارات می‌توان به موارد زیر اشاره کرد:

- هجای آخر مصراع همواره بلند است اما شاعر می‌تواند در آخر مصراع از هجای کوتاه استفاده کند (اختیار شاعری «ر»). لذا آخرین کاراکتر رشته همواره باید «-» باشد.
- شاعر می‌تواند در اول مصراع به جای فعلاتن از فاعلاتن استفاده کند (اختیار شاعری «خ»). با بررسی ۳۵ وزنی که سیستم قادر به تشخیص آن می‌باشد مشاهده می‌کنیم که در صورتی که رکن اول وزن فعلاتن

رشته‌ها است. این الگوریتم تعمیم یافته‌ی الگوریتم فاصله‌ی همینگ است.

۷-۲ الگوریتم پیشنهادی

با توجه به اهمیت و نقش اختیارات شاعری در تعیین وزن اشعار در الگوریتم پیشنهادی مراحل خاصی را به تعیین این اختیارات اختصاص یافته است. الگوریتم پیشنهادی برای تطبیق وزن شامل مراحل زیر می‌باشد:

۱. تشخیص اختیارات شاعری و اعمال آن‌ها در رشته‌ی هجاهای کوتاه و بلند
۲. مقایسه‌ی رشته تولید شده با وزن‌هایی با طول هجای برابر
۳. مشخص کردن وزن‌هایی با کمترین اختلاف
۴. بدست آوردن مکان‌های عدم تطابق وزن‌های مشابه
۵. تعیین کلمات نامنطبق
۶. مقایسه رشته اختیارات شاعری با رشته‌ی وزن‌های مشابه در محل کلمات غیر منطبق
۷. جایگزینی رشته‌ی هجاهای کوتاه و بلند اختیارات شاعری در صورت تطابق

۷-۲-۱ تشخیص اختیارات شاعری و اعمال آن‌ها در

رشته‌ی هجاهای کوتاه و بلند

در گام اول دو دسته‌ی دیگر از اختیارات شاعری را تشخیص و آن‌ها در رشته‌ی هجاهای کوتاه و بلند اعمال می‌کنیم.

برای این کار از الگوریتم XOR استفاده می‌کنیم. در این الگوریتم هر بار نویسه‌ی \bar{a} رشته‌ی اول را با نویسه‌ی a رشته‌ی دوم مقایسه می‌گردد و در صورتی که دو نویسه متفاوت باشند الگوریتم جواب ۱ و در صورت تشابه جواب ۰ را بر می‌گرداند.

بر اساس اختیار شاعری «د» شاعر می‌تواند بنا به ضرورت وزن گاهی به جای U - از \bar{U} استفاده کند یا بالعکس. برای تشخیص این اختیار شاعری می‌توانیم از الگوریتم XOR به این صورت استفاده کنیم: برای هر نویسه اگر پاسخ یک است بررسی می‌کند که اگر برای نویسه بعدی نیز به صورت معکوس پاسخ یک باشد نشان‌گر این است که شاعر از اختیار شاعری فوق استفاده کرده است.

- - \bar{U} - - - \bar{U} - - - \bar{U} - - - \bar{U} - -
- - \bar{U} - - - \bar{U} - - - \bar{U} - - - \bar{U} - -

همچنین بر اساس اختیار شاعری «د» شاعر می‌تواند گاهی به جای دو هجای کوتاه از یک هجای بلند استفاده کند. برای تشخیص این اختیار به وسیله‌ی الگوریتم XOR در صورتی که خروجی الگوریتم برای یک نویسه یک شد و نویسه‌ی رشته‌ی اول U بود، در صورتی که نویسه‌ی دوم رشته‌ی اول هم U باشد و نویسه‌ی سوم رشته‌ی اول برابر نویسه‌ی دوم رشته‌ی دوم بود در این صورت شاعر از اختیار شاعری استفاده کرده است.

باشد همواره رکن دوم نیز فعال‌تر می‌باشد (وزن ۵ و ۳۲) لذا در صورتی که دو رکن اول مصراع‌ی فعال‌تر فعال‌تر بود آن را به فعال‌تر فعال‌تر تبدیل می‌کنیم.

۶-۷ انطباق یکی از ۳۵ وزن موجود بر شعر

آخرین مرحله انطباق رشته تولید شده با یکی از ۳۵ وزن موجود است. در این قسمت رشته تولید شده با ۳۵ وزن مقایسه شده و با توجه به الگوریتم پیشنهادی نزدیک‌ترین وزن موجود به عنوان پاسخ نهایی اعلام می‌گردد.

۷- انطباق رشته

الگوریتم‌های انطباق رشته، مجموعه الگوریتم‌هایی هستند که برای مقایسه یک رشته با مجموعه‌ای از رشته‌ها به کار می‌رود. ورودی این الگوریتم‌ها مجموعه‌ای از رشته‌ها به عنوان پایگاه داده و یک رشته به عنوان مسأله‌ی اصلی است. هدف در این الگوریتم‌ها پیدا کردن رشته‌ای از میان پایگاه داده است که بیشترین شباهت را به رشته‌ی صورت مسأله دارد. میزان شباهت رشته‌ی صورت مسأله با رشته‌های پایگاه داده تحت عنوان برازندگی معرفی می‌گردند.

در این سامانه، ما مجموعه‌ای شامل ۳۵ وزن پر کاربرد شعر فارسی و رشته‌ی هجاهای بلند و کوتاهشان داریم. از طرفی پس از طی مراحل شش‌گانه تعیین وزن، بیت شعر ورودی به رشته‌ی اصلاح شده‌ی هجاهای بلند و کوتاه تبدیل می‌شود. در این‌جا وظیفه‌ی الگوریتم‌های انطباق رشته بیان برازندگی هر یک از ۳۵ وزن موجود با رشته‌ی هجاهای بلند و کوتاه شعر ورودی و معرفی وزنی که بیشترین برازندگی را دارد به عنوان پاسخ مسأله است.

۷-۱ معرفی الگوریتم‌های قابل استفاده

برای تطبیق رشته الگوریتم‌های متعددی وجود دارد که امکان استفاده از آن‌ها در تشخیص وزن عروضی اشعار وجود دارد. در این‌جا به معرفی دو دسته از این الگوریتم‌ها می‌پردازیم.

۷-۱-۲ فاصله‌ی همینگ

الگوریتم فاصله‌ی همینگ یکی از مجموعه الگوریتم‌هایی است که برای مقایسه‌ی رشته‌ها به کار می‌رود. این الگوریتم تنها برای رشته‌هایی با طول برابر کاربرد دارد.

۷-۱-۲ فاصله‌ی لون اشتاین

فاصله‌ی لون اشتاین بیانگر کمترین میزان عملیات برای تبدیل یک رشته به رشته‌ی دیگر است. این معیار به اقتضای ولادیمیر لون اشتاین، که این فاصله را در سال ۱۹۵۶ مطرح کرد، نام گذاری شده است [۱۴].

الگوریتم لون اشتاین تا حدی مشابه تابع difflib است با این تفاوت که difflib برای مقایسه انواع اشیاء است ولی لون اشتاین فقط برای

-- U -- U U -- U -- U --
-- U -- U -- U --

۷-۲-۶ مقایسه رشته اختیارات شاعری با رشته‌ی وزن‌های مشابه در محل کلمات غیر منطبق

پس از تعیین کلمات نامنطبق در صورتی که در ستون اختیارات شاعری پایگاه داده‌ی کلمات برای آن کلمه رشته‌ی هجاهای کوتاه و بلند حاصل از اختیار شاعری خاصی موجود باشد آن رشته را با رشته‌ی وزن یا اوزان مشابه در محل عدم تطابق مقایسه می‌کنیم. در صورتی که آن کلمه در پایگاه داده‌ی کلمات موجود نباشد و با استفاده از قوانین حرکت‌گذاری، حرکت‌گذاری شده باشد به بازگرداندن وزنی که کمترین اختلاف را دارد به عنوان پاسخ بسنده می‌کنیم.

۷-۲-۷ جایگزینی رشته‌ی هجاهای کوتاه و بلند اختیارات شاعری در صورت تطابق

در صورت تطابق رشته با یکی از اوزان می‌توان نتیجه گرفت که اختلاف موجود حاصل از اختیارات شاعری بوده و شعر در وزنی که تطابق صورت گرفته سروده شده است.

۷-۳ وزن‌های مشابه

اگر رشته‌ی هجاهای کوتاه و بلند همه‌ی ۳۵ وزن پرکاربرد شعر فارسی را بنویسیم با نتایج دور از ذهنی برخورداریم خورد. در میان این ۳۵ وزن، اوزانی وجود دارد که رشته‌ی هجاهای کوتاه و بلند آن‌ها دو به دو کاملاً مشابه است و در این صورت سامانه نمی‌تواند یکی از اوزان را به عنوان پاسخ بازگرداند. برای حل این مشکل از یک روش آماری استفاده می‌کنیم. در این روش میزان استفاده‌ی شعرا از هر یک از اوزان را بر حسب درصد بدست می‌آوریم که این کار بیش از این توسط زبان‌شناسان انجام شده است. حال در صورتی که رشته‌ی هجاهای کوتاه و بلند مصراع شعری برابر رشته‌ی هجاهای کوتاه و بلند اوزان مشابه شد سامانه هر دو وزن را به عنوان پاسخ بازگردانیده و احتمال وقوع هر یک را برای کاربر مشخص می‌کند.

۷-۴ وزن‌های نزدیک به هم که با اعمال اختیارات شاعری مشابه می‌شوند

پیش از این با اختیارات شاعری به‌طور کامل آشنا شدیم. استفاده از این اختیارات باعث می‌شد تا رشته‌ی هجاهای کوتاه و بلند شعر با رشته‌ی هجاهای کوتاه و بلند وزن عروضی آن منطبق نباشد. برای انطباق از روش‌هایی که بیش از این معرفی شد استفاده می‌کنیم. اما در میان اوزان شعر فارسی، وزن‌هایی وجود دارد که رشته‌ی هجاهای کوتاه و بلند آن‌ها در صورت اعمال تعداد معدودی از اختیارات شاعری به رشته‌ی هجاهای کوتاه و بلند وزن دیگری تبدیل می‌شود. برای مثال رشته‌ی هجاهای کوتاه و بلند وزن اول و دوم از ۳۵ وزن پرکاربرد را می‌نویسیم:

- U - - - U - - - U - - - U - - -
- - U - - - U - - - U - - - U - - -

بر اساس اختیار شاعری «ذ» شاعر می‌تواند بنا به ضرورت وزن گاهی به جای U - از - U استفاده کند یا بالعکس. حال در صورتی که این

۷-۲-۲ مقایسه‌ی رشته تولید شده با وزن‌هایی با طول هجای برابر

از میان اختیارات شاعری تنها دو اختیار شاعری وجود دارد که بر اساس آن‌ها شاعر به تولید مصراعی با تعداد هجای متفاوت نسبت به وزن شعر می‌پردازد.

مورد اول اختیار شاعری «د» که بر اساس آن شاعر می‌تواند گاهی به جای دو هجای کوتاه از یک هجای بلند استفاده کند و مورد دوم اختیار شاعری «ز» که بر اساس آن شاعر می‌تواند در پایان مصراع یک یا دو حرف صامت اضافه بر وزن بیاورد و به جای هجای بلند از هجای کشیده استفاده کند.

از آنجایی که هر دوی این اختیارات شاعری در مراحل قبلی تشخیص داده شده‌اند می‌توان ادعا کرد تعداد هجای مصراع شعر ورودی باید با تعداد هجای وزن پاسخ برابر باشد لذا مقایسه تنها باید با اوزانی انجام شود که دارای تعداد هجای برابر با مصراع ورودی هستند.

همین‌جا می‌توان به یکی از مشکلات الگوریتم فاصله‌ی لون اشتاین پی برد. در این الگوریتم ما طول رشته‌ها برای ما اهمیت نداشته و اختلاف در طول رشته از نگاه این الگوریتم با اختلاف هجای کوتاه و بلند ارزشی یکسان دارد. این در حالی است که با توجه به این که در زبان و ادب فارسی تعداد هجاها مبنای تعیین وزن اشعار است لذا تساوی تعداد هجاها از اهمیت دوچندانی برخوردار است.

۷-۲-۳ مشخص کردن وزن‌هایی با کمترین اختلاف

پس از مقایسه‌ی رشته‌ی تولید شده با وزن‌هایی با طول برابر در صورتی که شاهد تطابق کامل رشته بودیم وزن تطابق یافته به عنوان پاسخ برگردانده می‌شود. در غیر این صورت وزن یا اوزانی که کمترین تعداد اختلاف نویسه با رشته‌ی تولید شده را دارند و همچنین وزن یا اوزانی که تا یک نویسه بیشتر اختلاف دارند مشخص می‌شوند.

۷-۲-۴ بدست آوردن مکان‌های عدم تطابق وزن‌های مشابه

در این مرحله برای وزن یا اوزان مشابه نویسه‌هایی که در آن‌ها عدم تطابق اتفاق افتاده را مشخص می‌کنیم. برای این کار کافی است در هنگام مقایسه‌ی دو رشته مکان نویسه‌های نامنطبق را ذخیره کنیم.

۷-۲-۵ تعیین کلمات نامنطبق

حال باید کلماتی که عدم انطباق ناشی از آن‌ها بوده است را تعیین کنیم. برای این کار باید مسیری را که در راه حل پیشنهادی پیاده‌ایم نگهداری کنیم تا بتوانیم از رشته‌ی حاصل، کلمات شعر ورودی را مشخص کنیم.

کار را در ۴ مکان از مصراع انجام دهیم وزن اول معادل وزن دوم خواهد شد.

با بررسی اشعار شعر فارسی به این نتیجه می‌رسیم که شاعر در یک بیت از یک اختیار شاعری بیش از یک بار استفاده نخواهد کرد. بر این اساس برای جلوگیری از وقوع چنین اشتباهاتی در دو مورد از اختیارات شاعری که هر دو با الگوریتم XOR توسط سامانه تشخیص داده می‌شوند، تنها در صورتی این اختیار شاعری را به رشته‌ی هجاهای کوتاه و بلند بیت ورودی اعمال می‌کنیم که الگوریتم فقط در یک مورد به ما پاسخ ۱ برگرداند.

۸- نتیجه‌گیری

در این مقاله سعی بر آن شد تا ضمن آشنایی با اوزان شعر فارسی و مراحل تعیین وزن اشعار فارسی راه حلی عملی برای طراحی سامانه‌ی تشخیص اوزان شعر فارسی ارائه گردد. در این راه حل با تبدیل یک مصراع شعر به رشته‌ی اصلاح شده‌ی هجاهای بلند و کوتاه طی شش مرحله و انطباق آن با اوزان مختلف شعر فارسی به وسیله‌ی الگوریتم پیشنهادی تشخیص الگو وزن شعر را تشخیص دادیم. گام بعدی پیاده‌سازی این سامانه و آزمون تجربی برای بررسی میزان کارایی آن است.

۹- پیشنهادها

آنچه واضح است دامنه‌ی استثنائات شعر فارسی به قدری زیاد است که قانونمند کردن آن بسیار دشوار است. شناوری هجاهای واقعاً مانع قطعیت قوانین می‌شود و فقط با تحلیل‌هایی موسیقایی می‌توانیم از این سردرگمی‌هایی یابیم [۱]. این از یک سو و از سوی دیگر انجام نگرفتن پژوهش‌های رایانه‌ای در این زمینه باعث نقص‌های فراوانی در سامانه‌ی موجود گردیده است. در زمینه‌ی کاربرد رایانه در شعر فارسی در سال‌های اخیر پژوهش‌های اندکی انجام شده است که از آن جمله می‌توان به شاهمیری، بروجردی [۱۵] اشاره کرد اما در زمینه تشخیص وزن عروضی اشعار فارسی پس از مجیری، مینایی [۱۶] این مقاله دومین پژوهش انجام شده است. برای توسعه و بهبود این پژوهش پیشنهادات زیر ارائه می‌گردد:

۱. استفاده از تمامی اوزان شعر فارسی

۲. استفاده از اوزان شعر نو

۳. استفاده از ساختار شبکه‌های عصبی

۴. پیش‌بینی تمامی اختیارات شاعری

۵. استفاده از توابع دیگر برای تطبیق رشته

مراجع

[۱] محمد کاظم کاظمی، "روزنه؛ مجموعه آموزشی شعر"، انتشارات ضریح آفتاب، ۱۳۷۷.

[2] H. M. Manurung, G. Ritchie, and H. Thompson. A flexible integrated architecture for generating poetic texts. Informatics Research Report EDINF-RR-0016, University of Edinburgh, 2000.

[3] P. Gervas. An expert system for the composition of formal Spanish poetry. Journal of Knowledge-Based Systems 14(3-4):181-188, 2001.

[4] P. Gervas. Wasp: Evaluation of different strategies for the automatic generation of Spanish verse. In Proceedings of the AISB-00 Symposium on Creative & Cultural Aspects of AI, pages 93-100, 2000.

[5] H. M. Manurung, G. Ritchie, and H. Thompson. Towards a computational model of poetry generation. In Proc. of the AISB-00 Symposium on Creative and Cultural Aspects of AI, 2001. CAEPIA 2001.

[۶] سیروس شمیسا، "آشنایی با عروض و قافیه"، انتشارات فردوس، ۱۳۸۳.

[۷] پوران (صدیقه) علیپور، "ریخت شناسی معنایی وزن در شعر فارسی"، پژوهشنامه‌ی زبان و ادب فارسی (گهر گویا)، سال ۳، شماره‌ی ۳، ۱۳۸۸.

[۸] محمدرضا شفیعی کدکنی، "صور خیال در شعر فارسی"، انتشارات آگاه، ۱۳۸۸.

[۹] عباس ماهیار، "عروض فارسی شیوه‌ای نو برای آموزش عروض و قافیه"، نشر قطره، ۱۳۷۴.

[۱۰] تقی وحیدیان کامیار، "عروض و قافیه"، انتشارات آموزش و پرورش، ۱۳۷۹.

[11] S. Elwell, The Persian Metters, Cambridge, 1976.

[۱۲] م فرزاد، "مجموعه شعر فارسی"، مجله خرد و کوشش، ۱۳۴۹.

[۱۳] ابوالحسن نجفی، "درباره‌ی «قاعده‌ی قلب» و چند نکته‌ی عروضی دیگر"، مجله نامه فرهنگستان، سال ۵، شماره‌ی ۲، ۱۳۷۸.

[14] V. I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10 (1966):707-71.

[۱۵] امیرشهاب شاهمیری، محمدرضا مطش بروجردی، "تعیین شاعر به کمک روش‌های یادگیری ماشین"، سومین کنفرانس فناوری اطلاعات و دانش، دانشگاه فردوسی مشهد، ۱۳۸۶.

[۱۶] محمدمهدی مجیری، بهروز مینایی بیدگلی، "تشخیص وزن عروضی اشعار فارسی: کاربرد جدیدی از متن کلوی"، دومین همایش داده‌کلوی ایران، دانشگاه صنعتی امیرکبیر، ۱۳۸۷.