

# هوش مصنوعی و پردازش زبان و فرهنگ ایرانی

امیر شهاب شاهمیری



هوش مصنوعی و پردازش زبان و فرهنگ ایرانی

امیر شهاب شاهمیری

کانون نشم علوم



کانون نشم علوم

## Iranian Culture and Language Processing using Artificial Intelligence

Amir Shahab Shahmiri

هوش مصنوعی دانشی است که می‌کوشد کارها را در سطح و باروشی کمابیش همچو انسان انجام دهد. بخشی از کارهایی که هر فرد روزانه انجام می‌دهد، در حوزه مسایل فرهنگی و زبانی می‌گنجد و شناخت، تشخیص یا دسته‌بندی آنها به درجاتی از هوشمندی نیاز دارد. این کتاب به حل برخی مسایل در زمینه فرهنگ و زبان‌های ایرانی - مانند اشعار کلاسیک فارسی، خط میخی، بازی نرد، ویرایش هوشمند، ریشه‌یابی، و تولید نام‌ها و واژگان تازه در مجموعه زبان‌های ایرانی - به کمک روش‌ها و ابزارهای هوش مصنوعی می‌پردازد.



9789643271800



# هوش مصنوعی و پردازش زبان و فرهنگ ایرانی

مؤلف:

امیرشهاب شاهمیری

ناشر:

کانون نشر علوم



سرشناسه: شاهمیری، امیرشهاب.

عنوان و نام پدیدآور: هوش مصنوعی و پردازش زبان و فرهنگ ایرانی

تألیف: امیرشهاب شاهمیری

مشخصات نشر: تهران: کانون نشر علوم، ۱۳۹۶.

مشخصات ظاهری: ۲۰۶ ص.:

شابک:

وضعیت فهرست‌نویسی: فیا

موضوع: هوش مصنوعی

موضوع: فرهنگ و زبان ایرانی

موضوع: شعر کلاسیک فارسی

موضوع: خط میخی پارسی باستان - هخامنشی

موضوع: بازی تخته نرد

موضوع: روش‌های کشف و تصحیح خطا

شناسه افزوده:

رده‌بندی کنگره:

رده‌بندی دیویی:

شماره کتابشناسی ملی:

نام کتاب: هوش مصنوعی و پردازش زبان و فرهنگ ایرانی

ناشر: کانون نشر علوم

تألیف: امیرشهاب شاهمیری

صفحه‌آرا: فاطمه آدیگوزل‌پور اردبیلی

طرح جلد: بهنام شاه‌حسینی

چاپ اول: تابستان ۱۳۹۶

تیراژ: ۱۰۰۰

چاپ و صحافی: کانون نشر علوم

قیمت:

شابک:

دفتر مرکزی پخش: خیابان انقلاب، خیابان فرح‌رازی، خیابان وحید نظری شرقی، پلاک ۶۵، واحد ۱

تلفن: ۶۶۹۶۱۵۶۸-۶۶۴۹۴۹۱۱



## سخن ناشر

به نام خداوند بخشنده و مهربان

گستره دانش بشری در برابر علم بی‌پایان و نامتناهی حضرت حق (جلّ جلاله) بسیار ناچیز است؛ همچنان که کریمه «و ما عطا نکردیم به شما علم را جز اندکی» شاهدی است بر ضعف معرفت انسانی نسبت به حکمت صمدانی. با این حال چه نیکو سفارش فرموده است خاتم انبیاء، حضرت محمد مصطفی (ص)، که «بجوید علم را اگر چه به سرزمین چین». این توصیه، تمامی پیروان راستین دین حنیف اسلام را به فراگیری و اکتساب علوم و فنون مختلف در همه زمان‌ها و همه مکان‌ها فرا می‌خواند.

تلاش کانون نشر علوم، همواره معطوف به تولید آثار شایسته و بایسته در زمینه فنی و مهندسی بوده است؛ اگرچه که در این راه مشکلات و دشواری‌های فراوانی حادث شده و خواهد شد. این امر تا حدّی بسیار، تابعی از ماهیت رشته‌های فنی و مهندسی و معارف مربوط به آنهاست؛ چرا که این رشته‌ها به سرعت و بی‌وقفه در حال نو به نو شدن می‌باشند. این امر رسالت کانون نشر علوم را در نشر آثار مربوطه، اعم از تألیف یا ترجمه، خطیرتر کرده و می‌کند. در این ارتباط دغدغه اصلی کانون نشر علوم این بوده و هست که اولاً، از قافله علوم فنی و مهندسی عقب نماند؛ و ثانیاً، آثاری را به زیور طبع بیاراید که ضمن داشتن وجاهت علمی، در خور شأن مخاطبین اندیشمند و فاضل نیز باشد. امید است که خداوند متّان در این مهم مساعدت فرماید.

پایان سخن اینکه، کانون نشر علوم دست یاری و همّت تمامی مؤلفین و مترجمین علاقه‌مند در زمینه علوم فنی و مهندسی را به گرمی می‌فشارد و تمامی ایشان را خاضعانه و خاشعانه به همکاری فرا می‌خواند. همچنین این مجموعه از تمامی مخاطبین استدعا دارد که با نظرات صائب و راهگشای خود، در بهبود شکلی و محتوایی آثار منتشر شده مساعدت فرمایند؛ و صد البته که تمامی کاستی‌ها از این رهگذر تنها و تنها متوجه این مجموعه بوده و هست.

سید محمدحسین منّوری

کانون نشر علوم



به همسر مریم

که مشوق اصلی من برای تحصیل در گرایش هوش مصنوعی بود.



## فهرست مطالب

<b>فصل ۱: پیشگفتار</b>	<b>۹</b>
پشت پردهٔ مقالات.....	۱۲
<b>فصل ۲: شناسایی اشعار شاهنامه فردوسی به کمک شبکهٔ عصبی مصنوعی</b>	<b>۲۱</b>
۱- مقدمه.....	۲۲
۱-۱- پیشینهٔ پژوهش.....	۲۳
۲-۱- روند انجام کار.....	۲۴
۲- پیش‌نیازهای طرح.....	۲۵
۱-۲- پایگاه داده واژگان فارسی.....	۲۵
۲-۲- انتخاب نمونه‌های آموزشی.....	۲۶
۳-۲- دیگر پیش‌نیازهای علمی.....	۲۷
۳- استخراج ویژگی‌ها.....	۲۷
۱-۳- ویژگی‌های فیزیکی.....	۲۷
۲-۳- ویژگی‌های مفهومی.....	۲۹
۳-۳- ویژگی‌های آوایی.....	۳۰
۴-۳- کاهش ویژگی.....	۳۲
۴- آزمایش نمونه‌ها.....	۳۵
۱-۴- درخت تصمیم.....	۳۵
۲-۴- شبکهٔ عصبی مصنوعی.....	۳۶
۵- نتیجه‌گیری.....	۳۷
مراجع.....	۳۹
پیوست ۱.....	۴۱
<b>فصل ۳: تعیین شاعر به کمک روش‌های یادگیری ماشین</b>	<b>۴۳</b>
۱- مقدمه.....	۴۴
۲- پیشینهٔ پژوهش.....	۴۵
۳- استخراج ویژگی.....	۴۷
۱-۳- ویژگی‌های فیزیکی.....	۴۸
۲-۳- ویژگی‌های مفهومی.....	۴۸
۳-۳- ویژگی‌های آوایی.....	۴۸
۴- تعیین شاعر به کمک روش‌های دسته‌بندی.....	۴۹
۱-۴- درخت تصمیم.....	۴۹
۲-۴- شبکهٔ عصبی مصنوعی.....	۵۰
۳-۴- تعیین شاعر توسط انسان.....	۵۱
۵- نتیجه‌گیری و پیشنهاد.....	۵۳
مراجع.....	۵۴
<b>فصل ۴: تعیین ریشهٔ زبانی واژگان فارسی و عربی به کمک شبکه عصبی مصنوعی</b>	<b>۵۷</b>
۱- مقدمه.....	۵۸
۲- پیشینهٔ پژوهش.....	۵۹



۶۱	۳- شناخت مساله.....
۶۱	۳-۱- پیشینه زبان‌های جهان.....
۶۳	۳-۲- واژگان فارسی و لغات عربی.....
۶۴	۳-۲-۱- واژگان با ریشه ایرانی.....
۶۴	۳-۲-۲- واژگان با ریشه عربی.....
۶۴	۳-۲-۳- واژگان همانند.....
۶۵	۴- استخراج ویژگی.....
۶۷	۴-۱- کد دوگانه.....
۶۷	۴-۲- برآیند نفوذ.....
۶۹	۴-۳- کد دوگانه فشرده.....
۶۹	۴-۴- میانگین برآیند نفوذ.....
۷۰	۵- تعیین ریشه زبانی واژگان به کمک شبکه عصبی.....
۷۰	۵-۱- شبکه دوقطبی با ورودی ساده.....
۷۰	۵-۲- شبکه دوقطبی با ورودی فشرده.....
۷۰	۵-۳- شبکه با ورودی فازی.....
۷۲	۵-۴- نتایج شبیه‌سازی و دسته‌بندی.....
۷۲	۶- نتیجه‌گیری.....
۷۲	مراجع.....

**فصل ۵: تصحیح خودکار غلط‌های تایپی فارسی به کمک شبکه عصبی مصنوعی ترکیبی ۷۵**

۷۶	۱- مقدمه.....
۷۹	۲- پیشینه پژوهش.....
۸۰	۳- غلط تایپی.....
۸۰	۳-۱- روش‌های خودکار تصحیح غلط املایی.....
۸۱	۳-۲- غلط املایی در زبان فارسی.....
۸۲	۳-۳- غلط تایپی در زبان فارسی.....
۸۴	۳-۴- تصحیح املایی بر پایه فاصله.....
۸۶	۴- پیش‌نیازهای تصحیح خودکار غلط تایپی.....
۸۶	۴-۱- تعریف فاصله (شعاع همسایگی) کلیدها.....
۸۷	۴-۲- کدهای ورودی شبکه عصبی.....
۸۹	۴-۳- فرهنگ واژگان به کار گرفته شده در شبیه‌سازی.....
۸۹	۵- شبیه‌سازی.....
۸۹	۵-۱- شبیه‌سازی با شبکه عصبی هاپفیلد.....
۹۳	۵-۲- شبیه‌سازی با شبکه عصبی پرسپترون چندلایه.....
۹۵	۵-۳- ارزیابی نتایج.....
۹۵	۵-۴- بهبود مشکل ظرفیت شبکه.....
۱۰۰	۵- جمع‌بندی.....
۱۰۱	مراجع.....

**فصل ۶: بازشناسی حروف خط میخی پارسی باستان به کمک گراف‌های مکانی ۱۰۳**

۱۰۴	۱- مقدمه.....
۱۰۵	۱-۱- رده‌بندی خط میخی و زبان پارسی باستان.....



۲-۱- پیشینه خط میخی پارسی باستان ..... ۱۰۶

۳-۱- ویژگی‌های خط میخی پارسی باستان ..... ۱۰۶

۴-۱- مراحل یک روش بازشناسی نوری حروف ..... ۱۰۸

۲- پیشینه پژوهش ..... ۱۰۸

۳- گراف‌های فضایی و مکانی ..... ۱۱۰

۴- بازشناسی نوری خط میخی ..... ۱۱۱

۱-۴- ویژگی‌های خط میخی از دیدگاه پردازش تصویر ..... ۱۱۱

۲-۴- پیش پردازش ..... ۱۱۲

۳-۴- تشخیص اجزای حرف ..... ۱۱۲

۴-۴- بازشناسی حرف ..... ۱۱۲

۱-۴-۴- تشکیل گراف مکانی ..... ۱۱۲

۲-۴-۴- کاهش گراف مکانی ..... ۱۱۴

۵- پیاده‌سازی و نتایج تجربی ..... ۱۱۵

۶- نتیجه‌گیری و پیشنهاد ..... ۱۱۶

مراجع ..... ۱۱۷

**فصل ۷: تولید نام‌های زیبا و معنادار ایرانی به کمک الگوریتم ژنتیک با تابع برازندگی مبتنی بر شبکه عصبی مصنوعی**

۱۲۱

۱- مقدمه ..... ۱۲۲

۲- پیشینه پژوهش ..... ۱۲۴

۳- ویژگی نام‌ها و واژگان ایرانی ..... ۱۲۵

۱-۳- نام و معنای آن ..... ۱۲۵

۲-۳- نام از نظر ساختاری ..... ۱۲۶

۳-۳- ریشه زبانی نام‌های ایرانی ..... ۱۲۷

۴-۳- زبان‌های آریایی ..... ۱۲۷

۵-۳- زبان‌های ایرانی ..... ۱۲۸

۶-۳- واژه‌سازی در زبان فارسی ..... ۱۲۸

۷-۳- جابه‌جایی واژه‌ها در زبان فارسی ..... ۱۳۱

۴- روش انجام کار ..... ۱۳۲

۱-۴- مدل‌سازی ..... ۱۳۲

۲-۴- ساخت پایگاه نام‌های آغازین ..... ۱۳۴

۳-۴- تهیه پرسش‌نامه و نظرسنجی از جامعه آماری ..... ۱۳۵

۴-۴- استخراج ویژگی ..... ۱۳۵

۵-۴- شبکه عصبی به‌عنوان تابع برازندگی ..... ۱۴۰

۶-۴- بردارهای ورودی و هدف ..... ۱۴۰

۷-۴- الگوریتم ژنتیک ..... ۱۴۴

۸-۴- تحلیل و ارزیابی ..... ۱۴۵

۵- جمع‌بندی و کارهای آینده ..... ۱۴۷

منابع ..... ۱۴۸

**فصل ۸: پیش‌بینی نتایج بازی نرد به کمک روش‌های یادگیری ماشین**

۱- مقدمه ..... ۱۵۴





۱۵۵.....	۱-۱- تاریخچه .....
۱۵۵.....	۲-۱- تعاریف .....
۱۵۷.....	۳-۱- تفاوت‌های قوانین نرد ایرانی و دیگر کشورها .....
۱۵۷.....	۴-۱- راهبردهای نرد .....
۱۵۸.....	۲- پیشینه پژوهش .....
۱۵۹.....	۳- روش انجام کار .....
۱۵۹.....	۳-۱- کدگذاری .....
۱۶۰.....	۳-۲- استخراج ویژگی‌ها .....
۱۶۲.....	۳-۳- تهیه پایگاه دادگان .....
۱۶۲.....	۳-۴- کاهش و رتبه‌بندی ویژگی‌ها .....
۱۶۴.....	۳-۵- ابزار شبیه‌سازی .....
۱۶۴.....	۴- شبیه‌سازی و تحلیل .....
۱۶۴.....	۴-۱- تحلیل آمارهای اولیه .....
۱۶۵.....	۴-۲- شبکه عصبی مصنوعی .....
۱۶۵.....	۴-۳- دسته‌بندی بیزی ساده .....
۱۶۶.....	۴-۴- ارزیابی نتایج .....
۱۶۷.....	۵- جمع‌بندی .....
۱۶۸.....	منابع .....

**فصل ۹: کد پرسک: روشی نو در کشف و تصحیح خطا**

۱۷۲.....	۱- مقدمه .....
۱۷۴.....	۲- تعاریف .....
۱۷۵.....	۳- کد پرسک .....
۱۷۶.....	۳-۱- ساختار روش .....
۱۷۶.....	۳-۲- کدگذاری کد پرسک .....
۱۷۷.....	۳-۳- فرایند کشف خطا .....
۱۷۸.....	۳-۴- فرایند تصحیح خطا .....
۱۸۱.....	۳-۵- بیت‌های بازفرست .....
۱۸۱.....	۴- ویژگی‌های کد پرسک .....
۱۸۱.....	۴-۱- رابطه طول داده و سربراری .....
۱۸۲.....	۴-۲- شرایط رخداد حالت بازفرست .....
۱۸۴.....	۴-۳- احتمال رخداد حالت بازفرست .....
۱۸۵.....	۴-۴- شبیه‌سازی و نتایج .....
۱۸۷.....	۵- طراحی سخت‌افزاری .....
۱۸۷.....	۵-۱- مدار کدگذار .....
۱۸۵.....	۵-۲- مدار کدگشا .....
۱۸۹.....	۶- نتیجه‌گیری .....
۱۹۰.....	مراجع .....

**واژه‌نامه**



فصل

پیشگفتار

فرهنگ و تمدن کشورم و نیز دانش هوش مصنوعی همواره دو موضوع بسیار گیرا و جذاب برای من بوده‌اند و برایم بسیار لذتبخش بوده که بخشی از عمرم را به‌پای پژوهش و گسترش این دو گذاشته‌ام و خوشحال‌تر هم هستم که اکنون می‌توانم ترکیب این دو را در کتابی گرد آورم. این علاقه دو سویه انگیزه اصلی من برای نگارش چنین مقالاتی بوده است.

این کتاب در حقیقت مجموعه‌ای از مقالات علمی من و همکارانم را در نشریات علمی پژوهشی و کنفرانس‌ها در بر دارد که در هر کدام از آنها یک موضوع مرتبط با زبان یا فرهنگ ایرانی به‌کمک برخی ابزارهای هوش مصنوعی پردازش شده است.

## درباره عنوان کتاب

شاید عبارت «پردازش زبان و فرهنگ ایرانی» برای برخی از خوانندگان کتاب نامانوس باشد. پیشتر اصطلاحاتی مانند پردازش گفتار، پردازش زبان و پردازش دانش متداول شده‌اند، اما مگر پردازش فرهنگ هم داریم؟ پاسخ روشن است: زبان و دانش و هنر خود بخشی از فرهنگ هستند و گذشته از این، چه عنوان بهتری برای پردازش مجموعه‌ای از مولفه‌های فرهنگی می‌توان برگزید؟

از سوی دیگر، هم در مقالات و هم بر عنوان کتاب واژه «ایرانی» را بر «فارسی» ترجیح داده‌ام، زیرا زبان، نام و فرهنگ فارسی بخشی (هرچند شاخص‌ترین بخش) از مجموعه بزرگتر زبان و فرهنگ ایرانی است. برای مثال، نام‌های رایج در کشورمان مجموعه‌ای هستند از نام‌های کردی، گیلکی، لری، آذری، ... و البته فارسی.

## هدف از گردآوری کتاب

به‌نظر من یک پروژه و مقاله علمی هوش مصنوعی از دیدگاه روش و اجرا دو جنبه دارد: نخست درک و تعریف و تحلیل مساله به‌منظور یافتن راه حل، و دوم به‌کارگیری ابزاری برای شبیه‌سازی و اجرا. در هر یک از مقالات این کتاب، یک یا چند ابزار هوش مصنوعی یا یادگیری ماشین (مانند شبکه عصبی مصنوعی) به‌کار گرفته شده است. این ابزارها ممکن است در گذر زمان تغییر کنند، بهبود یابند، قدیمی شوند و یا حتی ممکن است چند دهه دیگر به‌کلی منسوخ شوند. مقالات این کتاب هم از این قاعده مستثنا نیستند.

اما تعریف و تحلیل مساله، به‌گونه‌ای که در این مقالات انجام شده، به‌نظر من و برخی دوستان و همکاران صاحب‌نظرم، نقطه قوت آنهاست. تعریف حدود و ثغور مساله، رفتار موضوع مورد بحث در شرایط گوناگون، تبدیل مفاهیم کیفی به مقادیر کمی، و به‌ویژه استخراج ویژگی‌ها در این پروژه‌ها به‌خوبی انجام شده و همچنان می‌تواند الگویی برای کارهای آینده بر همین مسایل یا مسایل همانند باشد. چنان‌که هنوز هم هر از گاه ایمیلی از دانشجویی دریافت می‌کنم مبنی بر این‌که استادش یکی از همین مقالات را موضوع پروژه وی تعیین کرده و از وی خواسته با ابزاری دیگر آن را شبیه‌سازی کند.

به همین دلیل گمان می‌کنم که بخش‌های آغازین هر مقاله برای هر خواننده علاقه‌مند — هر چند ناآشنا با مفاهیم هوش مصنوعی — جذاب و قابل درک باشد.

## بزرگ‌ترین مشکل اجرایی

در راه انجام بیشتر پروژه‌های انجام‌شده، بزرگ‌ترین مشکل من زمان، منابع مالی، دسترسی به پژوهش‌های دیگران، یا بهره‌مندی از همکاران باتجربه نبود، کم‌وبیش همه آنها را به اندازه کافی داشتم. بزرگ‌ترین مشکل من برای کار روی موضوعات مرتبط با زبان و ادبیات، نبود یک پایگاه دادگان جامع و در دسترس از واژگان فارسی بود به گونه‌ای که فیلدهایی همچون نقش‌های دستوری، آوانگاری، ارتباط معنایی با دیگر واژگان و ... را در بر داشته باشد. در برخی از این پروژه‌ها — مانند تعیین شاعر — ناگزیر خودمان دست به طراحی و ساخت چنین پایگاهی از واژگان می‌زدیم و واژگان مورد نیاز را تا جای ممکن به آن وارد می‌کردیم.

باور دارم اگر چنین پایگاه واژگانی در دسترس پژوهشگران زبان و ادب فارسی قرار گیرد، اجرای پروژه‌های پردازش زبان و تبدیل متن و گفتار به یکباره و به گونه‌ای باور نکردنی شکوفا خواهد شد.

## سخن آخر

در پایان این بخش چند نکته کوچک را به آگاهی می‌رسانم:

۱. برای درج مقالات کتاب، بر مقالات اصلی ویرایشی اندک — که برای تبدیل قالب مقاله به کتاب لازم بوده — اعمال شده است.
۲. در مقالات اصلی همواره نام من به‌عنوان نویسنده اصلی نوشته شده و این به دلیل نقش بیشتر من در طراحی، اجرا و نگارش مقالات بوده است. اما همیشه این تعیین ترتیب نام نویسندگان را به احترام، در اختیار استادانم گذاشته‌ام و نظر آنها اعمال شده است.
۳. اینجانب هنوز فهرستی از موضوعات گوناگون در زمینه پردازش هوشمند زبان و فرهنگ ایرانی را در سر دارم که در انتظار پژوهشگران جوان و کوشا برای پیاده‌سازی است. از این‌رو همواره پذیرای همکاری ایشان خواهم بود.

## پشت پردهٔ مقالات

### تعیین شاعر

بهمن ماه سال ۱۳۸۳ دانشجوی کارشناسی ارشد رشتهٔ هوش مصنوعی و رباتیک در دانشگاه صنعتی امیرکبیر بودم. در مجموع و طبق معمول همهٔ دانشگاه‌ها، دانشجویان از هیات علمی راضی نبودند و طرز فکر آنان را قدیمی می‌دانستند. در این میانه، چند استاد تازه‌وارد به جمع هیات علمی افزوده شدند و این توجه دانشجویان به ایشان را به امید گشایش پنجره‌های تازه به فناوری‌های روز برمی‌انگیخت. دکتر سعید شیری نیز یکی از این استادان بودند که من درس یادگیری ماشین را با ایشان می‌گذراندم.

روزی سر کلاس، ایشان به دانشجویان تشریح می‌زدند که از انتخاب پروژه‌های کلیشه‌ای و تکراری دست بردارند و به ایده‌های نو بپردازند. مثلاً به جای کار بر پردازش تصویر برای تشخیص چهره — که در آن زمان بسیار رایج بود — چرا بر اشعار حافظ کار نمی‌کنند؟!

این سخن چون جرقه‌ای برای ذهن من بود و با توجه به علاقه و اطلاعاتی که دربارهٔ زبان و ادب فارسی داشتم، توجه مرا جلب کرد. طی همان کلاس درس، مساله را بررسی کردم و هر الگویی که منجر به تشخیص اشعار حافظ از دیگر شاعران می‌شد را در ذهنم آزمودم و جمع‌بندی کردم. در پایان کلاس خدمت استاد رسیدم و عرض کردم که تشخیص شعر حافظ، به دلیل شباهت‌هایش با برخی اشعار دیگر شاعران، اگر ناممکن نباشد، بسیار دشوار است و من نمی‌توانم آن را به نتیجه برسانم (کاری که البته بعداً انجامش دادم) اما به گمانم می‌توانم کار بر اشعار شاهنامهٔ فردوسی را به فرجام برسانم و اگر ایشان موافق باشند، فردوسی را به جای حافظ برگزینم. پاسخ ایشان هم این بود: «چه فرقی می‌کند؟ هر شاعری را که انتخاب کنی خوب است!»

در آن زمان هم‌کلاسی صمیمی من آقای رسول دژکام بودند؛ پسری باهوش و مودب که برنامه‌نویسی‌اش هم از من بهتر بود. پس از گفت‌وگوها، استاد پذیرفتند به دلیل بنیادین بودن کار، پروژه را دو نفری انجام دهیم (هرچند که بعداً نمره‌اش را هم نصف کردند!) به طبع بیشتر کارهای پژوهشی و مدل‌سازی بر عهدهٔ من بود و من هم برای پژوهش روشی را پیش گرفتم که هنوز هم آن را پی می‌گیرم: درو کردن منابع. صدها مقاله و چندین کتاب را دیدم و خواندم که البته به علت بدیع بودن موضوع پروژه، بیشترشان به مباحث مرتبط با کار ما می‌پرداختند؛ نه دقیقاً خود کار. همچنین ناگزیر — و البته از روی علاقه — علم عروض و قافیه را نیز آموختم و از چند استاد برجستهٔ زبان و ادب فارسی یاری گرفتم که بیش از همه آقای دکتر مهدی ماحوزی (که بعداً استاد مشاور پایان‌نامهٔ من شدند) و پسرشان دکتر حسین ماحوزی و نیز جناب آقای فریدون جنیدی به من کمک کردند. نیاز دیگر این پروژه نیز به پایان رسانیدن یک کار قدیمی‌ام یعنی طراحی فونت خط اوستایی بود، زیرا باور داشته و دارم که این خط برای آوانگاری واژگان ایرانی بهتر از فونتیک لاتین عمل می‌کند.

سرانجام دست‌آورد این پژوهش و پروژه، در قالب مقاله‌ای با عنوان «شناسایی اشعار شاهنامهٔ فردوسی به کمک

شبکه عصبی مصنوعی» به یازدهمین کنفرانس انجمن کامپیوتر ایران (CSICC 2006) فرستاده شد. اما در این راه دو اتفاق افتاد که باعث افتخار و غرور ما و افزایش اعتماد آقای دکتر شیرینی به من شد، تا جایی که تا امروز افتخار دوستی و همکاری با ایشان را نصیب من کرده است. اول این که داوران به جای تذکر ایرادات و اشکالات مقاله — چنان که مرسوم کار داوران مقالات است — به تمجید و تعریف از مقاله پرداخته بودند و دوم این که کنفرانس خود درخواست کرد که با توجه به این که سطح مقاله بالاتر از کنفرانس بوده، در صورتی که نسخه کامل تری هم داریم، بفرستیم تا در نشریه انجمن علمی پژوهشی انجمن کامپیوتر ایران نیز منتشر شود. ما نیز با کمال میل چنین کردیم و مقاله‌ای که در فصل دوم این کتاب آمده، همان نسخه نشریه (JCSE 2006) است و از درج مقاله کنفرانس، به دلیل همانندی بسیار، خودداری کرده‌ام. البته هر چند تاریخ انتشار این شماره نشریه پاییز ۱۳۸۵ درج شده، اما در عمل چاپ آن یکی دو سال بعد رخ داد!

افزون بر این‌ها، این مقاله در کتاب «زبان فارسی و رایانه» — که مجموعه‌ای از مقالات برگزیده ادوار کنفرانس انجمن کامپیوتر ایران در زمینه پردازش زبان فارسی، به کوشش دکتر محمود بی‌جن‌خان و حسین صامتی بود — در سال ۱۳۸۹ درج شد.

اما داستان مقاله بعدی کمی متفاوت است. پس از آن که استاد راهنمای پایان‌نامه آقای دکتر بروجردی را برگزیدم، ایشان با دانشکده اختلافی پیدا کردند و حدود یک سالی کار پایان‌نامه دانشجویان ایشان — حتی در حد تایید عنوان پایان‌نامه — متوقف شد. از آنجا که من نمی‌خواستم زمان ارزشمندم را از دست بدهم، تصمیم گرفتم همزمان بر چند موضوع مختلف کار کنم تا پس از رفع کدورت‌ها و احتمالاً تایید یکی از عنوان‌ها توسط استاد، دستم پر باشد و بتوانم زودتر از پایان‌نامه دفاع کنم.

همین اتفاق هم افتاد و استاد موضوع «تایید شاعر» را پسندیدند. پس از کسب اجازه از آقای دکتر شیرینی — که با بزرگواری پذیرفتند، موضوعی که با ایشان کار کرده بودم، عنوان پایان‌نامه شود — شورای دانشکده مقرر کرد تعداد شاعران از یک (فردوسی) به چهار (حافظ شیرازی، مولوی بلخی و خیام نیشابوری) افزایش یابد و روش دسته‌بندی دیگری (درخت تصمیم) نیز افزوده و با نتایج شبکه عصبی مقایسه شود. البته تشخیص اشعار چهار شاعر در عمل ساده‌تر از کار بر اشعار یک شاعر بود. چون تشخیص شعر یک شاعر، مسأله «شناسایی» شعر وی از میان بی‌نهایت شعر و عبارات دیگر است، اما وقتی چهار شاعر داریم، مسأله به «تعیین» تبدیل می‌شود و جامعه آماری محدود به شعر خود آنهاست. از این‌رو به‌جز اندکی کار بر گردآوری پایگاه دادگان و پیاده‌سازی درخت تصمیم، بار چندانی بر کار من نیفزود.

البته خودم نیز برای محکم‌کاری از طریق پرسش‌نامه‌هایی، توان تشخیص انسان در سه جامعه آماری گوناگون را سنجیدم تا معیاری برای مقایسه عملکرد سیستم خودکار داشته باشم. جالب آن که دقت تشخیص انسان حدود یک‌سوم ماشین بود!

به هر حال کار به انجام رسید و نتیجه آن در پایان‌نامه و مقاله «تعیین شاعر به کمک روش‌های یادگیری

ماشین» مکتوب شد. طبق نظر استادان و دوستانم این مقاله قابلیت انتشار در یک نشریه علمی پژوهشی را داشت، اما از آنجا که به تجربه دریافته بودم از ارایه تا انتشار مقاله در نشریات دو سالی (و گاه تا چهار سال) طول می کشد، و از طرفی انتشار یک مقاله ماخوذ از پایان نامه از شرایط دفاع از پایان نامه و فارغ التحصیلی بود، آن را به کنفرانس فناوری اطلاعات و دانش — یا (IKT 2007) که آن هم از معتبرترین کنفرانس های کشور است — فرستادم تا زودتر منتشر شود.

در پایان این را هم بگویم که آرزو داشتم در مقطع دکترا باز هم در گرایش هوش مصنوعی و رباتیک تحصیل کنم و عنوان پایان نامه ام چیزی باشد شبیه: «کامپیوتر شاعر». اما به هر تقدیر چنان نشد و در رشته دیگری تحصیل کردم که خوشبختانه به آن نیز بسیار علاقه مندم. اما هنوز رویای به انجام رسانیدن آن طرح بر دلم مانده؛ هر چند آن را کاملاً عملی می دانم.

## تعیین ریشه زبانی واژگان فارسی و عربی

نیمسال دوم سال تحصیلی ۱۳۸۳-۴ دانشجوی آقای دکتر محمدرضا صفابخش — از بلندپایه ترین استادان دانشگاه — در درس شبکه های عصبی مصنوعی بودم. استنباط من این بود که استاد نظر خوشی نسبت به توانایی های من ندارند و این ممکن است بر نمره آخر ترم من تاثیر بگذارد، زیرا شایع بود که ایشان با پذیرش کسانی که هنگام تحصیل کار هم می کنند و نیز کسانی که دانش آموخته کارشناسی از دانشگاه آزاد باشند، در دانشگاه صنعتی امیرکبیر مخالف هستند و البته من هر دو ویژگی را داشتم! و خلق و خوی نسبتاً خشک استاد هم این ظن را تقویت می کرد. از این رو کوشیدم تا با کوشش بیشتر و به ویژه نوشتن مقاله نمرات بیشتری به دست آورم.

خوشبختانه یکی از وظایف اصلی شبکه های عصبی، دسته بندی بود و من هم نه تنها آن را خوب آموخته و کار عملی با انواع آن را خوب دریافته بودم، بلکه انبوه مسایل فرهنگی و ادبی را در ذهن داشتم که مرا برای طرح در یک پروژه عملی فراموش خواندند. در حقیقت از هنگامی که رشته هوش مصنوعی را آغاز کردم، هر پدیده ای را به شکل مجموعه ای از ویژگی ها<sup>۱</sup> می دیدم که اگر آن ویژگی ها خوب تعریف و کمی شوند، می توان با دسته بندی کننده ای<sup>۲</sup> مناسب آن را به طور خودکار بازشناسی<sup>۳</sup> کرد.

از آنجا که دانش و مهارتی در زمینه زبان های باستانی ایران و ریشه شناسی واژگان زبان های آریانی (هندواروپایی) و عربی داشتم، یکی از نخستین موضوعاتی که ذهنم را درگیر کرد، تشخیص ریشه زبانی واژگان فارسی و عربی بود که پس از تایید استاد، باز هم با همکاری آقای رسول دژکام، بدان مشغول شدیم و نتیجه آن در قالب مقاله «تعیین ریشه زبانی واژگان فارسی و عربی به کمک شبکه عصبی مصنوعی» به دوازدهمین کنفرانس انجمن کامپیوتر ایران (CSICC 2007) فرستاده و پذیرفته و منتشر شد.

1. feature
2. classifier
3. recognize

راستش همیشه فکر می‌کردم - به دلیل به کارگیری شبکه عصبی کوچک و ساده - در میان کارهایم این ساده‌ترین کار بوده، اما هنگامی که آن را برای درج در این کتاب ویرایش می‌کردم، کاری پربار یافتم که قابلیت نشر در یک نشریه علمی پژوهشی را نیز داشته است. این مقاله هم در کتاب «زبان فارسی و رایانه» گنجانیده شده است.

روش تشخیص ریشه زبانی ارایه شده در این مقاله را می‌توان به زبان‌های دیگر هم تعمیم داد و دست‌کم واژگان با ریشه ترکی و لاتینی را بدان افزود. تنها مساله، برگرفتن روشی واحد برای نگارش واژگان است که آن هم به سادگی حل می‌شود.

استاد محترم به خاطر نگارش این مقاله یک نمره به مجموع نمرات ما افزودند!

## تصحیح خودکار غلط تایپی

در همان کلاس درس شبکه عصبی برای محکم‌کاری مقاله دومی را هم کار کردیم! هنگامی که شبکه‌های انجمنی یا تداعی‌کننده<sup>۱</sup> - که قابلیت حذف نویز و اختلالات ورودی را دارند - تدریس می‌شد، به این فکر افتادم که می‌توان از آنها برای تصحیح غلط املایی و تایپی بهره جست. آقای دژکام مثل همیشه یار بود و پس از تایید استاد کار بر این پروژه درسی را نیز آغاز کردیم.

اکنون که این مقاله را می‌خوانم، می‌بینم که بخش‌های تعریف، شناخت و برکشیدن ویژگی‌ها از مساله، از بهترین کارهای من بوده است. دست‌آورد کار در مقاله‌ای با عنوان «تصحیح خودکار غلط‌های تایپی فارسی به کمک شبکه عصبی مصنوعی ترکیبی» به تشخیص آقای دکتر صفابخش به نشریه علمی پژوهشی انجمن مهندسين برق و الکترونیک ایران ارسال شد و پس از سال‌ها انتظار در سال ۱۳۸۷ به انتشار رسید. استاد یک نمره دیگر هم برای نگارش این مقاله به نمرات ما افزودند!

## بازشناسی خط میخی پارسی باستان

در نیمسال اول سال تحصیلی ۱۳۸۳ در محضر آقای دکتر محمدرحمتی، درسی داشتیم به نام «بازشناسی آماری الگو» درس بار ریاضی سنگینی داشت و درک آن به اندازه کافی دشوار بود، اما حضور چند دانشجوی دکترای ریاضی در کلاس - که درس را بهتر از دانشجویان کارشناسی ارشد هوش مصنوعی می‌فهمیدند - شرایط را برای ما سخت‌تر می‌کرد. گذشته از این در کشاکش آزمون‌های پایان ترم گرفتار بیماری شدم که داروهای مسکن قوی آن از توانایی ذهنی من می‌کاست. پروژه درسی که انتخاب کرده بودم، در زمینه «بازشناسی خط میخی» بود که آن را هم نتوانسته بودم به فرجام برسانم و بدون آن نمره من به مرز قبولی نمی‌رسید.



این عنوان پروژه را از آن‌رو برگزیده بودم که شناخت خوبی که از خط‌های ایران باستان (به‌ویژه پهلوی و اوستایی) داشتم و البته به دلیل ماهیت فیزیکی خط میخی پارسی باستان، بازشناسی آن ساده‌تر و خاص‌تر از دیگر خطوط بود.

به هر حال، خدمت ایشان رسیدم و برای نخستین بار در زندگی تحصیلی‌ام از استاد تقاضای مهلت بیشتر کردم. ایشان نیز با متانت و حتی شرم از این‌که در چنین موقعیتی قرار می‌گرفتم، تقاضایم را رد کردند و گفتند که نمرات باید زودتر رد شود. چاره‌ای نبود، روزی دیگر (برای دومین و آخرین بار در طول زندگی‌ام) تقاضایم را مطرح کردم و قول دادم که پروژه را پس از تخصیص نمرات به‌انجام خواهند رسانید و ایشان با آن‌که فرمودند به‌تجربه به ایشان ثابت شده دانشجوی پس از دریافت نمره کارش را انجام نمی‌دهد، با خواهش من موافقت کردند.

اما متأسفانه زمان به‌سرعت می‌گذشت و من درگیر تحصیل و کار و تدریس بودم و فرصتی برای ادای دینم پیش نمی‌آمد و من از این بابت هر روز شرمگین‌تر می‌شدم. تا آن‌که روزی از روزهای تدریس در دانشگاه علمی کاربردی واحد داده‌پردازی، دانشجویی باهوش و علاقه‌مند به‌نام آقای مجتبی فرامرزی از من تقاضای موضوع پایان‌نامه کردند. از آنجا که من بسیار گزیده و سخت‌گیرانه موضوع پایان‌نامه و دانشجو را انتخاب می‌کردم، همان‌که کسی سوی من می‌آمد نشان از تمایل وی برای کار واقعی (یا البته گاه بی‌اطلاعی کامل او از شرایط!) بود. از این‌رو همان عنوان پروژه را با وی مطرح کردم و او نیز آغاز به کار کرد.

هرچند تمایل من به استفاده از شبکه عصبی یا یکی از روش‌های ابزارهای دسته‌بندی‌کننده آماری بود، اما آقای فرامرزی نظرم را برگرداندند و گراف‌های مکانی را پیشنهاد دادند که انتخابی هوشمندانه و مناسب‌تر برای بازشناسی خط میخی بود. ایشان شناخت و استفاده از این ابزار را نزد آقای دکتر کاظم فولادی قلعه — که خودشان این ابزار را بر فرمول‌های ریاضی آزموده بودند — آموختند و این از معدود مواردی بود که دانشجویی نظر مرا تغییر می‌داد یا بر طرح پروژه اثر مهمی می‌گذاشت.

البته با آن‌که ایشان پروژه را به‌خوبی اجرا کردند، اما پایان‌نامه‌شان به‌هیچ وجه معیارهای مرا برآورده نمی‌ساخت و به‌ویژه در تعریف مساله و حدود و ثغور آن و بررسی پیشینه پژوهش به‌روشنی کاستی داشت. از این‌رو برای تهیه مقاله علمی و بازپرداخت بدهی به استادم — که در حین کار نیز از نظراتشان بهره‌مند می‌شدم — ناگزیر خودم دست به‌کار شدم و نقایص را برطرف کردم. نتیجه در قالب مقاله‌ای با عنوان «بازشناسی حروف خط میخی پارسی باستان به‌کمک گراف‌های مکانی» در آمد، اما من تاب چند سال منتظر نشستن به پای انتشار نشریات علمی پژوهشی را نداشتم و از این‌رو آن را به چهاردهمین کنفرانس بین‌المللی انجمن کامپیوتر ایران (CSICC 2009) فرستادم و آن نیز به‌انتشار رسید.

## تولید نام‌های زیبا و معنادار ایرانی

از هنگامی که با الگوریتم ژنتیک آشنا شدم، هویت هر پدیده‌ای را به‌شکل مجموعه‌ای از ژن‌ها می‌دیدم که اگر

خوب شناخته و تعریف شوند، می‌توان با یک تابع ارزیابی مناسب، نسل بعدی بهتری از آنها تولید کرد. صندلی، اتومبیل، کامپیوتر، شعر، موسیقی، نقاشی، برنامه تلویزیونی، غذاها و ... همه حامل ژن‌هایی هستند که ساختار وجودی و هویت آنها را بیان می‌کند و هر روز نمونه‌ها و گونه‌های جدیدی از آنها تولید می‌شود. تابع برآزندگی هم انتخاب مردم است، آن ژنی برای تولید انتخاب می‌شود که بهتر یا قوی‌تر باشد.

سال‌ها بود که می‌خواستیم پروژه‌های را در ترکیب الگوریتم ژنتیک و زبان و فرهنگ ایرانی طراحی کنیم، اما امکانش پیش نمی‌آمد. پروژه با توجه به آن‌که باید با همکاری دانشجویانم انجام می‌شد، نیاز به شرایط خاصی داشت: در سطح دوره کارشناسی باشد و طی یک ترم تحصیلی و توسط یک یا حداکثر دو تن انجام پذیرد. در حقیقت فهرستی از پروژه‌های هوش مصنوعی و ترکیب آن با مسایل فرهنگی و همیشه در دفترچه یادداشتیم همراه بود و هست. اما در دانشگاه‌هایی که تدریس می‌کردم، به‌ندرت دانشجویی که توان و تمایل انجام آنها را داشته باشد، پیدا می‌شد.

یکی از این موضوعاتی که در نظر داشتیم، واژگان و نام‌های ایرانی بود. با توجه به شناخت خوبی که از زبان‌های باستانی ایران و روند تکاملی و تغییرات دست‌ورزبان و واژگان ایرانی — از دوره باستان تا کنون — داشتیم، با خودم فکر می‌کردم چرا نتوانیم این روند تکاملی واژگان ایرانی را شبیه‌سازی کنیم و واژه‌ها و به‌ویژه نام‌های ایرانی جدید تولید نکنیم؟ سرانجام دانشجویی به‌نام خانم بهاره زمانی نظامی از دانشگاه آزاد واحد پرند به این موضوع علاقه‌مند شدند و روند طاقت‌فرسای انجام پروژه را پیمودند. از آنجا که لازم بود نظرات زیبایی‌شناسانه مردم نیز شبیه‌سازی شود، بخشی از فرایند کار به توزیع صدها پرسش‌نامه اختصاص داشت که با توجه به آن‌که چنین کارهایی در دانشکده‌های فنی و مهندسی مرسوم نیست، کار را دشوارتر می‌نمود.

به هر حال کار به انجام رسید و آنقدر خوب درآمد که می‌ارزید برای چاپ در نشریه معتبر چند سال منتظر بمانیم! پروژه در سال ۱۳۹۰ دفاع شد و مقاله‌اش با عنوان «تولید نام‌های زیبا و معنادار ایرانی به کمک الگوریتم ژنتیک با تابع برآزندگی مبتنی بر شبکه عصبی مصنوعی» در سال ۱۳۹۴ در فصلنامه علمی پژوهشی انجمن اطلاعات و فناوری ارتباطات ایران (IJCT) پذیرفته و در سال ۱۳۹۵ منتشر شد.

## پیش‌بینی نتایج بازی نرد

بارها گفته‌ام: تخته نرد مسأله مرگ و زندگی نیست که بشود راحت از کنارش گذشت! با توجه به این‌که اینجانب بهترین نردباز جهان هستم (حتما می‌دانید که این ادعای تقریبا تمام نردبازان تاریخ بشریت بوده و این‌که گاه بازی را می‌بازند، به‌خاطر ننشستن تاس و بدشانسی است، نه بد بازی کردن آنها!) وظیفه خود می‌دانستم بخشی از علم خود را در اختیار دیگران قرار دهم!

از شوخی گذشته، مثل بسیاری دیگر از مردم دنیا و به‌ویژه خاورمیانه، نرد از بازی‌های دلپذیر من نیز بوده است. به‌گمانم شناخت کافی از این بازی و راهبردها و ویژگی‌هایش دارم که بخش مهمی از آن را مرهون مرحوم

وجیه‌الله رضانی گیوی — شاعر چیره‌دست اما نسبتاً گمنام آذربایجان — هستم که از دوستان مرحوم پدرم بودند و هر از گاه به محل کار ایشان سر می‌زدند و من بعد از مدرسه به ایشان می‌پیوستم و بدین ترتیب نه‌تنها در آنجا هزاران دست نرد باختیم، بلکه من پای تخته درس تاریخ و سیاست و ادبیات و فرهنگ نیز می‌گرفتم. بخش دیگری از این مهارت و اعتماد به‌نفس (!) را نیز وامدار خویشاوند ارجمندم آقای ارد عطارپور هستم.

در مقوله هوش مصنوعی، بازی نرد یکی از مباحث مهم درسی به‌شمار می‌آید، زیرا این بازی گونه‌ای خاص از بازی‌هاست. در حالی که شترنج یک بازی رویت‌پذیر و مستقل از شانس و پوکر بازی نیمه‌رویت‌پذیر و وابسته به شانس است، نرد یک بازی رویت‌پذیر وابسته به شانس است و از این‌رو در کنار بازی‌های کلاسیک در سرفصل درس هوش مصنوعی جای دارد. البته این‌که می‌گوییم این بازی وابسته به شانس است، بدان معنا نیست که «شانسی» و مستقل از توانایی‌ها و مهارت‌های بازیکن یا برنامه است، بلکه در آن مهارت برنامه‌ریزی بازیکن باید چنان باشد که از پیشامد هر تاس خود یا حریف، بهترین استفاده را ببرد.

باری، نرد هم در کنار انبوه موضوعات دیگر روی دستم مانده بود و فقدان دانشجوی علاقه‌مند باعث شد که این موضوع — پیش‌بینی نتیجه بازی نرد به‌منظور انتخاب بهترین حرکت — را خودم به‌تنهایی به انجام برسانم. این بار خواستم به‌جز شبکه عصبی که همواره عصای دست من بوده، ابزاری دیگر را هم بیازمایم و آن دسته‌بندی‌کننده بیزی بود. همچنین ناشکیبایی من برای زودتر منتشر شدن این کتاب باعث شد که از ارسال آن به نشریات چشم‌پوشم (چون طبق تعهد، مقاله‌ای که به یک کنفرانس یا نشریه فرستاده می‌شود، نباید پیشتر در جایی منتشر شده باشد) و مقاله‌نهایی را با عنوان «پیش‌بینی نتایج بازی نرد، به‌کمک روش‌های یادگیری ماشین» به نهمین کنفرانس فناوری اطلاعات و دانش (IKT 2017) فرستادم.

## کد پرسک

این مقاله جزو آن دسته از مقالات من است که ارتباطی به هوش مصنوعی ندارد، اما ویژگی‌هایی داشت که باعث شد آن را در پایان این کتاب بگنجانم.

سال‌ها قبل، حدود سال‌های ۷۴ و ۷۵ هنگامی که دانشجوی کارشناسی مهندسی نرم‌افزار بودم، به الگوریتم‌هایی مانند فشرده‌سازی داده‌ها و کشف و تصحیح خطا علاقه داشتم و سعی می‌کردم الگوریتمی جدید بدین منظور طراحی کنم. بنیاد کار روش‌های کشف و تصحیح خطا بر آن است که اگر خطایی هنگام انتقال داده بر آن رخ دهد، گیرنده باید حتماً آن را تشخیص دهد و اگر توانست، آن را مکان‌یابی و در نتیجه تصحیح کند وگرنه کل داده باید دوباره از فرستنده به گیرنده ارسال شود. از دیدگاه من این به آن می‌مانست که هخامنشیان هنگامی که چند تن از تکاوران سپاه جاویدان بیمار، زخمی یا کشته می‌شدند، به‌جای جایگزین کردن آن افراد، کل سپاه را جایگزین کنند! (این سپاه را به آن دلیل جاویدان می‌گفتند که تعداد سربازانش همواره ده هزار تن ثابت بود و اگر کسی بازنشست، زخمی یا کشته می‌شد، فوراً دیگری را جایگزین وی می‌کردند.)

الگوریتمی که من روی آن کار می‌کردم، به‌شیوهٔ ارتشداران هخامنشی عمل می‌کرد و فقط بیت‌های مشکوک به بیماری را جایگزین می‌کرد، نه کل بستهٔ داده را. با شوق فراوان موضوع را با یکی از سخت‌گیرترین استادان دانشگاه، به نام آقای دکتر قاسم میرجلیلی، مطرح کردم تا آن را به‌عنوان موضوع پایان‌نامه برگزینم. اما ایشان به‌جای آن که مرا به‌سویی هدایت کند که وقتم را بر کار روی طرح صرف کنم، مرا به‌سوی یادگیری انجام درست پژوهش علمی راهنمایی کردند. نتیجهٔ آن شد که پس از نزدیک به یک سال عنوان پایان‌نامهٔ من شد: «روش‌های کشف و تصحیح خطا» و دیگر زمانی برای تعریف روش ابداعی خودم نماند!

سال‌ها بعد که به تحصیل دورهٔ کارشناسی ارشد مشغول بودم، باز یاد این الگوریتم افتادم و با توجه به آن که درس مرتبگی نداشتم، به آقای دکتر محمدکاظم اکبری مراجعه کردم. ایشان هرچند از استادان گروه تجارت الکترونیکی بوده و هستند، اما در شبکه و انتقال داده و پردازش توری تبحر دارند و چند پروژهٔ ابررایانهٔ کشور نیز زیر نظر ایشان به بهره‌برداری رسیده است. طرح را توضیح دادم و ایشان از آن استقبال کردند. از ایشان درخواست کردم که با توجه به مشغله‌های فراوانم، یک دانشجوی کاری و مستعد را برای انجام کارهای شبیه‌سازی معرفی کنند و ایشان هم آقای احسان نادری پاریزی را به جمع‌مان افزودند.

خوشبختانه آقای نادری از باهوش‌ترین و بااخلاق‌ترین کسانی بودند که من با ایشان آشنا شده‌ام. کار به‌خوبی به انجام رسید و نتیجهٔ اولیه را در قالب مقاله‌ای برای دستگرمی به یکی از کنفرانس‌های داخلی فرستادیم، اما در کمال تعجب داوران مقاله را رد کردند. گفتیم شاید اشتباهی شده و به کنفرانس دیگری فرستادیم، اما آنجا هم رد شد. من که فکر می‌کردم این کار یکی از بهترین ایده‌های من بوده، دلیل را از استاد پرسیدم و ایشان مساله را از ضعف داوران و عدم باورشان به یک کار نوآورانه دانستند و گفتند که کار اصلی را به خارج از کشور بفرستیم. باز هم عجله کردم و به‌جای آن که نتیجه را به نشریه‌ای بفرستیم، آن را در قالب مقالهٔ "A New Error Correction Code" به International Conference on the Latest Advances in Networks (ICLAN 2007) در فرانسه فرستادیم و مقاله پذیرفته و با موفقیت توسط آقای نادری ارائه شد. چندی پیش هم پیاده‌سازی سخت‌افزاری آن طی مقاله‌ای با عنوان «یک طراحی سخت‌افزاری برای کد کشف و تصحیح خطای پرسک» در کنفرانس بین‌المللی مهندسی و علوم کامپیوتر (CCSE2017) ارائه شد و مقاله‌ای که در این کتاب درج شده در عمل ترکیبی است از ترجمهٔ مقالهٔ اول و مقالهٔ دوم.

نام «پرسک» را هم از این بابت برگزیدم که هم کوتاه‌نوشت per second است که واحد سنجش سرعت انتقال داده است. هم این که persec خود یک واحد نجومی است و نیز کوتاه‌نوشت PERSian Error correction Code نیز هست.

## سیاسگزاری

در پایان بر خود واجب می‌دانم از همهٔ استادان، دانشجویان و دوستانم که مرا در نگارش مقالات این کتاب یاری رسانیده‌اند، سپاسگزاری کنم. همچنین سپاسی ویژه دارم از دوست فرهیخته‌ام جناب آقای دکتر محمدحسین منوری که امکان نشر این کتاب را فراهم آوردند.

امیرشهاب شاهمیری

تابستان ۱۳۹۶

amir@shahmiri.ir

## واژه‌نامه

$k$ -Nearest Neighbor ( $k$ -NN)	$k$ مین نزدیک‌ترین همسایه
Epoch	اپک
Feature extraction	استخراج ویژگی
Pattern recognition	بازشناسی الگو
Optical character reading (OCR)	بازشناسی نوری حروف
Natural language processing / understanding (NLP)	پردازش (فهم) زبان طبیعی
Multi-layer Perceptron	پرسپترون چندلایه
Back propagation	پس‌انتشار
Feed forward	پیشخور
Fitness function	تابع برازندگی
Activation function	تابع فعالیت
Text to speech (TTS)	تبدیل متن به گفتار
Fault tolerance	تحمل پذیری خطا
Recurrent Jordan	جردن بازگشتی (شبکه عصبی)
Associative memory	حافظه یادآور (شبکه عصبی)
Self-organizing	خودسازمانده (شبکه عصبی)
Decision tree	درخت تصمیم
Bayesian classifier	دسته‌بندی بیزی
Naïve Bayesian classifier	دسته‌بندی‌کننده بیزی ساده
Artificial neural network	شبکه عصبی مصنوعی
Simulation	شبیه‌سازی
Text categorization	طبقه‌بندی متن
Error detection and correction	کشف و تصحیح خطا

Spatial graph	گراف مکانی (فضایی)
$n$ -grams algorithm	الگوریتم $n$ گرام
Genetic algorithm	الگوریتم ژنتیک
Hidden layer	لایه پنهان
Correlation matrix	ماتریس همبستگی
Support vector machine (SVM)	ماشین بردار پشتیبان
Text mining	متن کاوی
Hidden Markov model (HMM)	مدل پنهان مارکوف
Backgammon	نرد
Hopfield	هافیلد
Artificial intelligence (AI)	هوش مصنوعی
Machine learning	یادگیری ماشین

# نمایه

## الف

- اسلواکی ۶۲
- اشکانی / اشکانیان ۱۲۴، ۱۰۶، ۴۰
- انزانی ۱۰۷
- انگلیسی ۲۹، ۵۹، ۶۰، ۶۱، ۷۴، ۷۹، ۸۱، ۸۲، ۱۰۹، ۱۱۰، ۱۲۷، ۱۲۹
- اورال - آلتایی ۶۲
- اوستایی ۱۶، ۲۵، ۳۸، ۴۰، ۴۱، ۶۲، ۶۳، ۷۴، ۱۰۵، ۱۲۳، ۱۲۴، ۱۲۸، ۱۲۹، ۱۴۸
- اوکراینی ۶۲
- اویغوری ۶۰، ۷۴
- ایتالیایی ۶۱
- ایرانویج ۷۴
- ایسلندی ۶۲
- ایران ۴۰، ۵۹، ۶۳، ۱۲۵، ۱۵۵، ۱۷۴
- ایران باستان ۱۱۷، ۱۵۵
- ایرانویج ۱۱۷
- ایرانی ۱۰، ۱۷، ۲۴، ۲۷، ۲۸، ۳۸، ۶۰، ۶۲، ۶۳، ۶۴، ۷۴، ۱۱۷، ۱۲۱، ۱۲۳، ۱۲۴، ۱۲۵، ۱۲۶، ۱۲۷، ۱۲۸، ۱۳۴، ۱۳۵، ۱۳۷، ۱۴۷، ۱۵۴، ۱۵۷، ۱۶۹
- ایرانی باستان ۱۰۴، ۱۰۵، ۱۲۸
- ایرانی میانه ۱۰۵، ۱۲۸
- ایرانی نو ۱۰۵، ۱۲۸
- ایسلند ۱۲۷
- ایشتویگو ۱۰۶
- آذربایجان ۱۸، ۷۴، ۱۲۷
- آذری ۱۰، ۱۲۷
- آریانی / آریایی ۱۴، ۶۲، ۶۳، ۱۰۵، ۱۲۷، ۱۲۸، ۱۳۱
- آریاییان ۷۴
- آسی ۱۲۸
- آسیای صغیر ۱۲۷
- آسیای مرکزی ۱۲۷
- آشوری ۱۰۷
- آفریقا / آفریقای ۶۲، ۱۲۷
- آلمانی ۶۰، ۶۱، ۱۰۷، ۱۲۷، ۱۲۹
- آمریکا ۱۰۴، ۱۲۷
- ارتودکس ۱۰۹
- اردو ۶۱
- ارمنی ۵۹، ۶۱، ۷۴، ۱۲۴
- اروپا / اروپایی ۲۹، ۳۴، ۶۲، ۸۲، ۱۲۵، ۱۲۷
- ازبکستان ۱۲۷
- اسپانیایی / اسپانیولی ۶۰، ۶۱، ۶۳، ۱۲۶
- اسپرانتو ۶۲، ۷۴
- استرالیا ۶۲، ۱۲۷
- استونیایی ۶۲
- اسکاندیناوی ۶۲
- اسلام ۲۹، ۶۰، ۶۳، ۶۴، ۱۱۷
- اسلاوی ۶۲



ترکمنستان ۱۲۷	ایلامی ۱۰۷، ۱۰۴
ترکمنی ۱۲۴	<b>ب</b>
ترکی ۳۴، ۲۹، ۶۰، ۶۲، ۶۴، ۱۲۴، ۱۲۵	بابلی ۱۰۷
ترکیه ۱۲۷، ۱۰۶	براهمی سینهایلی ۱۰۹
تلوگو ۱۰۹	بزرگمهر بختگان ۱۵۵
<b>ج</b>	بلغاری ۶۲
جیرفت ۱۵۵	بلوچی ۱۲۸
<b>چ</b>	بنگالی ۱۰۹
چک ۶۲	بنگلایی ۱۱۰، ۱۰۹، ۷۹
چین / چینی ۶۰، ۶۲، ۷۴، ۷۹، ۱۰۷، ۱۰۹، ۱۲۷	بیلوروسی ۶۲
چینی - تبتی ۶۲	بیزانسی ۱۰۹
<b>ح</b>	بیستون ۱۱۷
حافظ شیرازی ۱۲، ۱۳، ۲۵، ۲۶، ۲۹	<b>پ</b>
<b>خ</b>	پارس ۱۲۸
خراسانی ۱۲۶	پارسی ۳۷، ۲۹
خسرو یکم ساسانی ۱۵۵	پارسی باستان ۱۶، ۶۲، ۱۰۳، ۱۰۴، ۱۰۵، ۱۰۶، ۱۰۷، ۱۱۱، ۱۱۲
خیام نیشابوری ۱۳، ۲۵، ۲۶	پاسارگاد ۱۱۷
<b>د</b>	پرتغالی ۶۰، ۶۱
داریوش هخامنشی ۱۰۶	پشتو ۱۲۸
دانمارکی ۶۲	پهلوی ۱۶، ۴۰، ۶۲، ۶۳، ۷۴، ۱۱۷، ۱۲۳، ۱۲۴، ۱۲۹، ۱۵۵
دری ۶۲، ۱۲۸	<b>ت</b>
دریای پارس ۶۰، ۷۴	تاتی ۱۲۷، ۱۲۸
دیوانی ۶۴	تاجیکی ۱۲۸
<b>ر</b>	تازیان ۱۰۶
روسی ۶۲	تایلندی ۱۱۰
رومی ۶۱	تخت جمشید ۱۰۴
	ترکان ۲۹

ز

زرتشتی ۱۵۵

زروان / زروانی ۱۶۸، ۱۵۵

ژ

ژاپنی ۱۰۹، ۱۰۷، ۷۹، ۶۲، ۶۰

ژرمنی ۱۲۷، ۶۱

س

ساسانی ۱۵۵، ۱۲۸، ۱۲۴، ۴۰

سامی ۶۴، ۶۳، ۶۲، ۲۹

سانسکریت / سنسکریت ۱۲۹، ۱۲۸، ۱۲۴، ۶۲، ۵۹

سپاه جاویدان ۱۷۴، ۱۷۲، ۱۸

سریانی ۶۴، ۶۲

سريلانكا ۱۰۹

سعدی شیرازی ۲۶، ۲۵

سغدی ۱۲۴

سکایی ۱۰۵

سوئدی / سویدی ۷۴، ۶۲، ۶۰

سومری ۷۴

سومریان ۵۹

سومری ۱۱۷

سومریان ۱۰۶، ۱۰۴

سین‌کیانگ ۱۲۷

ص

صربی ۶۲

ع

عبری ۱۲۴، ۶۴، ۶۲

عرب ۷۴، ۶۰، ۲۶

عربی ۱۴، ۲۹، ۳۵، ۵۸، ۶۰، ۶۱، ۶۲، ۶۳، ۶۴، ۶۵، ۶۶، ۶۷

۶۸، ۷۰، ۷۲، ۷۴، ۸۱، ۸۲، ۱۰۹، ۱۱۰، ۱۲۴، ۱۲۵

۱۲۶، ۱۳۰، ۱۴۸

غ

غزنویان ۱۰۶

ف

فارسی ۱۰، ۱۱، ۱۳، ۱۴، ۲۲، ۲۵، ۲۷، ۲۹، ۳۰، ۳۱، ۳۲، ۳۸

۴۰، ۵۸، ۵۹، ۶۰، ۶۱، ۶۳، ۶۴، ۶۵، ۶۶، ۶۷، ۶۸، ۷۰

۷۲، ۷۴، ۷۵، ۷۸، ۸۱، ۸۲، ۸۳، ۸۴، ۸۷، ۱۰۶، ۱۲۲

۱۲۳، ۱۲۴، ۱۲۷، ۱۲۸، ۱۲۹، ۱۳۰، ۱۳۱، ۱۳۲، ۱۳۴

۱۳۵، ۱۳۶، ۱۳۷، ۱۳۸، ۱۴۱، ۱۴۲، ۱۴۳، ۱۴۴، ۱۴۸

فارسی باستان ۱۲۸، ۱۲۹

فرانسه ۱۹، ۲۹، ۶۰، ۶۱، ۸۲، ۱۲۶

فردوسی ۱۲، ۱۳، ۲۱، ۲۲، ۲۴، ۲۵، ۲۶، ۲۷، ۲۹، ۳۰، ۳۲

۳۵، ۳۶، ۳۷، ۴۰، ۱۴۸، ۱۵۵، ۱۶۸، ۱۶۹

فنلاندی ۶۲

ق

قبطی ۶۴، ۶۲

قرآن مجید ۶۰، ۶۳، ۷۴

ک

کردی ۱۰، ۱۲۴، ۱۲۶، ۱۲۸

ش

شاهنامه ۱۲، ۲۱، ۲۲، ۲۴، ۲۵، ۲۶، ۲۹، ۳۰، ۳۲، ۳۵، ۳۷

۳۸، ۴۰، ۷۴، ۱۴۸، ۱۵۵، ۱۶۸، ۱۶۹

شهر سوخته ۱۵۵، ۱۶۸

شوش ۱۱۷

شیکاگو ۱۰۴

هـ

هخامنشی / هخامنشیان ۱۸، ۱۹، ۱۰۴، ۱۰۶، ۱۱۷، ۱۲۴،

۱۲۸، ۱۷۴

هگمتانه ۱۰۶

هلندی ۲۹

هلنی ۱۰۹

همدان ۱۰۶

هند / هندوستان ۱۲۷، ۱۵۵

هندواروپایی ۱۴، ۱۰۵، ۱۲۶، ۱۲۷، ۱۳۱، ۱۴۸

هندوایرانی ۱۰۵، ۱۲۷، ۱۳۱

هندوژرمنی ۱۲۷

هندی ۶۰، ۶۱، ۷۹، ۱۰۹، ۱۵۵

ی

یونانی ۲۴، ۵۹، ۶۱، ۶۲، ۶۴، ۱۱۰، ۱۰۶، ۱۲۴، ۱۲۹

کره‌ای ۶۰، ۶۲

کورش (بزرگ) هخامنشی ۱۰۶، ۱۱۷

کوفی ۶۴

گ

گرجستان ۱۲۷

گیلکی ۱۰، ۱۲۶

ل

لاتین / لاتینی ۵۹، ۶۲، ۶۵، ۱۰۹، ۱۱۰، ۱۲۴، ۱۲۹

لری ۱۰، ۱۲۴

لهستانی ۶۲

م

ماد / مادی ۱۰۵، ۱۰۶، ۱۲۴

مازندران / مازندرانی ۷۴، ۱۲۴

مزدیسنا ۱۴۸

مصر ۱۰۶

مغولان / مغولی ۲۹، ۱۰۶

مقدونیان ۱۰۶

مولوی بلخی ۱۳، ۲۵، ۲۶

میخی ۱۵، ۱۶، ۱۰۳، ۱۰۴، ۱۰۵، ۱۰۶، ۱۰۷، ۱۱۱، ۱۱۴،

۱۱۵، ۱۱۶

ن

نبطی ۶۲، ۶۴

نرد ۱۷، ۱۵۳، ۱۵۴، ۱۵۵، ۱۵۶، ۱۵۷، ۱۶۰، ۱۶۷

نردشیر ۱۵۴، ۱۵۵

نروژی ۶۲

نسخ ۶۴

نیواردشیر ۱۵۵، ۱۶۸

# Persec Code: A New Error Correction Code

## Abstract

Designed procedures simplify fault diagnosis or detection, so errors can be automatically detected and/or corrected by use of coded inputs. In general, codes are commonly classified in terms of their ability to detect or correct classes of errors that affect some fixed number of bits in a word. Many codes have been developed that can be used in the design of self-checking circuits. Type of codes may vary depending on the type of circuits. For data-transmission busses, a parity-check code may be adequate, for other types of functions, however, we may wish to use a code by which the check bits of the result can be determined from the check bits of the operands.

In this study, a new error detection and correction code (ED/CC), called "Persec code" is developed, which is mathematically proved to be better in comparison with other candidates and also adaptive to environmental changes. Theoretically, this code is able to detect several errors, and correct more than one error of data-packet as well. This paper successfully demonstrates 1-error correcting scenario, via simulation and validation processes and also encoding, decoding, and detection hardware circuits.

## keywords

- Error detection and correction codes
- Data transmission
- Testability
- Fault Tolerance
- Achaemenian immortal army.

# Backgammon Results Prediction using Machine Learning Methods

## Abstract

Backgammon is an ancient Iranian board game which is still popular in the world and is played widely on the internet. This paper discusses a model for predicting final results of backgammon games by using Naïve Bayesian classifier and artificial neural network, according to current states in the middle of the match or final statistics. By the extracted features, Bayesian classifier could predict actual results from current states with 82% in average and it achieved up to 97% of accuracy by final statistics; while neural network reached to 86% and 96% subsequently. Neural network doesn't have significant difference with Bayesian classifier performance.

## keywords

- Artificial Intelligence
- Naïve Bayesian classifier
- Artificial Neural Network
- Nard
- Nardshir
- Iranian board games.

# **Intelligent Production of New Beautiful and Meaningful Iranian Names using Genetic Algorithm with Neural Network-Based Fitness Function**

## **Abstract**

Spoken languages around the world have wended an evolutionary flow, during the past thousands of years. On such a way, words and also common names in different cultures and languages have evolved and will continue their way. Some of these names have more popularity and spread out the society, according to the aesthetic sense of individuals and their phonetic features which are variable depending on geographical places and time.

In this paper we propose a model which can produce new Iranian names having meaning and acceptable degree of beautiness in comparison with human opinion. The procedure of name production is based on evolutionary processing (genetic algorithm) and the fitness function is simulated by a feed-forward back-propagation neural network (except for the first generation) with two hidden layer which can examine human opinion up to 89%. The best generated name gains rank 5 among 252 names of initial population. Training and test data, the best samples of initial population, examination of fitness function (for first generation) and also determination of beautiness degree of the best final generated names have done by using a statistical analysis on questionnaires data.

## **keywords**

- The Encyclopedia of Iranian Names
- Artificial Intelligence
- Etymology
- Namology
- Multi-layered Perceptron
- Natural language processing.

# Optical Old-Persian Cuneiform Recognition using Spatial Graphs

## Abstract

Cuneiforms are one of the oldest scripts created by human to conveying concepts. Because of the need for expertise and difficulty of reading and understanding of cuneiform texts, thousands of ancient tablets are still unrecognized, nevertheless by recent archeological explorations, its number is increasing.

This paper proposes an optical character recognition project on Old Persian (Achaemenian) cuneiform by use of spatial graphs and nearest neighbor matrix. Results shows this method has achieved 98% of accuracy on machine-typed text reading.

## keywords

- Artificial Intelligence
- Optical character reading
- Achaemenian cuneiform
- Iranian literary heritage
- Intelligent Production of

# Automatic Farsi Typo Correction using Hybrid Neural Network

## Abstract

Automatic correction of typos in the computer-typed texts is a goal of researchers in artificial intelligence, data mining and natural language processing. Most of the existing methods are based on searching in dictionaries and determining the similarity of the dictionary entries and the given word.

This paper presents the design, implementation, and evaluation of a Farsi typo correction system using the Hopfield and Multi-Layer Perceptron (MLP) neural networks. The results show that for learning a dictionary from 4 up to 256 words having 4 to 6 characters, the correction accuracy of the Hopfield network is 55% to 100% and for the multilayer perceptron 80% to 100%. The hybrid network achieved a correction accuracy from 80% to 100% for over 3000 words.

## keywords

- Artificial intelligence
- Persian words
- Hopfield neural network
- Multi-layer Perceptron,
- Hybrid neural network
- Natural language processing



# Persian-Arabic Etymology using Artificial Neural Networks

## Abstract

Etymology and lexicology have a deep influence in many sciences ranging from humanities (Linguistics, History, archeology, etc.) to text classification (translation, language identification, author identification, etc.) and also speech processing (correct spelling of the words).

This paper studies a new application of artificial neural networks and text categorization to etymologize and find historical roots of Persian and Arabic words existing in Persian language. An MLP network was designed and used to solve the problem, achieving a correct recognition rate of over 92%.

## keywords

- Artificial intelligence
- Lexicology
- Farsi
- Text classification
- Text categorization
- Language identification
- Multi-layer Perceptron.

# Poet Identification using Machine Learning Methods

## Abstract

This paper discusses automated and intelligent identification of four great Iranian poets (Ferdowsi, Hafez, Khayyam, and Mawlavi) by use of two machine learning methods: decision tree and artificial neural networks. For identification of network performance, precision of human decision and identification has been examined and the results have been compared with machine learning methods efficiencies.

In phase of feature extraction examination by machine, the way of human thinking has been tried for simulation. Above 50 varies features are extracted from Persian poets and categorized in three physical, conceptual and vocal classes.

Results shows decision tree method has achieved 94% of correctness while classification and artificial neural network 95.9% in average, however human identification precision, according to our statistics, was only about 34.4%.

## keywords

- Artificial intelligence
- Artificial neural networks
- Decision tree
- Shahnameh of Ferdowsi
- Divan of Hafez
- Rubáiyát of Omar Khayyám
- Masnavi of Mawlavi (Roomi).
- Iranian literature.

# Poet Identification for Shahnameh of Ferdowsi using Artificial Neural Network

## Abstract

"Poet identification" is classified as "author identification", an important subject in "text classification" and "natural language processing". This paper discusses intelligent recognition of Ferdowsi's Shahnameh by use of feed-forward back-propagation (FF/BP) artificial neural network (ANN). Decision tree system (DTS) results has been used to comparison. In poetic feature extraction phase, human thinking about poem understanding and recognition is simulated. More than 64 data inputs from 22 various features is extracted which categorize physical, conceptual, and rhythmic classes. In this manner, specified database of Persian words has been designed and implemented, for quantitative measuring and examination of samples. Afterward, to make an automatic feature reduction and determining usefulness of every poetic feature, single Perceptron network was used.

After the training and testing phases, it was distinguished that DTS method achieves more than 85% of accuracy on classifying poems into the Shahnameh and non-Shahnameh classes, whereas ANN accomplished 100%. Combination of this network with above methods keeps accuracy of classifiers, while decreases loadings of FF/BP-ANN.

## keywords

- Artificial intelligence
- Text categorization
- Natural language Processing
- Author Identification,
- Persian poetry
- Iranian literature.

**Copyright © 2017 by Kanoon Nashre Oloom Publications.**

Shahmiri, Amir Shahab, 1975-

Iranian Culture and Language Processing using Artificial Intelligence / Amir Shahab Shahmiri.

ISBN 978-9-64-327175-6

Printed in Iran.

# **Iranian Culture and Language Processing using Artificial Intelligence**

**Amir Shahab Shahmiri**



*Kanoon Nashre Oloom Publications*