دانشگاه فردوسی مشهد

مهندسی کامپیوتر

تمرین سری ۳ – شبکه‌های عصبی
تعلیم مدل شبکه عصبی برای تصمیم گیری regression
بر پایه اطلاعات هیپوکامپ

متین نظام‌الملکی

۴۰۱۱۲۶۲۳۵۰

مقایسه نهایی در میان هر سه سری آزمایش

فهرست مطالب

۱ مقدمه و بیان مسئله ۲
۱.۱ توضیحات دادگان ۲

۲ معماری مدل ۳

۳ سری ۱: مدل مستقل برای هر موش ۴
۱.۳ خلاصه ۴
۲.۳ نتایج ۴
۳.۳ تحلیل ۵

۴ سری ۲: RNN مشترک با رمزگذارهای اختصاصی برای هر موش ۶
۱.۴ خلاصه ۶
۲.۴ نتایج ۶
۳.۴ تحلیل ۶

۵ سری ۳: یادگیری انتقالی (Transfer Learning) ۷
۱.۵ چکیده ۷
۲.۵ نتایج ۷
۳.۵ تحلیل ۷

۶ نتیجه‌گیری ۹

۱. مقدمه و بیان مسئله

این گزارش به ارائه طراحی، پیاده‌سازی و ارزیابی مدل‌های شبکه عصبی بازگشتی برای رمزگشایی موقعیت مکانی یک موش صحرایی (rat) روی یک مسیر خطی با استفاده از داده‌های spike هیپوکامپ می‌پردازد. هیپوکامپ ناحیه‌ای از مغز است که به دلیل کدگذاری اطلاعات فضایی از طریق فعالیت سلول‌های مکان (place cells) شناخته شده است: نورون‌هایی که زمانی که حیوان در مکان خاصی قرار می‌گیرد، ترجیحاً شلیک می‌کنند. با استفاده از داده‌های ثبت‌شده از چهار موش (Achilles, Buddy, Cicero Gatsby) به حل مسئله رگرسیون برای پیش‌بینی موقعیت موش (یک مقدار اسکالر در بازه $[0, 1.6]$ متر) در هر گام زمانی از روی بردار جمعیت فعالیت عصبی می‌پردازیم.

۱.۱. توضیحات دادگان

داده‌های مربوط به هر موش با دو آرایه ذخیره شده است:

- neural: با ابعاد $[K \times M]$، که در آن K تعداد کل گام‌های زمانی و M تعداد نورون‌های ثبت‌شده همزمان است.
- behavioral: با ابعاد $[K \times 1]$، موقعیت موش در هر گام زمانی.

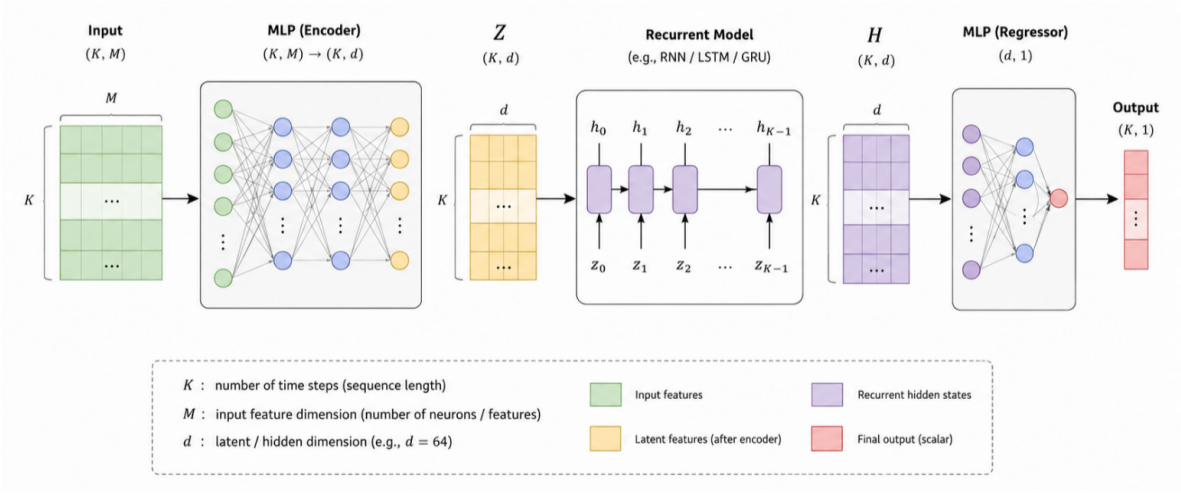
جدول ۱: آمار دادگان برای هر موش

موش	تعداد نورون‌ها (M)	گام‌های زمانی (K)	تعداد آزمایش‌ها
Achilles	۱۲۰	۱۰،۱۷۸	۸۵
Buddy	۴۸	۶،۵۷۷	۵۸
Cicero	۵۵	۴۷،۴۳۱	۹۳
Gatsby	۶۶	۱۷،۰۲۶	۸۵

داده‌های پیوسته به دوره‌های آزمایش (trials) تقسیم شده‌اند: هر آزمایش متناظر با یک بار طی کردن مسیر است (از چپ به راست یا راست به چپ). بخش‌بندی با تشخیص کمینه‌ها و بیشینه‌های محلی موقعیت در نزدیکی مرزهای مسیر (کمتر از 0.025 متر یا بیشتر از 1.575 متر) در یک پنجره لغزان با ± 20 گام زمانی انجام می‌شود و فقط گذرهایی که در آن‌ها $|\Delta y| \geq 1$ متر است، نگه داشته می‌شوند. آزمایش‌هایی با طول بیشتر از 500 گام زمانی برش می‌خورند تا زمان آموزش کاهش یابد (این موضوع عمدتاً بر Cicero تأثیر می‌گذارد که طولانی‌ترین آزمایش آن ۵۵۱۴ گام است). دادگان با نسبت ۸۰٪ / ۲۰٪ و با استفاده از یک سید تصادفی ثابت (42) به مجموعه‌های آموزش و اعتبارسنجی تقسیم شده‌اند.

۲. معماری مدل

همه مدل‌ها از یک پایپ‌لاین سه‌بلوکی مشترک استفاده می‌کنند:



شکل ۱: ساختار کلی معماری مدل های سری اول آزمایش

۳. سری ۱: مدل مستقل برای هر موش

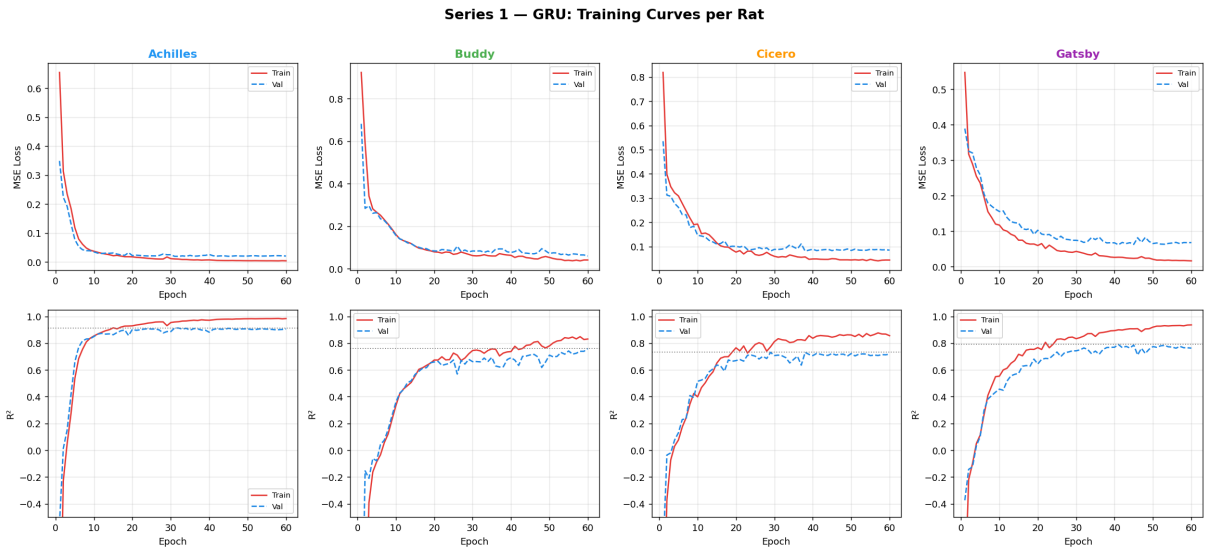
۱.۳. خلاصه

در این سری، یک RatModel مجزا برای هر یک از چهار موش و دو معماری (GRU, RNN) آموزش داده می‌شود که در مجموع منجر به ۸ مدل می‌گردد. هر مدل به‌طور مستقل روی داده‌های موش مربوطه آموزش می‌بینند و هیچ اطلاعاتی بین حیوانات به اشتراک گذاشته نمی‌شود. معماری مدل از شکل ۱ پیروی می‌کند که شامل یک لایه encoder اختصاصی برای موش ($M_{\text{rat}} \to 64$)، یک لایه بازگشتی ($64 \to 128$) و یک رگرسور مشترک ($128 \to 1$) است.

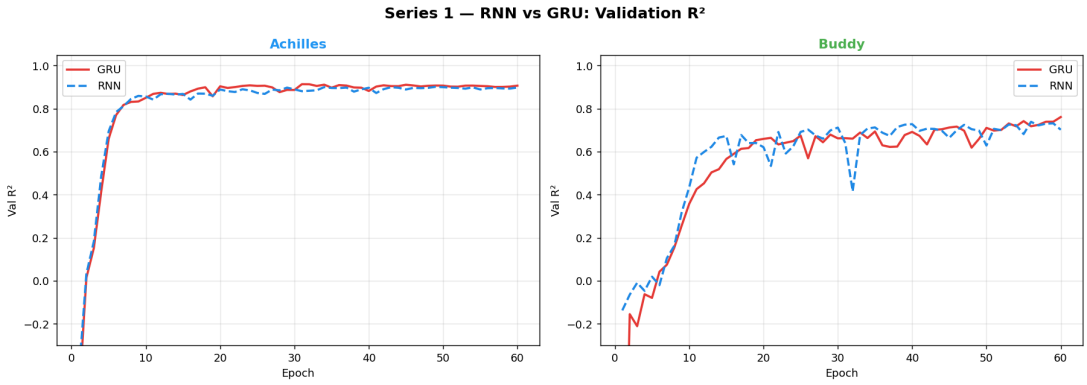
۲.۳. نتایج

موش	معماری	بهترین R^2 اعتبارسنجی	R^2 نهایی آموزش	Δ(GRU−RNN)
Achilles	GRU	**0.9139**	0.9843	+0.0122
	RNN	0.9018	0.9833	
Buddy	GRU	**0.7613**	0.8316	+0.0221
	RNN	0.7392	0.7984	
Cicero	GRU	**0.7317**	0.8566	+0.0336
	RNN	0.6981	0.8400	
Gatsby	GRU	**0.7907**	0.9371	+0.0026
	RNN	0.7881	0.8761	

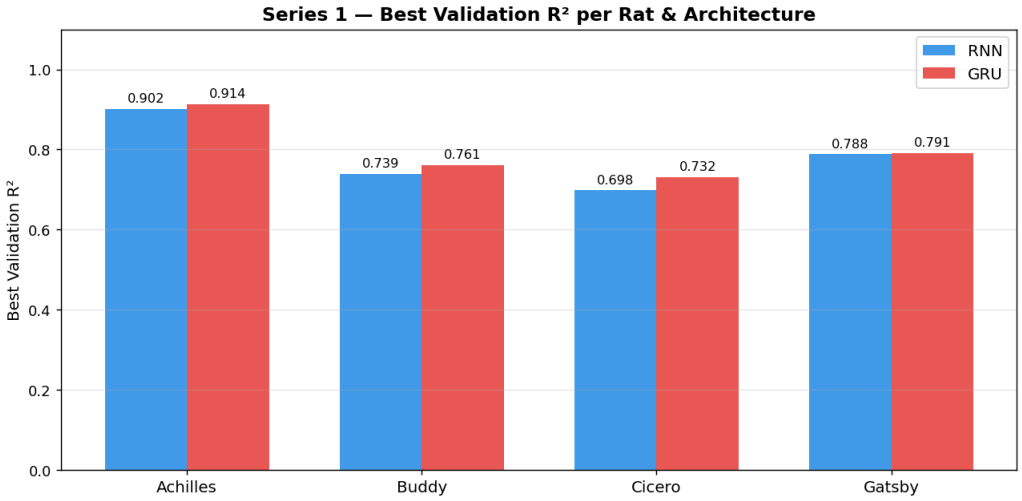
جدول ۲: نتایج سری ۱: بهترین R^2 اعتبارسنجی و R^2 نهایی آموزش



شکل ۲: سری ۱ — منحنی‌های آموزش GRU. ردیف بالا: خطای MSE (آموزش/اعتبارسنجی). ردیف پایین: R^2 (آموزش/اعتبارسنجی). خط‌چین عمودی دوره مربوط به بهترین R^2 اعتبارسنجی را نشان می‌دهد.



شکل ۳: سری ۱ – مقایسه R^2 اعتبارسنجی بین GRU و RNN برای Achilles (چپ) و Buddy (راست).



شکل ۴: سری ۱ – بهترین R^2 اعتبارسنجی به تفکیک موش و معماری.

۳.۳. تحلیل

برتری مداوم GRU بر RNN. در تمامی چهار موش، GRU به بهترین R^2 اعتبارسنجی بالاتری نسبت به RNN ساده دست می‌یابد (جدول ۲). بیشترین میزان بهبود مربوط به موش Cicero است ($+0.034$)، زیرا آزمایش‌های طولانی‌تر آن فرصت بیشتری برای وابستگی‌های زمانی فراهم می‌آورد که مکانیزم گیت‌ها می‌تواند از آن بهره‌برداری کند. برای Gatsby این اختلاف ناچیز است ($+0.003$)، احتمالاً به این دلیل که آزمایش‌های این موش دارای طول متوسطی هستند و سیگنال زمانی به‌خوبی توسط یک RNN ساده قابل مدل‌سازی است.

سادگی بیشتر رمزگشایی برای Achilles. موش Achilles دارای بیشترین تعداد نورون ($M = 120$) است و به بهترین R^2 اعتبارسنجی دست می‌یابد($\text{GRU} = 0.914$). ثبت همزمان جمعیت بزرگتری از نورون‌های هیپوکامپ، اطلاعات فضایی غنی‌تری فراهم کرده و دشواری مسئله رگرسیون را کاهش می‌دهد.

شکاف بزرگتر بین آموزش و اعتبارسنجی در Cicero. مقدار R^2 نهایی آموزش برای مدل GRU موش Cicero برابر با 0.857 است در حالی که در اعتبارسنجی در مقدار 0.732 به اوج می‌رسد. این شکاف بزرگتر در مقایسه با Achilles (0.984 در برابر 0.914) نشان‌دهنده یک Overfitting خفیف است، احتمالاً به این دلیل که Cicero آزمایش‌های پرت (از نظر طول) زیادی دارد و برش داده‌ها روی ۵۰۰ گام زمانی باعث از بین رفتن اطلاعات می‌شود.

دینامیک یادگیری (شکل ۲). هر چهار موش تقریباً در عرض ۲۰ تا ۳۰ دوره به همگرایی می‌رسند. Achilles سریع‌ترین و هموارترین همگرایی را دارد که نشان‌دهنده فراوانی نورون‌های حاوی اطلاعات مفید است.

۴. سری ۲: RNN مشترک با رمزگذارهای اختصاصی برای هر موش

۱.۴. خلاصه

در این سری، یک SharedRNNModel به‌طور همزمان روی هر چهار موش آموزش داده می‌شود. معماری مدل شامل موارد زیر است:

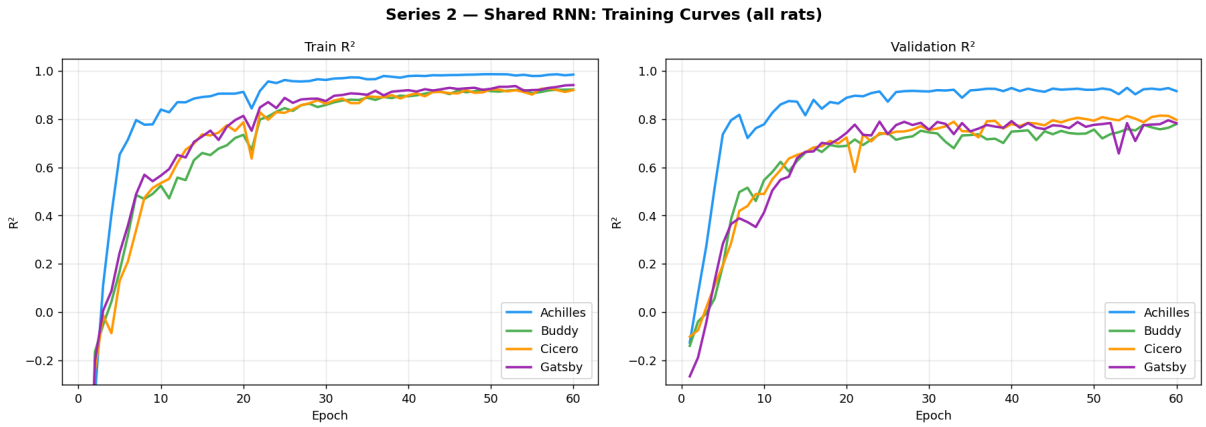
- چهار encoder مستقل، یکی برای هر موش: $\text{Encoder}_r : \mathbb{R}^{M_r} \to \mathbb{R}^{64}$.
- یک GRU مشترک $(64 \to 128)$، که با استفاده از گرادیان‌های به‌دست‌آمده از هر چهار موش به‌روزرسانی می‌شود.
- یک رگرسور مشترک $(128 \to 1)$.

در هر گام آموزشی، یک مینی‌بچ از هر موش پردازش می‌شود؛ هر کدام از طریق encoder مختص به خود، سپس از طریق GRU و رگرسور مشترک. زیان کل برابر با مجموع خطاهای MSE ماسک‌دار روی همه موش‌هاست و یک `Adam.step()` مشترک، همه پارامترها را به‌روزرسانی می‌کند.

۲.۴. نتایج

موش	R^2 اعتبارسنجی سری ۲	R^2 آموزش سری ۲	R^2 اعتبارسنجی سری ۱ (GRU)	Δ(سری ۲ − سری ۱)
Achilles	**0.9292**	0.9839	0.9139	**+0.0153**
Buddy	**0.7794**	0.9221	0.7613	**+0.0181**
Cicero	**0.8139**	0.9198	0.7317	**+0.0822**
Gatsby	**0.7949**	0.9406	0.7907	**+0.0042**

جدول ۳: نتایج سری ۲ و مقایسه با سری ۱ (GRU)



شکل ۵: سری ۲ — مقدار R^2 آموزش (چپ) و اعتبارسنجی (راست) برای هر چهار موش در طول آموزش مدل RNN مشترک.

۳.۴. تحلیل

آموزش مشترک عملکرد را بهبود می‌بخشد. مدل GRU مشترک نسبت به مدل پایه مستقل GRU (سری ۱) برای همه موش‌ها عملکرد بهتری نشان می‌دهد. این بهبود بیشترین مقدار را برای Cicero (+0.082) دارد: با داشتن ۹۳ آزمایش و متنوع‌ترین طول‌های آزمایش، Cicero بیشترین بهره را از داده‌های آموزشی اضافی که توسط موش‌های دیگر فراهم می‌شود می‌برد.

این نتایج از این فرضیه پشتیبانی می‌کند که مدل GRU مشترک، یک مدل زمانی عمومی از encoder ها یاد می‌گیرد، طوری که علی‌رغم تفاوت های هاencoder با یک دیگر و سازگاری encoder هر نمونه آزمایشی با هیپوکامپ(اطلاعات)، باز هم مدل عمومی می‌تواند دقت بیشتری از مدل های خاص سری ۱ داشته باشد

توجه کردن به این نکته نیز که مقادیر R^2 آموزش در سری ۲ به‌طور قابل‌ملاحظه‌ای بالاتر از سری ۱ است ،خالی از لطف نیست (برای: Buddy 0.922 در برابر 0.831؛ Cicero: 0.920 در برابر 0.857)، که نشان می‌دهد مدل مشترک ظرفیت بیشتری دارد

و بهتر روی داده‌های آموزشی برازش می‌شود. با این حال عملکرد اعتبارسنجی نیز بهبود می‌یابد که نشان می‌دهد پارامترهای اضافی باعث تعمیم بهتر شده و منجر به بیش‌برازش مخرب نمی‌شوند.

۵. سری ۳: یادگیری انتقالی (Transfer Learning)

۱.۵. چکیده

در این سری بررسی می‌شود که آیا دانش آموخته‌شده از سه موش مبدأ می‌تواند به یک موش جدید که قبلاً دیده نشده، منتقل شود یا خیر. روند کار به این صورت است:

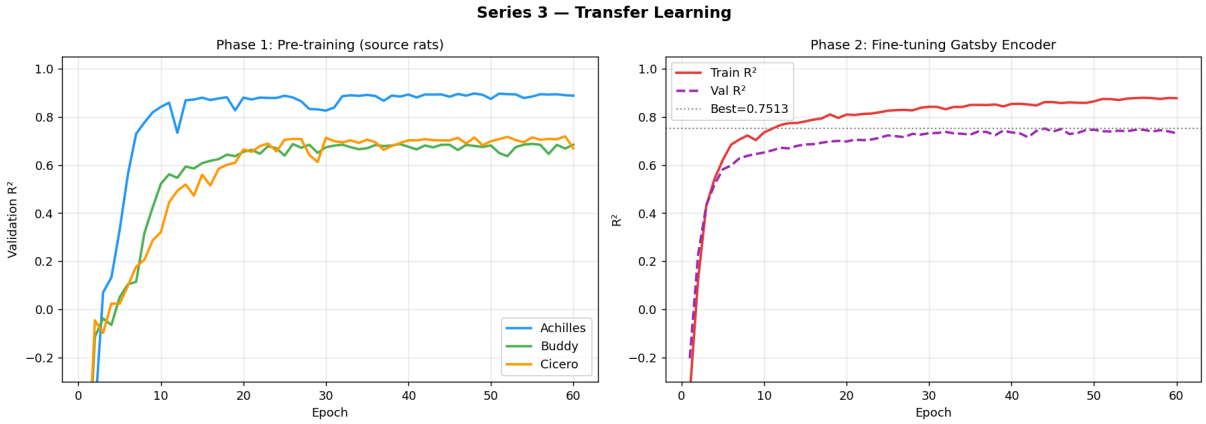
۱. فاز ۱ — پیش‌آموزش: یک SharedRNNModel فقط روی موش‌های Buddy Achilles، و Cicero (به مدت ۶۰ دوره) آموزش می‌بینند. این مدل دارای سه encoder اختصاصی و یک GRU + رگرسور مشترک است.
۲. فاز ۲ — تنظیم دقیق (Fine-tuning): تمام پارامترهای GRU مشترک و رگرسور فریز (ثابت) می‌شوند . یک رمزگذار جدید برای Gatsby متصل می‌شود . سپس تنها رمزگذار جدید روی داده‌های Gatsby به مدت ۶۰ دوره آموزش می‌بینند.

— تعداد پارامترهای قابل آموزش در فاز ۲ تنها ۴۶،۰۱۴ پارامتر (فقط رمزگذار جدید) از مجموع ۱۲۸،۸۳۱ پارامتر کل مدل است یعنی فقط ۳۵.۷ درصد از کل مدل به‌روزرسانی می‌شود.

۲.۵. نتایج

موش / فاز	بهترین R^2 اعتبارسنجی	R^2 نهایی آموزش
Achilles (پیش‌آموزش)	0.8966	0.9504
Buddy (پیش‌آموزش)	0.6875	0.7668
Cicero (پیش‌آموزش)	0.7187	0.8173
Gatsby (تنظیم دقیق)	**0.7513**	0.8769

جدول ۴: نتایج سری ۳-



شکل ۶: سری ۳ — سمت چپ: R^2 اعتبارسنجی موش‌های مبدأ در طول پیش‌آموزش. سمت راست: R^2 آموزش و اعتبارسنجی برای تنظیم دقیق Gatsby. خط‌چین بهترین R^2 اعتبارسنجی بدست آمده (0.7513) را نشان می‌دهد.

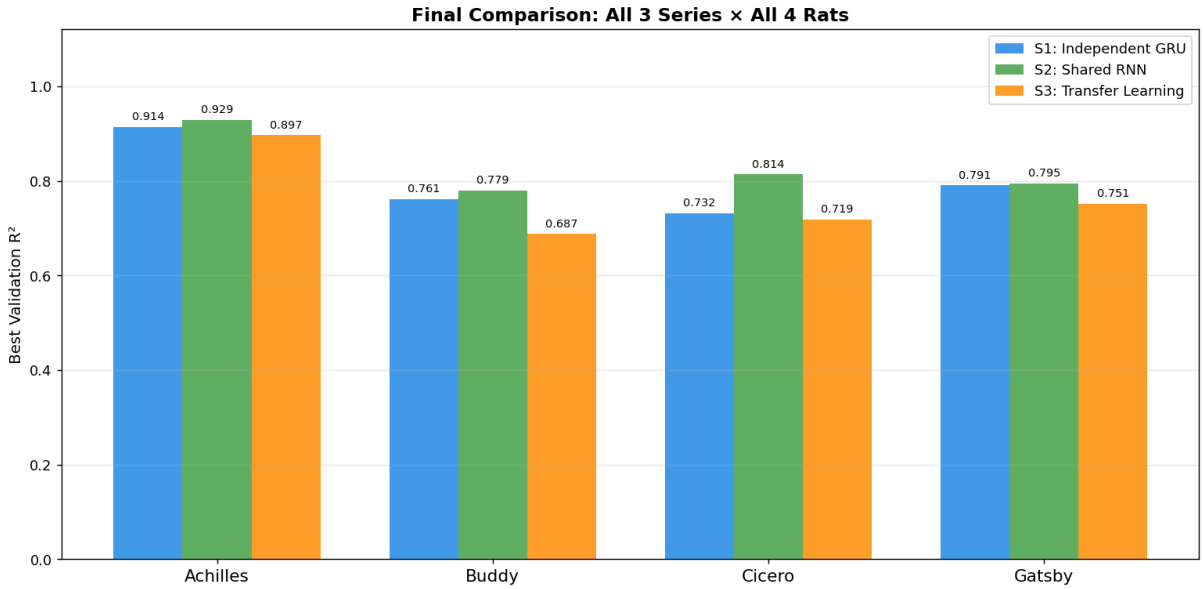
۳.۵. تحلیل

یادگیری انتقالی مؤثر است. با وجود آموزش تنها یک encoder کوچک روی داده‌های Gatsby و ثابت نگه داشتن کامل ساختار شبکه، مدل تنظیم دقیق‌شده به مقدار $R^2 = 0.751$ روی مجموعه اعتبارسنجی Gatsby دست می‌یابد. این مقدار بسیار بالاتر از شانس بوده و نشان می‌دهد که دینامیک‌های زمانی یادگرفته‌شده از سه موش دیگر حاوی سوگیری‌های استنتاجی و درک الگو های مفیدی برای یک حیوان جدید است.

تفاوت عملکرد نسبت به سری ۲. مقدار R^2 تنظیم دقیق‌شده (0.751) کمتر از مدل GRU سری ۱ (0.791) و مدل سری ۲ (0.795) برای موش Gatsby است. علتی که حدس زده می‌شود این است که: ساختار شبکه فریز شده هرگز با داده‌های Gatsby به‌روزرسانی نشده است، بنابراین دینامیک‌های پنهان کاملاً با کدهای عصبی این موش تطابق ندارند. در مقابل، سری ۲ تمامی پارامترها از جمله GRU مشترک را با استفاده از داده‌های Gatsby به‌روزرسانی می‌کند.

عملکرد موش‌های مبدأ در فاز ۱. مدل پیش‌آموزش دیده در مقایسه با سری ۲ (0.779 و 0.814) به ترتیب دارای R^2 اعتبارسنجی کمتری روی موش‌های Buddy (0.688) و Cicero (0.719) است، زیرا بدون داده‌های Gatsby سیگنال آموزشی کمتری دریافت کرده و تنوع پایین‌تری دارد.که این موضوع نیز منطقی و قابل پیشبینی می‌باشد.

۶. نتیجه‌گیری



شکل ۷: مقایسه نهایی بهترین مقادیر R^2 اعتبارسنجی در میان هر سه سری آزمایش و برای هر چهار موش. آبی: مدل مستقل GRU در سری ۱؛ سبز: شبکه RNN مشترک در سری ۲؛ نارنجی: یادگیری انتقالی در سری ۳.

موش	RNN سری ۱	GRU سری ۱	مدل مشترک سری ۲	انتقالی سری ۳
Achilles	0.9018	0.9139	**0.9292**	0.8966
Buddy	0.7392	0.7613	**0.7794**	0.6875
Cicero	0.6981	0.7317	**0.8139**	0.7187
Gatsby	0.7881	0.7907	**0.7949**	0.7513
میانگین	0.7818	0.7994	**0.8294**	0.7635

جدول ۵: خلاصه بهترین مقادیر R^2 اعتبارسنجی در تمام سری‌ها

این پژوهش نشان داد که شبکه‌های عصبی بازگشتی قادرند با دقت بالایی موقعیت مکانی را از روی فعالیت جمعی هیپوکامپ رمزگشایی کنند. نتیجه‌گیری های کلیدی این پژوهش به شرح زیر می‌باشد:

- معماری GRU، به لطف داشتن مکانیزم گیت‌ها برای مدیریت وابستگی‌های زمانی بلندمدت، پیوسته بهتر از RNN ساده عمل می‌کند.
- آموزش مشترک روی چندین موش با استفاده از رمزگذارهای خاص هر حیوان و یک بستر شبکه بازگشتی مشترک، مؤثرترین استراتژی است: این کار باعث دستیابی به یک میانگین R^2 اعتبارسنجی به میزان 0.829 می‌شود که نسبت به مدل‌های مستقل تا ۲.۸ درصد بهبود نشان می‌دهد.
- یادگیری انتقالی — ثابت نگه داشتن بستر مشترک و آموزش تنها یک encoder جدید برای موش دیده‌نشده — با موفقیت دینامیک‌های زمانی را با یک هزینه پارامتری ۷.۳۵ درصدی منتقل می‌کند و موفق به کسب $R^2 = 0.751$ برای Gatsby بدون نیاز به آموزش مجدد بخش‌های بازگشتی می‌شود.
- برتری GRU بر RNN. در همه موش‌ها، معماری GRU عملکرد بهتری از RNN ساده از خود نشان داد. میزان این مزیت با پیچیدگی ساختار زمانی داده‌های هر موش متناسب است: بیشترین مقدار برای Cicero ($+0.034$) و کمترین مقدار

برای Gatsby ($+0.003$). این مسئله تأیید می‌کند که بازگشت‌های دارای گیت (gated) انطباق بهتری با ماهیت طول متغیر و ساختار زمانی دنباله سلول‌های مکان هیپوکامپ دارند.

- آموزش مشترک، تعمیم را بهبود می‌بخشد. آموزش مشترک همزمان روی چند موش همواره از آموزش روی یک موش منفرد پیشی می‌گیرد. میانگین بهبود نسبت به مدل GRU در سری ۱ برابر با $+0.030$ در معیار R^2 است. موش Cicero بیشترین سود را برده است ($+0.082$)، چرا که طول متغیر آزمایش‌های آن باعث می‌شود رمزگشایی مستقل آن دشوارتر باشد؛ ساختار مشترک به نحو مؤثری داده‌های آموزشی این موش را با اطلاعات استخراج شده از حیوانات دیگر تقویت می‌کند.

- یادگیری انتقالی میتواند مؤثر اما غیربهینه باشد . آموزش تنها یک encoder ، در این مثال یعنی ۳۵.۷ درصد از پارامتر ها روی یک موش جدید برای موش Gatsby منجر به کسب $R^2 = 0.751$ شد – نتیجه‌ای رقابتی با RNN سری ۱ (0.788) که با پارامترهای بسیار کمتر و بدون نیاز به آموزش مجدد مدل زمانی حاصل شده است. این امر طراحی ماژولار را اعتبارسنجی می‌کند: ساختار بازگشتی دینامیک ناوبری کلی را در خود نگه می‌دارد و یک رمزگذار سبک وزن کافی است تا شکاف بین این ساختار را با داده‌های عصبی یک موش جدید پر کند. با این حال برای رسیدن به بالاترین سطح عملکرد، در صورت وجود داده از موش جدید، آموزش مشترک (سری ۲) ترجیح داده می‌شود.اینطور میتوان استدلال نمود که استفاده از یادگیری انتقالی، در مواردی که یک شبکه از پیش تعلیم دیده شده داریم و میخواهیم برای یک خروجی خاص مورد بررسی قرار دهیم، مورد استفاده قرار خواهد گرفت.پیشبینی می‌شود که در اکثر مواقع تعیلم مدل با همه نمونه هایی که داریم میتواند بهینه ظاهر گردد.

- حجم داده‌ها اهمیت دارد. موش Achilles بالاترین R^2 را در تمام تنظیمات بدست آورد، که دلیل عمده آن داشتن بیشترین تعداد نورون ثبت شده به صورت همزمان است (120). از طرفی Buddy که فقط ۴۸ نورون و ۵۸ آزمایش دارد همواره پایین‌ترین R^2 را ثبت کرد. این مسئله اهمیت پوشش الکترودی و مدت زمان ثبت در کیفیت رمزگشایی عصبی را نشان می‌دهد.