



Közzététel: 2022. október 28.

A tanulmány címe:

Magyar halandósági ráták előrejelzése visszacsatolt neurális hálózatokkal

Szerzők:

SZENTKERESZTI GÁBOR,

az UNION Biztosító Zrt. aktuáriusa és a Budapesti Corvinus Egyetem PhD-hallgatója

E-mail: gabor.szentkereszti@uni-corvinus

DR. VÉKÁS PÉTER,

a Budapesti Corvinus Egyetem egyetemi docense

E-mail: peter.vekas@uni-corvinus.hu

DOI: <https://doi.org/10.20311/stat2022.10.hu0905>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) *Statisztikai Szemle* c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Sztj.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Sztj. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:
„*Forrás: Statisztikai Szemle* c. folyóirat 100. évfolyam 10. számában megjelent, **Szentkereszti Gábor – Vékás Péter** által írt, **Magyar halandósági ráták előrejelzése visszacsatolt neurális hálózatokkal** című tanulmány (link csatolása)”
7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem feltétlenül esnek egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Szentkereszti Gábor – Vékás Péter

Magyar halandósági ráták előrejelzése visszacsatolt neurális hálózatokkal

Forecasting Hungarian mortality rates with Recurrent Neural Networks

Szentkereszti Gábor, az UNION Biztosító Zrt. aktuáriusa és a Budapesti Corvinus Egyetem PhD-hallgatója

E-mail: gabor.szentkereszti@uni-corvinus

Dr. Vékás Péter, a Budapesti Corvinus Egyetem egyetemi docense

E-mail: peter.vekas@uni-corvinus.hu

A Lee–Carter-modell régóta a legnépszerűbb halandóság-előrejelző modell, azonban az utóbbi években az aktuáriustudomány területén is egyre gyakrabban készülnek előrejelzések gépi tanulási módszerekkel. Tanulmányunkban visszacsatolt neurális hálózatokkal modellezzük a magyar férfi- és női halandósági adatokat az 1950-es évektől 2020-ig, 0 és 99 éves életkorok között. A neurális hálózatokkal kapott előrejelzéseket összevetjük a klasszikus Lee–Carter-modell előrejelzéseivel, majd elemezzük, hogy az egyes életkorok és naptári évek mentén melyik modell teljesít jobban. Eredményeink azt mutatják, hogy jelentősen javíthatók a Lee–Carter-modell előrejelzései neurális hálózatok felhasználásával. A pontosabb előrejelzéseket alkalmazva újra lehet gondolni egyes aktuáriusi és nyugdíjszámításokat.

Tárgyszavak: halandósági ráták, gépi tanulás, neurális hálózat

The Lee–Carter model is one of the most popular models for mortality forecasting, but machine learning solutions are increasingly popular in actuarial science. In recent years, authors have tried to forecast mortality rates with different types of neural networks. This paper introduces these neural network approaches on Hungarian mortality rates between age 0 and 99 from 1950 to 2020. We use the classical Lee–Carter model and LSTM neural networks for mortality modeling. Our paper analyses the residuals by age group and calendar year on the test set (from 2005 to 2020) and compares the results of the Lee–Carter and LSTM models. Our results show that LSTM networks give more accurate forecasts on Hungarian mortality data. With these forecasts, one can reconsider the expectations about the pension system or improve the accuracy of actuarial calculations.

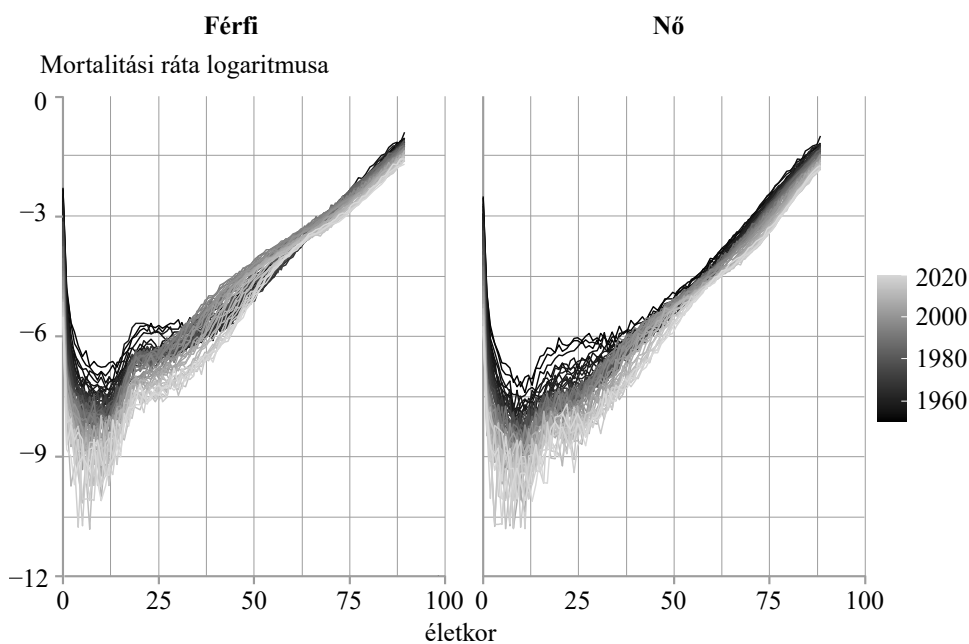
Keywords: mortality rates, machine learning, neural network

Az utóbbi évtizedekben a legtöbb országban folyamatosan javulnak a halandósági mutatók. Így van ez Magyarország esetében is. Nagyon izgalmas kérdés, hogy a trend miképpen folytatódhat. Így aztán nem csoda, hogy a mortalitási ráták elemzése, modellezése és előrejelzése központi témává vált az aktuáriusok és a demográfusok körében. Azonkívül, hogy halandó mivoltunk az egyént mélyen foglalkoztató kérdés, az életbiztosításokat kínáló vállalatok számára is fontos bizonyos populációk halandóságának az alakulása, illetve az állami folyó finanszírozású nyugdíjrendszerben is központi szerepet játszik a népesség számának és életkor szerinti összetételének jövőbeli alakulása.

1. ábra

A magyar mortalitási ráta logaritmusának nemek szerinti bontásában

Logarithms of Hungarian mortality rates by sex



Az életbiztosítások többnyire hosszú távú szerződések, nem ritkán több évtizedre szólnak, ezért a díjkalkuláció és a pénzáram-modellezés szempontjából kiemelt jelentőségű a megfelelő halandósági feltételezések kialakítása. Az életbiztosítási számításokat sok esetben egy ún. „periódus” halandósági tábla használ-

latával végezzük, tehát feltételezzük, hogy az adott korokhoz tartozó halálozási valószínűségek a következő években változatlanok maradnak (*Vékás, 2017*). A valóság ezzel szemben azt mutatja, hogy az egyes korokhoz tartozó halálozási valószínűségek a naptári évek előrehaladtával folyamatosan javulnak. Ezt szemlélteti az 1. ábra.

Szembetűnő, hogy a középkorú népesség halandósági mutatóinak javuló tendenciája megtorpan egy átmeneti időszakban, a férfiaknál egyértelműen látszik, hogy romlik a korábbi évekhez képest, míg a nőknél stagnál. Az 1960–1990-es évek között Magyarországon az elhízás, az alkoholfogyasztás, a dohányzás és a kevés mozgás igen erőteljesen befolyásolta a középkorú népesség egészségi állapotát, és a halandósági mutatók átmeneti javulása megállt (*Óri–Spéder, 2020*).

A biztosítási díjkalkulációban statikus és dinamikus díjszámítást különböztünk meg. A statikus díjszámítás esetében rögzített halandósági táblával számolunk, sok esetben a legfrissebbel. A dinamikus díjszámításnál viszont előrejelzésekre van szükségünk a halálozási valószínűségekről, s ezeket használjuk a kalkulációhoz. A statikus megközelítés azoknál az életbiztosításoknál okozhat gondot, amelyeknél a biztosító számára kedvezőtlen, ha a biztosított sokáig él. Ebből a szempontból a legnagyobb kockázattal az elérési életbiztosítások és a különböző járadékstruktúrák rendelkeznek. Ezt a veszélyt a hazai szakirodalom élet-tartam-kockázatnak nevezi (*Májner–Kovács, 2011*).

A halandóság modellezése és előrejelzése mellett tanulmányunk másik nagy témaköre a gépi tanulási (*machine learning – ML*) eszközök aktuáriustudományi alkalmazása. A mesterséges intelligencia (*artificial intelligence – AI*) és a gépi tanulás, csakúgy, mint sok más területen, a biztosítóknál is központi témává vált az utóbbi években. Egy biztosítótársaság sokféle módon használhatja a mesterséges intelligencia által nyújtott megoldásokat. Tanulmányunk kifejezetten az aktuáriusi alkalmazásokat állítja a középpontba. *Richman (2018)* átfogó tanulmányt készített az AI és az ML aktuáriusi felhasználhatóságáról. Lehetőségként említi többek között a halandóság-előrejelzést, a kártartalékok becslését, a törlesztésmodellezést és a nem életbiztosítási díjkalkulációt. Mivel a biztosítóknál általában sok adat áll rendelkezésre, és a különböző gépi tanulási módszerek elsődleges alkalmazási feltétele egy megbízható és nagy méretű adathalmaz, az aktuáriusi területen ezek a megoldások széleskörűen alkalmazhatók.

Az előbb említett alkalmazási lehetőségek közül a továbbiakban a halandóság-előrejelzésre fogunk koncentrálni, bemutatjuk az eddigi kutatásokat, valamint ML-eszközzel modellezzük a magyar halandósági adatokat.

1. Szakirodalmi áttekintés

A legelterjedtebb halandóság-előrejelző modell a *Lee–Carter (1992)* szerzőpáros nevéhez kötődik. A központi mortalitási ráta logaritmusára javasolnak egy log-bilineáris modellt naptári évek és korcsoportok szerint. Az információt két életkortól és egy naptári évtől függő paramétervektorba tömörítik, majd a naptári évtől függő paraméterre egy ARIMA⁻¹ (autoregresszív integrált mozgóátlagolású) folyamatot illesztnek a jövőbeli mortalitási ráták előrejelzése céljából. Lee és Carter a paramétereket a legkisebb négyzetek módszerével becsülték, de mivel tényleges megfigyelhető magyarázó változók nincsenek a logaritmikus halandósági rátákat leíró egyenlet jobb oldalán, ezért az általános regressziós becslések nem működnek. Így szingulárisérték-felbontással kapják meg a paraméterek legkisebb négyzetes becslését.

Brouhns–Denuit–Vermunt (2002) említik, hogy a Lee–Carter- (LC-) modellben a hibatagok normális eloszlása és homoszkedaszticitása irreális feltevés, így a szerzők egy másik megközelítést javasolnak. Feltételezésük szerint az egyes naptári évekhez és korcsoportokhoz tartozó halálozások száma Poisson-eloszlást követ. A modell paramétereit maximum likelihood becsléssel határozzák meg. Ezt nevezzük Poisson-féle Lee–Carter-modellnek.

Richman–Wütrich (2019) visszacsatolt neurális hálózatokkal modellezik a halandóságot. Tanulmányuk középpontjában a hosszú rövid távú memória (*long short-term memory – LSTM*) és a kapuval rendelkező visszacsatolt egység (*gated recurrent unit – GRU*) áll. Kimeneti adatként – a Lee–Carter-modellhez hasonlóan – ők is a központi halandósági ráta logaritmusát alkalmazzák. Bemeneti adatnak pedig a kimeneti adat naptári évek szerinti késleltetett értékeit használják, valamint a szomszédos életkorok azonos naptári évhez tartozó adatait, így elérnek egy simítást is az életkorok szerint. Empirikusan elemzik külön-külön a svájci nőkre és férfiakra vonatkozó adatokat, majd egy közös, mindkét nemet tartalmazó modellt építenek. 1950 és 2016 közötti adatokat használnak, a 0 évestől 99 évesig terjedő életkorokra. Eredményeik azt mutatják, hogy az LSTM- és a GRU-modellek jobban teljesítenek a Lee–Carter-modellnél mindkét nem, illetve a teljes népesség esetében is. Kiemelik, hogy a neurális hálózatok nem stabilak (más és más eredményeket adnak) a különböző tanítások során, ezért ensemble módszerek használatát javasolják.

Richman–Wütrich (2021) a korábbi visszacsatolt neurális hálózatos megközelítéssel ellentétben egyszerű előrecsatolt neurális hálózatokat alkalmaznak multi-populációs halandósági modellezés céljára. Bemeneti adatnak a naptári évet, az

¹ Az ARIMA-modell részletes leírása megtalálható például *Asteriou–Hall (2015)* könyvében.

életkort, a régiót, illetve a nemet használják. A naptári évet leszámítva minden változót kategorikusként kezelnek, és beágyazóréteget (*embedding layer*) alkalmaznak rá. Az ehhez tartozó kimeneti adat a központi mortalitási ráta logaritmus-a az adott naptári évben és életkorban. A modellezéshez 41 ország férfi- és női halandósági adatait használják, és több neurális hálózatot építenek különböző hiperparaméter²-beállításokkal. A multipopulációs neurális hálózatos halandósági modelleket klasszikus modellekkel vetik össze. *Li és Lee (2005)* a Lee–Carter-modellt kiterjesztik több populációra, és ezeket „közös faktor” modellnek (*common factor model – CFM*), illetve „kiterjesztett közös faktor” modellnek (*augmented common factor model – aCFM*) nevezik. Ezenkívül *Kleinow (2015)* bevezeti a „közös korhatás” modellt (*common age effect – CAE*). *Richman–Wütrich (2021)* ezeket a modelleket használják a neurális hálózatokkal való összehasonlításban. Eredményeik azt mutatják, hogy a neurális hálózatok teljesítenek a legjobban a mintán kívüli előrejelzéseknél.

Nigri et al. (2019) más megközelítést alkalmaznak, mint *Richman és Wütrich*. Először felépítik a Lee–Carter-modellt, majd a naptári évtől függő paramétert modellezik LSTM-hálózatokkal, és ezekkel a neurális hálózatos modellekkel készítik el az előrejelzéseket. A neurális hálózatoknál a kimeneti adat az adott naptári évhez tartozó paraméter, a bemeneti adatok pedig ennek a paraméterértéknek a késleltetettjei. A szerzők kiemelik, hogy az LSTM az ARIMA-val ellentétben képes nemlineáris trendeket is megragadni, ezért megfelelőbb választás lehet a modellezésre. Empirikus vizsgálatokat végeznek hat ország halandósági adatai alapján, és eredményeik azt mutatják, hogy az LC–LSTM-modell kisebb hibájú előrejelzéseket ad, mint a tradicionális ARIMA-megközelítés.

Perla et al. (2021) szintén neurális hálózatokkal modellezik több ország mortalitási rátáit. Munkájukban a már említett LSTM-modellen túl konvolúciós neurális hálózatokat (*convolutional neural network – CNN*) is alkalmaznak. Fontos kiemelni, hogy az ország és a nem változókra nem dummy változókat készítenek, hanem beágyazóréteget alkalmaznak. CNN- és LSTM-modelljeik egyaránt számottevő eredményeket érnek el a mortalitási ráták előrejelzésében.

Schnürch–Korn (2022) szintén adnak egy CNN-hálózatos megközelítést multipopulációs mortalitásmodellezésre. Modelljüket Lee–Carter-, aCFM-, előrecsatolt neurális hálózatos és LSTM-modellekkel mérik össze. Eredményeik alapján az általuk létrehozott CNN-modell bizonyult a legjobbnak.

Richman–Wütrich (2021) és *Nigri et al. (2019)* is említik, hogy a neurális hálózatos megközelítések nagyon precíz pontbecsléseket tudnak adni, de a becslések varianciáját nehéz meghatározni. *Marino–Levantesi–Nigri (2021)* az

² Előzetesen rögzített, nem tanulás útján meghatározandó paraméterek.

LC–LSTM-modell varianciájára adnak becslést bootstrap módszerrel, amely megközelítést *Koissi et al. (2006)* alkalmazza az eredeti Lee–Carter-modellekre.

Deprez–Shevchenko–Wüthrich (2017) gépi tanulási módszereket alkalmaznak a halandósági modellezés fejlesztésére. Munkájukban döntési fákat használnak a klasszikus halandósági modellek visszatesztelésére. Ezáltal rámutatnak egyes modellek gyengeségeire, növelik az illeszkedés jóságát, és megteremtik a lehetőséget egyéb változók bevonására is (pl. iskolázottság, jövedelmi szint stb.). Erre a tanulmányra épít *Levantesi–Pizzorusso (2019)*, akik az előrejelzések terén is alkalmazzák a klasszikus modellek gépi tanulós kiterjesztését. Ők már nemcsak döntési fákat használnak, hanem véletlen erdőt (*random forest*) és *gradient boosting* módszereket is. *Levantesi–Nigri (2020)* folytatják ezt a modellezési elképzelést, de újítként P-spline függvényekkel extrapolálják a gépi tanulással fejlesztett klasszikus modell korrekciós paraméterét, s így pontosítják az előrejelzéseket.

A hazai irodalomban *Májer–Kovács (2011)* a Lee–Carter-modell felhasználásával elemzik az élettartam-kockázat hatását. A szerzők a 65 év feletti korosztályra fókuszálnak, és 1970-től 2006-ig használják az adatokat. *Vékás (2017)* egy lépéssel már továbbmegy, és az általánosított korcsoport-időszak-kohorsz (*generalized age-period-cohort – GAPC*) modellesalád egyes modelljeivel elemzi az élettartam-kockázat hatását.

Egy cikkben magyar szerzők (*Petneházi–Gáll, 2019*) is alkalmaznak LSTM-hálózatokat halandóságmodellezésre. Multipopulációs megközelítéssel dolgoznak, eredményeik igazodnak a bemutatott tanulmányokhoz, és cikkükben az LSTM-modell jobb teljesítményt nyújt a Lee–Carter-féle előrejelzésnél.

2. A Lee–Carter-modell

A Lee–Carter-modell a központi mortalitási rátát (jelölje m) modellezi és jelzi előre. Ez utóbbi kiszámítható az alábbi formulával,

$$m = \frac{D}{E},$$

ahol D az adott időszakban elhalálozottak száma, E pedig az adott időszak alatt élők átlagos száma. A mortalitási ráták számításáról bővebben *Ágoston–Kovács (2000)* írnak.

Mivel a későbbiekben a Poisson-féle Lee–Carter-modellt fogjuk használni, ezért ezt a modellt mutatjuk be, és eltekintünk az eredeti, 1992-es verzió tárgyalásától.

A Lee–Carter-modell az x korcsoporttól és t naptár évtől függő mortalitási rátákat jelzi előre, melyeket jelölje $m_{x,t}$ ($x = 1, 2, \dots, X$ és $t = 1, 2, \dots, T$). *Brouhns–Denuit–Vermunt (2002)* feltételezik, hogy $D_{x,t} \sim \text{Poisson}(E_{x,t} m_{x,t})$, ahol $m_{x,t} = \exp(a_x + b_x k_t)$, a logaritmus mortalitást tehát a következőképpen adják meg:

$$\ln(m_{x,t}) = a_x + b_x k_t$$

Identifikációs probléma miatt be kell vezetni a következő paramétermegkötéseket:

$$\sum_{t=1}^T k_t = 0 \text{ és } \sum_{x=1}^X b_x = 1$$

A paramétereket a szerzők az alábbi log-likelihood függvény maximalizálásával határozzák meg:

$$l(a_x, b_x, k_t) = \sum_{x,t} [D_{x,t}(a_x + b_x k_t) - E_{x,t} \exp(a_x + b_x k_t)] + \text{konstans}$$

A megoldáshoz numerikus módszereket alkalmaznak.

3. Neurális hálózatok

Ebben a fejezetben a neurális hálózatokat mutatjuk be röviden, azokra a hálózatokra, illetve a hálózatok azon elemeire fókuszálva, amelyeket majd alkalmazni fogunk a későbbi fejezetek során.

3.1. Előre-csatolt neurális hálózatok

A neurális hálózatok a bemeneti és a kimeneti adatok között egy nemlineáris leképezést valósítanak meg. A neurális hálózatok két fő alkotóeleme a neuron és a réteg. A neuronok matematikai műveleteket hajtanak végre, súlyozzák a bemeneti adatokat, összegzik őket, majd átfuttatják a kapott értéket egy nemlineáris függvényen, amelyet aktivációs függvénynek nevezünk. Az egyes neuronokat réte-

gekbe rendezzük, hogy megkapjuk a hálózat felépítését. Egy réteg lehet bemeneti, kimeneti vagy rejtett réteg. A neurális hálózat topológiája alatt a hálózat felépítését, struktúráját értjük. A neurális hálózatot gyakran irányított gráffal ábrázoljuk. A csomópontok a neuronokat, a gráf élei a neuronok közötti kapcsolatokat jelölik. Ha a hálózatot leíró gráf nem tartalmaz hurkot, akkor előreccsatolt neurális hálózatról beszélünk, ellenkező esetben visszacsatolt neurális hálózatról (*recurrent neural network – RNN*) (Altrichter et al., 2006).

Az egyik legkorábbi és legegyszerűbb neurális hálózatot Rosenblatt (1958) írta le, ez az úgynevezett Rosenblatt-perceptron. A hálózat egy bemeneti és egy kimeneti réteget tartalmaz, és egy klasszifikációs feladat megoldására képes. Formalizálva:

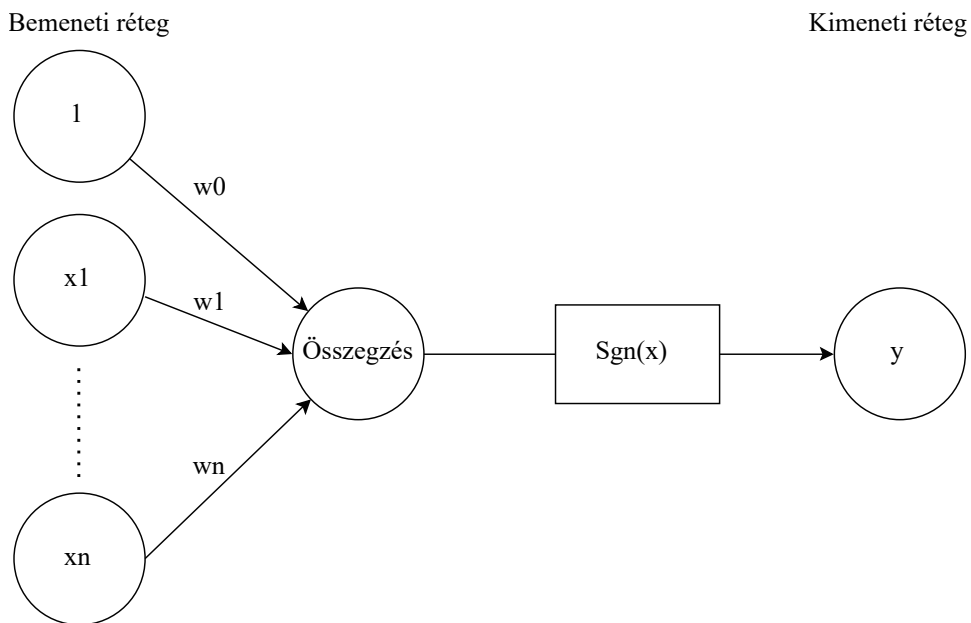
$$y = f(\mathbf{w}^T \mathbf{x}) = f\left(\sum_{i=0}^n w_i x_i\right),$$

ahol f az előjelfüggvény (sgn).

2. ábra

Rosenblatt-perceptron gráffal szemléltetve

Rosenblatt-perceptron on a graph



Forrás: saját szerkesztés.

Természetesen ez csak az alapeset, amikor egy bemeneti és egy kimeneti réteg alkotja a neurális hálózatot. A legnépszerűbb hálózatok az előre-csatolt többrétegű neurális hálózatok, itt több réteg alkotja a neurális hálózatot, és a köztes rétegeket rejtett rétegeknek nevezzük. Ezen belül is a teljesen összekötött hálózatokat alkalmazzák a leggyakrabban, ami azt jelenti, hogy egy (bemeneti vagy rejtett) réteg minden neuronja kapcsolódik a következő réteg minden neuronjához. Formalizálva:

$$\text{Első réteg: } \mathbf{z}_1 = f_1(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1)$$

$$\text{Második réteg: } \mathbf{z}_2 = f_2(\mathbf{W}_2 \mathbf{z}_1 + \mathbf{b}_2)$$

$$\dots$$

$$\text{Kimeneti réteg: } \hat{y} = f_k(\mathbf{W}_k \mathbf{z}_{k-1} + \mathbf{b}_k)$$

ahol $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_k$ a súlymátrixokat, $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ a konstans eltolásokat, valamint f_1, f_2, \dots, f_k az aktivációs függvényeket jelöli.

Aktivációs függvényként a későbbiekben a szigmoid és a hiperbolikus tangens függvényeket fogjuk alkalmazni, amelyek rendre a következők:

$$\text{Sigmoid: } \sigma(x) = \frac{1}{1 + e^{-x}}$$

$$\text{Hiperbolikus tangens: } \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

A hálózat topológiájának meghatározásán kívül a másik nehéz feladat a súlyok optimalizálása. Erre különböző gradiens alapú módszereket szoktak alkalmazni, hiba-visszaterjesztéses módszerrel kiegészítve. Erről részletesebben *Altrichter et al. (2006)* tanulmányában lehet olvasni, itt nem tárgyaljuk.

Regressziós feladatoknál (amilyen a mortalitás modellezése is) kedvelt veszteségfüggvény az átlagos négyzetes hiba (*mean squared error – MSE*):

$$L(y_i, \hat{y}_i) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

3.2. Visszacsatolt neurális hálózatok

Az előbb bemutatott előre-csatolt neurális hálózatok főként keresztmetszeti adatokon alkalmazhatók, felügyelt tanulóssal.³ A gyakorlatban nagyon sokszor szekvenciális adatok állnak rendelkezésre, tehát az adatok valamilyen sorrendben értel-

³ Felügyelt tanulásról beszélünk, amikor egyaránt rendelkezésre állnak a bemeneti és a hozzájuk tartozó kimeneti adatok.

mezhetők és kezelhetők. Gondolhatunk szövegekre, ahol a szavak sorrendje fontos és nem felcserélhető, de ilyenek a hang- vagy képalapú inputok is. Ezek mellett természetesen a különböző idősoros adatok is nagyon jól modellezhetők ezekkel a típusú hálózatokkal, mint például egy árfolyam vagy a mortalitási ráta alakulása. Az RNN gyakori alkalmazásai közé tartozik a szentimentelemzés, az idősorok előrejelzése, a gépi fordítás, a képfeliratozás vagy a szöveg- vagy videógenerálás.

A visszacsatolt neurális hálózatok bemeneti adatként használják a hálózat korábbi állapotát is. Strukturájuk a következő egyenletekkel adható meg:

$$\mathbf{h}^{(t)} = f(\mathbf{W}_h \mathbf{h}^{(t-1)} + \mathbf{W}_x \mathbf{x}^{(t)} + \mathbf{b}_x)$$

$$\mathbf{y}^{(t)} = \mathbf{W}_y \mathbf{h}^{(t)} + \mathbf{b}_y,$$

ahol a súlyok, eltolások és aktivációs függvények jelölése megegyezik az előző alfejezetben leírtakkal. Az egyenletekben látható, hogy az előrecsatolt neurális hálózatokhoz képest a visszacsatolt neurális hálózat a $\mathbf{h}^{(t)}$ közbülső állapotot is felhasználja inputnak a következő kimenet becslésére.

3.3. LSTM-hálózatok

Hochreiter–Schmidhuber (1997) írják le az LSTM-hálózatot, ami a visszacsatolt neurális hálózatok egy speciális típusa. Az egyszerű RNN a hosszú távú hatásokat nem kezeli megfelelően, a tanítás során előkerül az ún. „eltűnő gradiens probléma” (*vanishing gradient problem*), erre mutatnak rá *Bengio–Simard–Frasconi (1994)*. Ennek a kezelésére alkották meg az LSTM-hálózatokat. Az LSTM négy-szer annyi paramétert használ, mint az RNN, és „kapus” kialakításának köszönhetően meg tudja őrizni a korábbi adatokból a későbbi kimenetek szempontjából releváns információt. Az LSTM az alábbi elemekből épül fel:

$$\text{Felejtéskapu (forget gate): } \mathbf{f}^{(t)} = \sigma(\mathbf{W}_{f,h} \mathbf{h}^{(t-1)} + \mathbf{W}_{f,x} \mathbf{x}^{(t)} + \mathbf{b}_f)$$

$$\text{Bemeneti kapu (input gate): } \mathbf{i}^{(t)} = \sigma(\mathbf{W}_{i,h} \mathbf{h}^{(t-1)} + \mathbf{W}_{i,x} \mathbf{x}^{(t)} + \mathbf{b}_i)$$

$$\text{Kimeneti kapu (output gate): } \mathbf{o}^{(t)} = \sigma(\mathbf{W}_{o,h} \mathbf{h}^{(t-1)} + \mathbf{W}_{o,x} \mathbf{x}^{(t)} + \mathbf{b}_o)$$

$$\mathbf{c}'^{(t)} = \tanh(\mathbf{W}_{c,h} \mathbf{h}^{(t-1)} + \mathbf{W}_{c,x} \mathbf{x}^{(t)} + \mathbf{b}_c)$$

$$\mathbf{c}^{(t)} = \mathbf{f}^{(t)} * \mathbf{c}^{(t-1)} + \mathbf{i}^{(t)} * \mathbf{c}'^{(t)}$$

$$\text{A végső kimenet: } \mathbf{h}^{(t)} = \mathbf{o}^{(t)} * \tanh(\mathbf{c}^{(t)}),$$

ahol * jelöli az Hadamard-szorzatot (elemenkénti szorzat), σ a szigmoid és tanh a hiperbolikus tangens függvényt.

4. A magyar halandósági adatok modellezése

A magyar halandósági adatok modellezésére a Human Mortality Database (*HMD*, 2022) nyilvánosan elérhető adatbázisból töltöttük le a naptári évek és életkorok szerinti központi kitettségek és halálozások számát. Mind a férfi-, mind a női adatok 1950-től állnak rendelkezésre egészen 2020-ig ($1950 \leq t \leq 2020$).⁴ Korosztályok tekintetében a 0 évesektől a 99 évesekig használtuk az adatokat ($0 \leq x \leq 99$). Ezekből kiszámítottuk az $\ln(m_{x,t})$ logaritmizált központi halandósági rátákat, amelyeket ezután lineáris transzformációval a $[0, 1]$ intervallumba transzformáltunk. A bemeneti adatok egyszerűen az adott naptári évhez és életkorhoz tartozó mortalitási ráta logaritmusának a naptári év szerinti késleltetett értékei, illetve értelemszerűen a kimeneti adat az adott naptári év és életkor mortalitási rátájának a logaritmusa. Egyéb életkorok mortalitási rátáit és további magyarázóváltozókat ebben az elemzésben nem alkalmaztunk. A bemeneti és a kimeneti adatok az alábbiak szerint foglalhatók össze:

$$\mathbf{z}_{x,t} = (\log(m_{x,t-k}), \dots, \log(m_{x,t-2}), \log(m_{x,t-1})),$$

$$y_{x,t} = \log(m_{x,t}),$$

ahol k a maximális késleltetés számát jelöli.

A modellezés során $k = 10$ -es beállítással dolgoztunk. Ez azt is jelenti, hogy minden életkor esetében az első 10 naptári évhez tartozó megfigyelés elvész, mivel hozzájuk nem tudunk 10 elemű bemeneti adatsort készíteni. Tehát a kimeneti adathalmaz az 1960-as évek mortalitási rátáival kezdődik.

A modellek előrejelzéseinek a kiértékelése céljából célszerű tanuló- és tesztelőhalmazra bontani a rendelkezésre álló adatokat, így elkerülhető a túlillesztés, és a tanulásban részt nem vevő adatokon tudjuk kiértékelni az egyes modelleket, ami realitásabb képet ad a teljesítményről. A tanuló- és tesztelőhalmazokat naptári évek alapján határoztuk meg, 1960-tól 2004-ig tart a tanuló (*TANULÓ*), valamint 2005-től 2020-ig a tesztelő (*TESZTELŐ*) adathalmaz. Formalizálva:

$$TANULÓ = \{(\mathbf{z}_{x,t}, y_{x,t}); 1950 + k \leq t \leq 2004, 0 \leq x \leq 99\}$$

$$TESZTELŐ = \{(\mathbf{z}_{x,t}, y_{x,t}); 2005 \leq t \leq 2020, 0 \leq x \leq 99\}$$

Fontos kiemelni – ahogy *Richman (2019)* is megteszi –, hogy az előrejelzéseket mindig csak lépésenként, a következő évre szabad elvégezni. Ez azért szükséges, mert a későbbi évek előrejelzéseikhez szükségünk van a korábbi évek becsléseire. A 2005-ös év mortalitási rátáit előre tudjuk jelezni a 2004 és 1994 közötti tényleges mortalitási rátákkal. A 2006-os év előrejelzésénél azonban már nem

⁴ A továbbiakban nem jelöljük, hogy a férfi- vagy a női adatokra vonatkozik-e az adott lépés, mert mindkét nem esetében hasonlóképpen jártunk el.

használhatjuk a megfigyelt 2005-ös mortalitási rátákat, hanem a 2005-ös előrejelzéseket kell használnunk. Így kapjuk meg az 1950–2004 közötti évek információtartalmát alapján modellünk előrejelzéseit a következő 16 naptári évre.

Az LSTM neurális hálózatokhoz az R programozási nyelv *keras* (Allaire–Chollet, 2022) csomagját használtuk. A modellhez egy 12 neuronból álló LSTM-cellát, valamint egy 1 neuronból álló kimeneti réteget alkalmaztunk. A tanításhoz szükséges paraméterek száma 685. Egy népszerű „hüvelykujjszabály” alapján a neuronok számát úgy állítottuk be, hogy a paraméterek száma ne haladja meg a megfigyelések számának az egyötödét. A neurális hálózat felépítéséhez a következő hiperparamétereket alkalmaztuk: aktiválófüggvénynek hiperbolikus tangens, veszteségfüggvénynek átlagos négyzetes hibát, illetve optimalizálás céljára az Adam-optimalizálót (Kingma–Ba, 2014). Az Adam-algoritmusban a kötegek (*batches*) méretét 600 eleműnek, a korszakok (*epochs*) számát pedig 50-nek választottuk. Validációs halmaznak a tanulólármaz utolsó 10%-át használtuk. Ezenkívül ún. korai megállítást (*early stopping*) alkalmaztunk, ami az előre megadott epoch-számnál korábban befejezi a tanítást, amennyiben a hiba elkezd növekedni a validációs halmazon. A korai megállítással a futások gyorsabbak, az egyes eredmények pedig kiegyensúlyozottabbak lettek. Az előbbiekben definiált beállításokkal készítettük el a modellt külön-külön mindkét nemre.

A Lee–Carter-modellhez az R *StMoMo* csomagjának (Villegas–Kaishev–Milossovich, 2016) segítségével építettünk egy Poisson-féle Lee–Carter-modellt a férfi- és női halandósági adatokra.

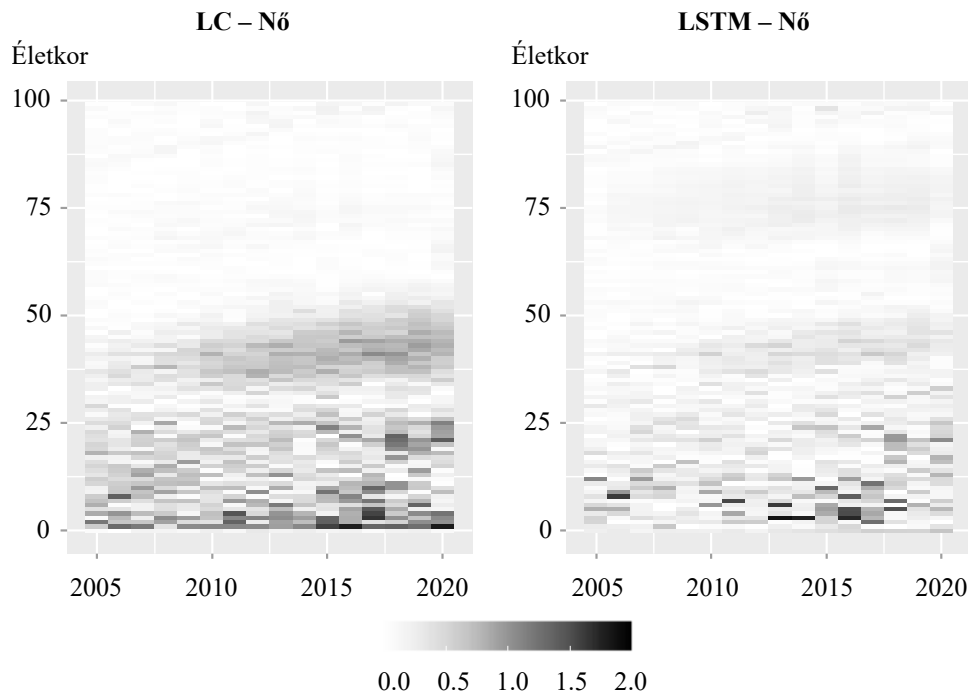
A 3. és a 4. ábrán látható, hogy az LSTM- és a Lee–Carter-modellek milyen hibákat vétenek abszolút értékben az egyes életkorokban és naptári években a tesztelőhalmazon.

Az eredmények azt mutatják, hogy az LSTM hibái sokkal egyenletesebbek. Jól látható, hogy mindkét modell a fiatal korcsoportban véti a legnagyobb hibákat, de az LSTM jelentősen jobb teljesítményt nyújt. A Lee–Carter-modell a fiatalok körében felülbecsüli a mortalitást, az LSTM esetében vegyes a kép. A legnagyobb különbség a középkorúak adatainál tapasztalható, itt a Lee–Carter-modell alulbecsül, ahogy az LSTM is, bár az utóbbi sokkal kisebb mértékben. Az idős korosztályban a Lee–Carter-modell nemcsak felveszi a versenyt az LSTM-mel, de minimálisan jobb teljesítményt is nyújt. Az LSTM kicsit felülbecsüli a mortalitást az idősek körében. Összességében az LSTM jóval kisebb hibákkal jelzi előre a mortalitást a női adatokon, a tesztelőhalmazon az LSTM MSE-értéke 0,081, míg a Lee–Carteré 0,175.

3. ábra

Az LSTM- és a Lee–Carter-modell reziduálisainak alakulása abszolút értékben a női adatok teszhalmazán

The absolute value of the Lee–Carter and the LSTM residuals on the female test set



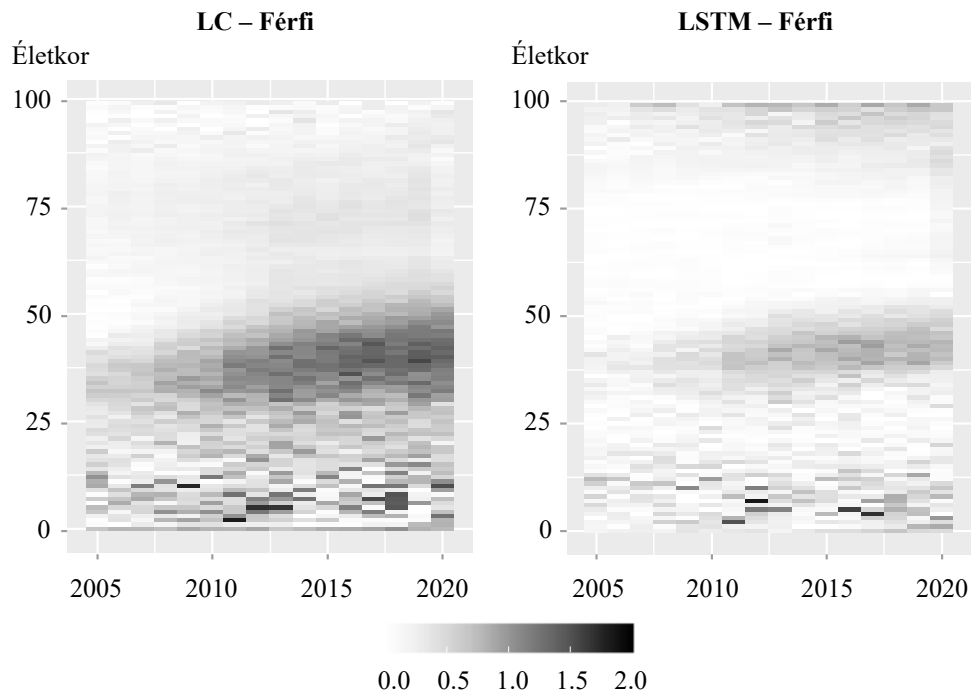
Forrás: saját szerkesztés.

A férfiak adatai esetében is jóval kedvezőbb az LSTM eredménye. Hasonlóan a női adatokhoz, itt is egyenletesebbek a hibák a tesztelőhalmazon. A fiatal és a középkorú generációk esetében egyértelműen jobb az LSTM előrejelzése, mint az LC-é. A Lee–Carter-modell mindenhol szisztematikusan felülbecsüli a mortalitási rátákat. Az LSTM a fiataloknál – a női modellhez hasonlóan – vegyes képet ad a hibatagok tekintetében, hol felül-, hol alulbecsüli a tényleges értéket. A középkorúaknál az LSTM is felülbecsül, de jóval kevésbé. A férfiak adatainál az LSTM jobb az időskorúak esetében, kb. 85 éves korig, utána azonban elkezd egyre inkább alulbecsülni a mortalitást. A Lee–Carter-modell az idősek halandóságát is felülbecsüli, de 85 éves kor felett kisebb hibákat eredményez, mint az LSTM. Összességében itt is sokkal jobb eredményt mutat az LSTM, amelynek MSE-értéke a tesztelőhalmazon 0,104, míg a Lee–Carteré 0,348.

4. ábra

Az LSTM- és a Lee–Carter-modell reziduálisainak alakulása abszolút értékben a férfiak adatainak tesztalmazán

The absolute value of the Lee–Carter and the LSTM residuals on the male test set



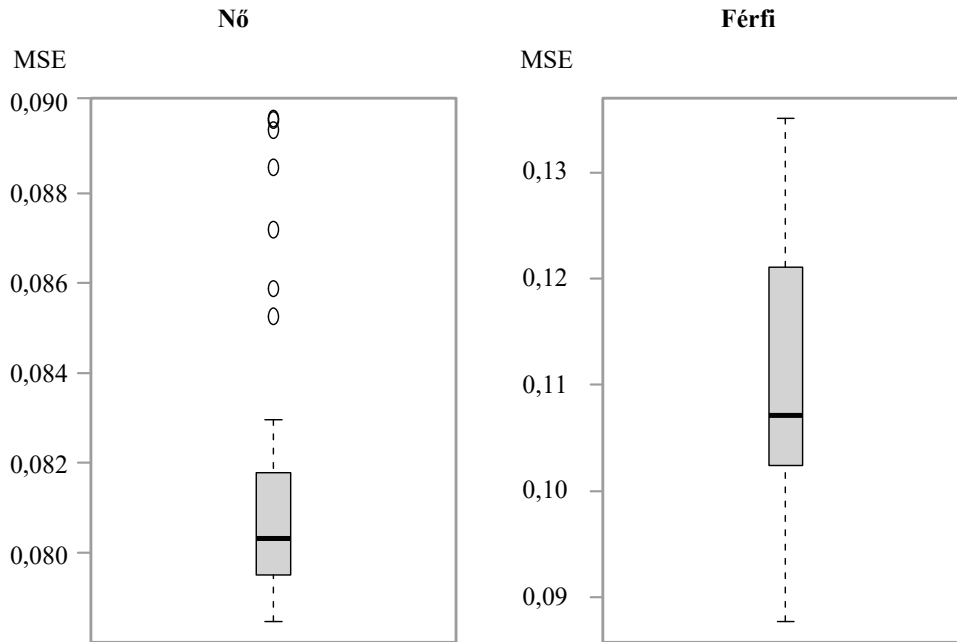
Forrás: saját szerkesztés.

Azt is érdemes kiemelni, hogy az LSTM eredményei a férfi- és női adatok esetében jóval közelebb állnak egymáshoz, míg a Lee–Carter-modell a férfiak adatain jelentősen nagyobb hibákat vétett.

A neurális hálózatok optimalizálása során eltérő eredményekre jutunk az egyes futtatásoknál. Ezért mind a férfi-, mind a női adatokra tanítottunk 50-50 különböző LSTM-modellt, hogy látható legyen a futások okozta variancia a modellekben. Az eredményeket az 5. ábra foglalja össze.

Az ábrán látható, hogy a modellek teljesítménye összességében elég egyenletes. A női adatokon az MSE-értékek variációja kisebb. Ami lényeges, hogy az 50-50 futás legrosszabb eredményei is jelentősen jobbak a Lee–Carter-modelléhez képest.

5. ábra

50-50 különböző LSTM-optimalizáció MSE-értékei a teszhalmazon*MSE values of 50 different LSTM optimization runs for each sex on the test set*

Forrás: saját szerkesztés.

1. táblázat

Az 50-50 LSTM-modell minimum, maximum és átlagos MSE-értékei, valamint a Lee-Carter-modell MSE-értéke a tesztelőhalmazon*Minimum, mean and maximum MSE values of 50 different LSTM optimization runs and Lee-Carter MSE values for each sex on the test set*

Nem\ MSE	LSTM minimum	LSTM átlag	LSTM maximum	LC
Férfi	0,088	0,110	0,135	0,348
Nő	0,078	0,081	0,090	0,175

Forrás: saját számítás.

A gyakorlatban természetesen azokat a modelleket fogjuk használni, amelyek a legkisebb hibát vétették, tehát a legjobban tanultak. Így a táblázat LSTM-minimumértékei tekinthetők mérvadónak.

5. Összefoglalás

Tanulmányunkban a mostanában igen népszerű neurális hálózatos megközelítéseket mutattuk be a halandóság modellezése kapcsán. A szakirodalomban megjelent korábbi tanulmányok alapján világosan látszik, hogy ez az irány nem téves, és az egyes hálózatstruktúrák igen jól megragadják azokat a szekvenciális mintázatokat, amelyek a halandósági adatokban is kirajzolódnak.

Elkészítettünk egy LSTM-hálózatot, amellyel a magyar halandósági adatokat modelleztük, külön-külön a két nemre. Az eredményeink azt mutatják, hogy a neurális hálózatok képesek nagyon jól leírni a magyar halandóságot. Az előrejelzések eredményét a népszerű Lee–Carter-modell alapján készített projekciókkal vetettük össze. Eredményeinkből látszik, hogy a fiatal és a középkorú népességet illetően egyértelműen jobb előrejelzéseket lehet adni LSTM-hálózatok segítségével. A férfiak esetében az idős korcsoportban is valamivel jobb az LSTM-modell, a nők adatain viszont az LC-modell teljesített jobban.

Tanulmányunkban pontbecsléseket adtunk az egyes mortalitási rátákra. Ahogy a szakirodalmi összefoglalóban is jeleztük, a becslések varianciája szintén fontos kérdés az aktuáriusi gyakorlatban. A későbbiekben érdemes lehet meghatározni ezt is, például a *Marino–Levantesi–Nigri (2021)* által javasolt bootstrap technikával.

A neurális hálózatos modellek teljesítménye jó eséllyel javítható. A hiperparaméterek optimalizációja a dolgozatunkban nem volt teljes körű, mivel nem ez állt a középpontban. A hiperparamétereket a korábbi szakirodalmi ajánlások, valamint saját tapasztalat alapján állítottuk be a végső modellhez.

Tanulmányunkban a mortalitási ráták minél pontosabb előrejelzésére összpontosítottunk, miközben háttérbe szorultak a gyakorlati alkalmazások. Ahogyan például *Májner–Kovács (2011)* és *Vékás (2017)* cikkében olvasható, a mortalitási ráták előrejelzése kulcsfontosságú az élettartam-kockázat, az életbiztosítási díjkalkuláció, valamint az életjáradékok esetében. E területeken érdemes lehet megismételni a korábbi elemzéseket a neurális hálózatokkal nyert előrejelzések felhasználásával.

Irodalom

- Ágoston K. Cs. – Kovács E. (2000): *Halandósági modellek*. Aktuárius jegyzetek. 3. kötet. Budapesti Corvinus Egyetem. Budapest.
- Allaire, J. – Chollet, F. (2022): *keras: R Interface to 'Keras'*. R package version 2.9.0. <https://CRAN.R-project.org/package=keras>
- Altrichter M. – Horváth G. – Pataki B. – Strausz Gy. – Takács G. – Valyon J. (2006): *Neurális hálózatok*. Panem Kft. Budapest.

- Asteriou, D. – Hall, S. G. (2015): *Applied Econometrics*. 3rd Edition. Palgrave MacMillan. London.
- Bengio, Y. – Simard, P. – Frasconi, P. (1994): Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*. Vol. 5. No. 2. pp. 157–166. <https://doi.org/10.1109/72.279181>
- Brouhns, N. – Denuit, M. – Vermunt, J. K. (2002): A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*. Vol. 31. No. 3. pp. 373–393. [https://doi.org/10.1016/S0167-6687\[02\]00185-3](https://doi.org/10.1016/S0167-6687[02]00185-3)
- Deprez, P. – Shevchenko, V. P. – Wüthrich, V. M. (2017): Machine learning techniques for mortality modeling. *European Actuarial Journal*. Vol. 7. pp. 337–352. <https://doi.org/10.1007/s13385-017-0152-4>
- Hochreiter, S. – Schmidhuber, J. (1997): Long Short-Term Memory. *Neural Computation*. Vol. 9. No. 8. pp. 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). www.mortality.org (a letöltés dátuma: 2022. 06. 05.)
- Kingma, D. P. – Ba, J. (2014): Adam: A method for stochastic optimization. *arXiv*. arXiv:1412.6980. <https://doi.org/10.48550/arXiv.1412.6980>
- Kleinow, T. (2015): A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*. Vol. 63. pp. 147–152. <https://doi.org/10.1016/j.insmatheco.2015.03.023>
- Koissi, M. C. – Shapiro, A. F. – Hognas, G. (2006): Evaluating and Extending the Lee–Carter Model for Mortality Forecasting Confidence Interval. *Insurance: Mathematics and Economics*. Vol. 38. Issue 1. pp. 1–20. <https://doi.org/10.1016/j.insmatheco.2005.06.008>
- Lee, R. D. – Carter, L. R. (1992): Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association*. Vol. 87. No. 419. pp. 659–671. <https://doi.org/10.2307/2290201>
- Levantesi, S. – Nigri, A. (2020): A random forest algorithm to improve the Lee–Carter mortality forecasting: impact on q-forward. *Soft Computing*. Vol. 24. Issue 12. pp. 8553–8567. <https://doi.org/10.1007/s00500-019-04427-z>
- Levantesi, S. – Pizzorusso, V. (2019): Application of machine learning to mortality modeling and forecasting. *Risks* Vol. 7. No. 1.: 26. <https://doi.org/10.3390/risks7010026>
- Li, N. – Lee, R. (2005): Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography*. Vol. 42. No. 3. pp. 575–594. <https://doi.org/10.1353/dem.2005.0021>
- Marino, M. – Nigri, A. – Levantesi, S. (2021): Deepening Lee–Carter for longevity projections with uncertainty estimation. *arXiv*. arXiv:2103.10535. <https://doi.org/10.48550/arXiv.2103.10535>
- Májér I. – Kovács E. (2011): Élettartam-kockázat – a nyugdíjrendszerre nehezedő egyik teher. *Statisztikai Szemle*. 89. évf. 7–8. sz. 790–812. old.
- Nigri, A. – Levantesi, S. – Marino, M. – Scognamiglio, S. – Perla, F. (2019): A Deep Learning Integrated Lee–Carter Model. *Risks*. Vol. 7. No. 1.: 33. <https://doi.org/10.3390/risks7010033>
- Őri P. – Spéder Z. (2020): Folytonos átmenet: Magyarország népesedése 1920 és 2020 között. *Statisztikai Szemle*. 98. évf. 6. sz. 481–521. old. <https://doi.org/10.20311/stat2020.6.hu0481>

- Perla, F. – Richman, R. – Scognamiglio, S. – Wüthrich, M. V. (2021): Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*. Vol. 2021. No. 7. pp. 572–598. <https://doi.org/10.1080/03461238.2020.1867232>
- Petneházi G. – Gáll J. (2019): Mortality rate forecasting: can recurrent neural networks beat the Lee–Carter model? *arXiv*. arXiv:1909.05501. <https://doi.org/10.48550/arXiv.1909.05501>
- Richman, R. (2018): *AI in Actuarial Science*. SSRN. 24.07.2018. <http://dx.doi.org/10.2139/ssrn.3218082>
- Richman, R. – Wüthrich, M. V. (2019): Lee and Carter go Machine Learning: Recurrent Neural Networks. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3441030>
- Richman, R. – Wüthrich, M. V. (2021): A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*. Vol. 15. No. 2. pp. 346–366. <https://doi.org/10.1017/S1748499519000071>
- Rosenblatt, F. (1958): The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*. Vol. 65. No. 6. pp. 386–408. <https://doi.org/10.1037/h0042519>
- Schnürch, S. – Korn, R. (2022): Point and interval forecasts of death rates using neural networks. *ASTIN Bulletin*. Vol. 52. No. 1. pp. 333–360. <https://doi.org/10.1017/asb.2021.34>
- Vékás P. (2017): Nyugdíjcélú életjáradékok élettartam-kockázata az általánosított korcsoport-időszak-kohorsz modellkeretben. *Statisztikai Szemle*. 95. évf. 2 sz. 139–165. old. <https://doi.org/10.20311/stat2017.02.hu0139>
- Villegas, A. M. – Kaishev, V. – Millossovich, P. (2016): *StMoMo: An R package for stochastic mortality modelling*. SSRN. 13.11.2016. <http://dx.doi.org/10.2139/ssrn.2698729>