

Úvod

Statistické jazykové modelování je úloha hledání modelu, který přiřadí vysokou pravděpodobnost větám častým v daném přirozeném jazyce, zatímco větám vzácným nebo vyloženě nesprávným přiřadí nízkou pravděpodobnost. Kvalitní jazykový model je důležitou součástí systémů pro automatický přepis řeči nebo kompresi textu.

Jako základní měřítko kvality statistických jazykových modelů se používá slovní průměr entropie daného testovacího textu.

$$\text{per-word entropy} = -\frac{1}{N} \sum_{i=1}^N \log_2 p(w_i | w_1^{i-1}) \quad (1)$$

Základní model

V posledních několika letech dosahují nejlepších výsledků jazykové modely založené na rekurentních neuronových sítích [Mik12]. První takový úspěšný model je popsán dvojicí rovnic:

$$\mathbf{h}_t = \sigma(\mathbf{U}\mathbf{x}_t + \mathbf{W}\mathbf{h}_{t-1}) \quad (2)$$

$$\mathbf{y}_t = \text{softmax}(\mathbf{V}\mathbf{h}_t) \quad (3)$$

Studované rozšíření

V této práci bylo navrženo rozšíření základního modelu, kdy je rovnice pro výpočet skrytého stavu neuronové sítě (2) nahrazena následující rovnicí:

$$\mathbf{h}_t = \sigma(\mathbf{U}\mathbf{x}_t + (\mathbf{W}_m \odot \mathbf{W})\mathbf{h}_{t-1}) \quad (4)$$

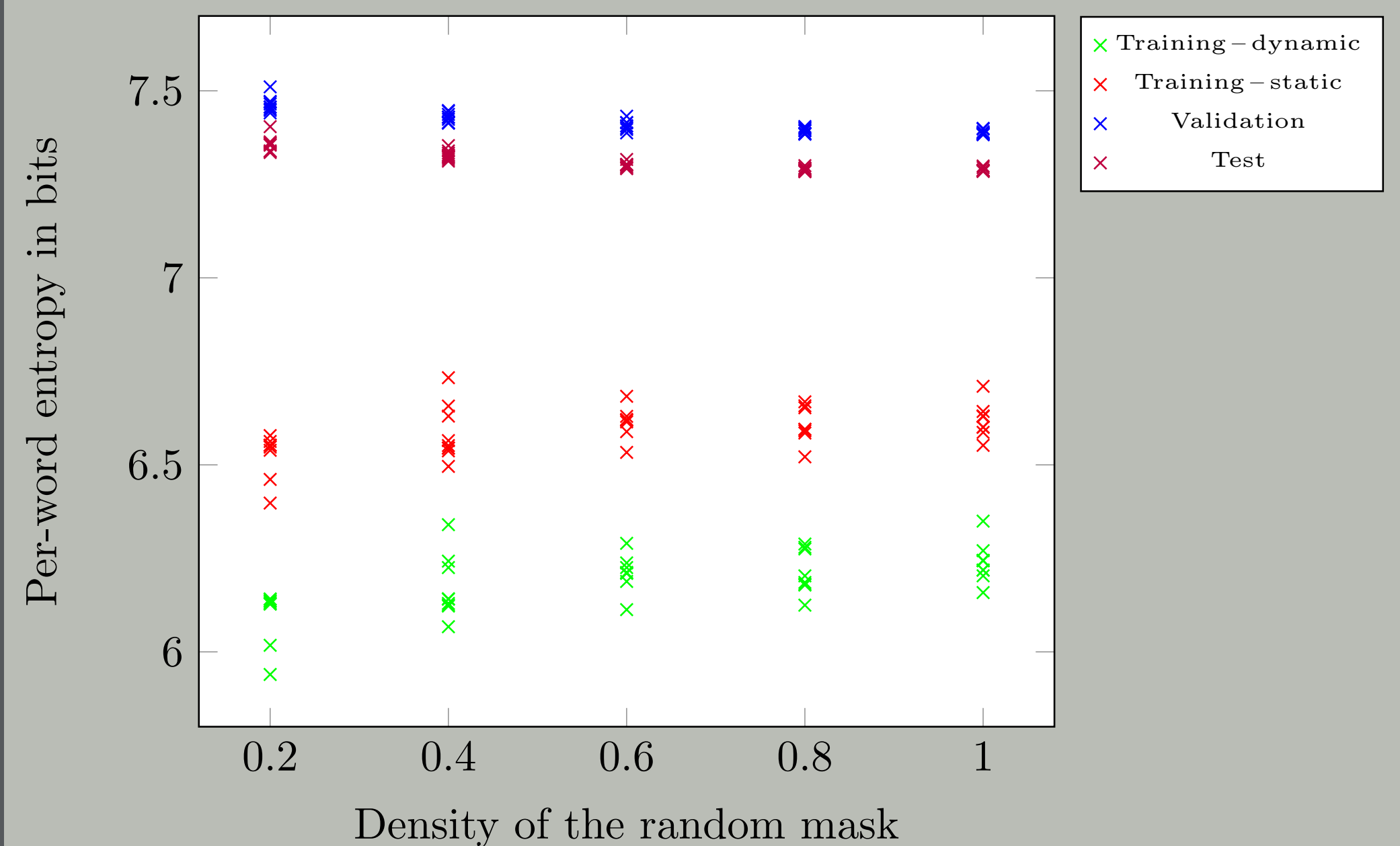
Maticе \mathbf{W}_m je binární maska o stejných rozměrech jako \mathbf{W} . Je nastavena na pevné hodnoty, takže omezuje počet aktivních rekurentních synapsí.

Toto rozšíření bylo inspirováno především modely Echo State Network [Jae03] a Structurally Constrained Recurrent Network [MJC⁺14].

Odkazy na literaturu

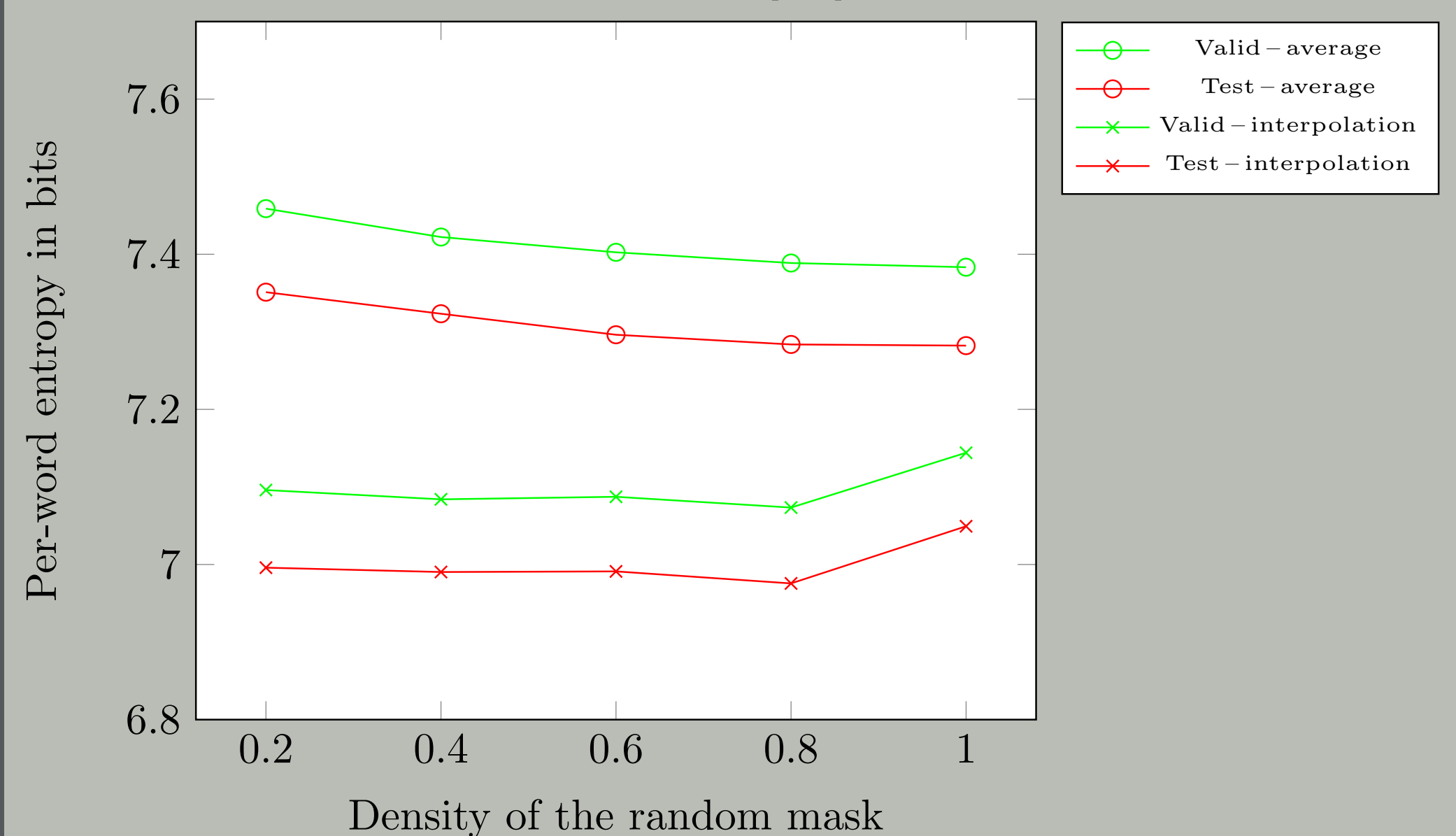
- [Jae03] Herbert Jaeger, *Adaptive nonlinear system identification with echo state networks*, Advances in Neural Information Processing Systems 15 (S. Becker, S. Thrun, and K. Obermayer, eds.), MIT Press, 2003, pp. 609–616.
- [Mik12] Tomáš Mikolov, *Statistical language models based on neural networks*, Ph.D. thesis, 2012, p. 129.
- [MJC⁺14] Tomas Mikolov, Armand Joulin, Sumit Chopra, Michaël Mathieu, and Marc'Aurelio Ranzato, *Learning longer memory in recurrent neural networks*, CoRR [abs/1412.7753](https://arxiv.org/abs/1412.7753) (2014).

Výsledky jednotlivých modelů



Kombinace modelů

$$p(w_t | p_1^{t-1}) = \frac{1}{N} \sum_{i=1}^N p_i(w_t | p_1^{t-1}) \quad (5)$$



Závěr

Navržený model sám o sobě nedosahuje žádných zlepšení. Kombinace prořídlejších modelů však dosahuje lepších výsledků než kombinace stejného množství plně propojených rekurentních sítí. Pokračováním práce je tak další studium vnitřního chování prořídlejších rekurentních neuronových sítí.