

Simulation von Nutzerverhalten bei der Interaktion mit Sprachdialogsystemen mittels Künstlicher Neuronaler Netzwerke

Stefan Hillmann, Benjamin Weiss, Thilo Michael, Sebastian Möller
TU Berlin

vorname.nachname@tu-berlin.de, sebastian.moeller@tu-berlin.de

Bei der Entwicklung von Sprachdialogsystemen (SDS) wird die Simulation von Nutzerverhalten im Wesentlichen für zwei Zwecke verwendet, zum einen für das Training von Dialogmanagern, die auf statistischen Ansätzen oder maschinellem Lernen beruhen. Zum anderen können solche Simulationen für die Evaluierung von Mensch-Computer-Interaktion verwendet werden [4, 5]. Zur Evaluation werden Dialoge zwischen simulierten Nutzer und SDS, meist auf der Ebene der Intention, generiert und anschließend analysiert. Intention meint, dass der Dialog nicht in natürlicher Sprache, sondern abstrakt mittels Dialogakten und ausgetauschten Informationen abgebildet wird (z.B. adressierte Slots und entsprechende Werte).

Zu den ersten Ansätzen in der Nutzersimulation zählen einfache Bi-Grammmodelle (jeweils zwei aufeinanderfolgende Dialogschritte werden betrachtet), die zu immer komplexeren (z.B. [6, 8]), aber dennoch stets probabilistischen oder regelbasierten Simulationen weiterentwickelt wurden. Die darauffolgende Generation von Nutzersimulationen verwendet Ansätze des maschinellen Lernens, wie bspw. Bayessche Netze [7], Hidden Markov Models [9] oder Reinforcement Learning [3 (Kap. 4)].

Die auch zum maschinellen Lernen zählenden neuronalen Netze haben seit den ca. 2010er Jahren eine große Verbreitung gefunden, da die Verfügbarkeit notwendiger Rechenkapazitäten und die Entwicklung effektiver Trainingsmethoden zusammenfielen. Im Kontext von SDS bilden künstliche neuronale Netze (KNN) heute den Stand der Wissenschaft im Bereich der Spracherkennung und des Sprachverstehens. Auch für die Modellierung von Dialogverläufen bilden sie eine interessante Möglichkeit, da KNN prinzipiell dazu geeignet sind, die im Dialogverlauf und dem gegebenen Kontext implizit enthaltenen Informationen abzubilden. Dies gilt vor allem für Rekurrente Neuronale Netze (RNN) und LSTM-RNN (Long Short-Term Memory RNN), welche die über die Zeit abgebildeten Einflussgrößen auf den Dialogverlauf (z.B. häufige Rückfragen durch das System) erlernen können [2, 1].

In dem Vollbeitrag soll die Leistungsfähigkeit von mehrschichtigen KNN und LSTM-RNN bei der Vorhersage von Nutzeraktionen mit einer rein probabilistischen Baseline verglichen werden. Dazu werden die Modelle mit Dialogdaten aus der Domäne Restaurantauskunft trainiert und validiert. Die Dialoge werden auf der Ebene der Intention modelliert, wobei sowohl Dialogakte alleine, als auch zusammen mit adressierten Systemslots betrachtet werden. Die trainierten Netzwerke können perspektivisch zu kompletten Nutzersimulationen ausgebaut werden. Der Beitrag diskutiert weiterhin mögliche Einsatzszenarien von KNN/RNN basierten Nutzersimulationen, wie auch die Vor- und Nachteile gegenüber anderen Methoden des maschinellen Lernens bei der Nutzung in Nutzersimulationen.

[1] P. Crook und A. Marin (2017): "Sequence to Sequence Modeling for User Simulation in Dialog Systems", ISCA 2017, S. 1706–10. ISCA, 2017.

[2] L. Asri, J. He, und K. Suleman (2016): "A Sequence-to-Sequence Model for User Simulation in Spoken Dialogue Systems", ISCA 2016, S. 1151–55, 2016.

[3] H. Cuayáhuitl (2009): "Hierarchical Reinforcement Learning for Spoken Dialogue Systems". Diss. University of Edinburgh

[4] K.-P. Engelbrecht (2011): "Estimating Spoken Dialog System Quality with User Models", Diss., Technische Universität Berlin

[5] Hillmann, S. (2017): "Simulation-Based Usability Evaluation of Spoken and Multimodal Dialogue Systems", Diss. Technische Universität Berlin, Springer [in Druck]

[6] S. Hillmann und K.-P. Engelbrecht. (2013): „Aufgabenmodellierung in der Simulation von Interaktionen mit Sprachdialogsystemen“. ESSV 2013, Hrsg. Petra Wagner, 20–27. TUDpress.

[7] O. Pietquin. (2004): "A Framework for Unsupervised Learning of Dialogue Strategies". Diss., Faculté Polytechnique de Mons.

[8] J. Schatzmann, B. Thomson, K. Weilhammer, H. Ye und S. Young. (2007): "Agenda-based user simulation for bootstrapping a POMDP dialogue system", Proc. ACL. Morristown, NJ, USA, S. 149–152.

[9] J. Schatzmann und S. Young. (2009): "The Hidden Agenda User Simulation Model". In: IEEE Trans. on Audio, Speech, and Language Processing 17(4), S. 733–747.