# Investigating Variable Dependencies in Dialogue States

**Anh Duong Trinh** †, **Robert J. Ross** †, **John D. Kelleher** ‡
† School of Computer Science
‡ Information, Communications & Entertainment Institute
Technological University Dublin
ADAPT Centre, Ireland
anhduong.trinh@mydit.ie, {robert.ross, john.d.kelleher}@dit.ie

## Abstract

Dialogue State Tracking is arguably one of the most challenging tasks among dialogue processing problems due to the uncertainties of language and complexity of dialogue contexts. We argue that this problem is made more challenging by variable dependencies in the dialogue states that must be accounted for in processing. In this paper we give details on our motivation for this argument through statistical tests on a number of dialogue datasets. We also propose a machine learning-based approach called energy-based learning that tackles variable dependencies while performing prediction on the dialogue state tracking tasks.

## 1 Introduction

Dialogue Systems have a wide application in the modern world to assist users with conversational activities. Among dialogue processing tasks dialogue state tracking is the process of identifying user intents within the dialogue contexts. Generally task-oriented dialogue systems define dialogue states as a combination of slot-value pairs. We argue that there exist relationships among the slots, that must be taken into account in the dialogue state tracking process to reflect the natural human way of processing information.

The idea of leveraging variable dependencies in the dialogue state tracking process is not new to the research community. There have been several published works around this phenomenon such as in the multi-task learning model (Trinh et al., 2018), the language modelling tracker (Platek et al., 2016), Conditional Random Fields (Kim and Banchs, 2014), Attention-based Sequence-to-Sequence model (Hori et al., 2016), and the work by Williams (2010). We find that these approaches are good at leveraging variable dependencies at different stages of the architecture.

In this paper we perform statistical tests on spoken dialogue data of Dialogue State Tracking Challenge (DSTC) series including the second challenge (Henderson et al., 2014a) and the third challenge (Henderson et al., 2014b). We demonstrate that there exist strong dependencies between dialogue slots that validate the motivation for our research direction. Moreover, we present the energy-based learning approach to predicting dialogue states while accounting for the variable relationships.

## 2 Categorical Data Analysis

To investigate the presence or not of variable dependencies, we perform statistical tests pairwise on labels for bivariate statistics. The chosen method is Pearson's chi-square test, which is effective for categorical data. There exist several measurements of association strength between variables directly related to the chi-square test statistics such as $\phi$ coefficient, contingency coefficient $C$, and Cramer's $V$. These measures are scaled between 0 and 1 indicating that 1 is the perfect relationship and 0 is no relationship between variables.

We report the statistics of DSTC2 and 3 data in table 1. The variable dependencies are reported with the chi-square test-based Cramer's $V$ coefficient.

In the result we observe that all statistical significance values $p < 0.05$, that confirms the existence of variable dependencies within dialogue data. We also find that these dependencies are stably strong ($V \geqslant 0.15$).

To expand on this, let us consider the case of the DSTC2 data. Here, the analysis shows that slot *food* is strongly dependent on slots *price range* and *area* in the domain. This implication indicates that when processing a restaurant search query, the

| | DSTC2 | | | | | DSTC3 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | food | price | area | | food | price | area | type |
| food | - | | | food | - | | | |
| price | 0.305 | - | | price | 0.248 | - | | |
| area | 0.269 | 0.214 | - | area | 0.163 | 0.232 | - | |
| | | | | type | 0.300 | 0.195 | 0.220 | - |

Table 1: Statistical tests on DSTC2 & 3 data. The results are reported with the Cramer's $V$ coefficient.

system should not process *food* without considering *price range* or *area* and vice versa. For example, a query such as *"expensive French food in city centre"* should return more results than *"expensive fast food"*. On the other hand, the relationship between *price range* and *area* is weaker than with slot *food*, but still relatively strong.

Overall, the data analysis results validate our motivation of accounting variable dependencies in dialogue state predictions. By adding these dependencies as extra information in the interpretation process, we argue that we can enhance the dialogue state tracker on tracking more challenging situations.

## 3 Energy-based Dialogue State Tracker

Given the strong dependencies seen between these dialogue state tracking variables, we argue that it is important that any tracking of dialogue state variables should take such dependencies into account as to ignore these dependencies is to assume an Independence that is not valid. To that end we propose a machine learning-based approach, that are notable for tackling the associations between variables, to the dialogue state tracking task . Currently we are investigating the appropriateness of this approach to the dialogue state tracking challenge series.

The core of our on-going work is based on a structured prediction methodology based on Energy-Based Learning. Energy-based models are notably good at structured prediction tasks such as in our case where there are dependencies between a set of predicted variables (LeCun et al., 2006).

The main concept of this method is to associate a structured output $Y$ and the input vector $X$ with a scalar energy value $E = E(X, Y)$ and to measure their goodness of fit using an appropriate loss function $L(E, E^*)$ on those energy values (figure 1).

Currently we are developing energy-based dialogue state tracking models based on a number of

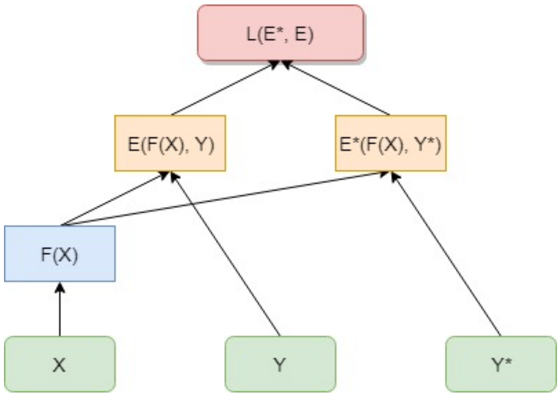


Figure 1: An example of Energy-Based Model, that consists of a feature network $F(X)$, an energy function $E(F(X), Y)$, and an objective function $L(E, E^*)$, where $X$ is input variable, $F(X)$ is a feature representation generated by a feature network, $Y$ is predicted output variable, and $Y^*$ is a gold standard label output variable.

energy-based architectures such as Structured Prediction Energy Networks (SPEN) (Belanger and McCallum, 2016; Belanger et al., 2017) and Deep Value Networks (DVN) (Gygli et al., 2017). Following these approaches we build our energy networks on top of a LSTM-based (Hochreiter and Schmidhuber, 1997) analyser that builds a feature representation for individual dialogue turns.

## 4 Conclusion

To date our approach has shown a lot of promise in improving on models where variable dependencies are otherwise ignored. In details our energy-based tracker outperforms a LSTM-based multi-task model (Trinh et al., 2018) on both DSTC2 & 3 main tasks. The SPEN methodology helps to improve DSTC2 performance measured with accuracy metric by 3%, while the DVN algorithm increases DSTC2 result by 5% and DSTC3 by 9%.

The observed improvement is achieved mainly due to the energy function and inference process of the energy-based learning approach that takes advantage of target variable dependencies.

## Acknowledgments

This research was supported by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

## References

David Belanger and Andrew McCallum. 2016. Structured Prediction Energy Networks. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48.

David Belanger, Bishan Yang, and Andrew McCallum. 2017. End-to-End Learning for Structured Prediction Energy Networks. In *Proceedings of the 34th International Conference on Machine Learning*.

Michael Gygli, Mohammad Norouzi, and Anelia Angelova. 2017. Deep Value Networks Learn to Evaluate and Iteratively Refine Structured Outputs. In *Proceedings of the 34th International Conference on Machine Learning*.

Matthew Henderson, Blaise Thomson, and Jason D. Williams. 2014a. The Second Dialog State Tracking Challenge. In *Proceedings of the SIGDIAL 2014 Conference*, pages 263–272.

Matthew Henderson, Blaise Thomson, and Jason D. Williams. 2014b. The Third Dialog State Tracking Challenge. In *Proceedings of 2014 IEEE Workshop on Spoken Language Technology*, pages 324–329.

Sepp Hochreiter and Jurgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Takaaki Hori, Hai Wang, Chiori Hori, Shinji Watanabe, Bret Harsham, Jonathan Le Roux, John R. Hershey, Yusuke Koji, Yi Jing, Zhaocheng Zhu, and Takeyuki Aikawa. 2016. Dialog State Tracking With Attention-Based Sequence-To-Sequence Learning. In *Proceedings of 2016 IEEE Workshop on Spoken Language Technology*, pages 552–558.

Seokhwan Kim and Rafael E. Banchs. 2014. Sequential Labeling for Tracking Dynamic Dialog States. In *Proceedings of the SIGDIAL 2014 Conference*, pages 332–336.

Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu Jie Huang. 2006. A Tutorial on Energy-Based Learning. *Predicting Structured Data*.

Ondrej Platek, Petr Belohlavek, Vojtech Hudecek, and Filip Jurcicek. 2016. Recurrent Neural Networks for Dialogue State Tracking. In *Proceedings of CEUR Workshop, ITAT 2016 Conference*, volume 1649, pages 63–67.

Anh Duong Trinh, Robert J. Ross, and John D. Kelleher. 2018. A Multi-Task Approach to Incremental Dialogue State Tracking. In *Proceedings of The 22nd workshop on the Semantics and Pragmatics of Dialogue, SEMDIAL*, pages 132–145.

Jason D. Williams. 2010. Incremental Partition Recombination For Efficient Tracking Of Multiple Dialog States. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5382–5385.