

Topic Switch Models for Dialogue Management in Virtual Humans

Wenjue Zhu¹, Andry Chowanda^{1,2*}, Michel Valstar¹

¹ School of Computer Science, The University of Nottingham, Nottingham, UK-GB
{psywz2,psxac6,pszmv}@nottingham.ac.uk

² School of Computer Science, Bina Nusantara University, Jakarta, ID

Abstract. This paper presents a novel data-driven Topic Switch Model based on a cognitive representation of a limited set of topics that are currently in-focus, which determines what utterances are chosen next. The transition model was statistically learned from a large set of transcribed dyadic interactions. Results show that using our proposed model results in interactions that on average last 2.17 times longer compared to the same system without our model.

Keywords: Social Relationship, Framework, Game-Agents, Interactions

1 Introduction

Current Dialogue Management (DM) systems “*are not natural enough*” and cannot sustain a coherent conversation with humans [2]. One issue is that systems are not able to “stay on topic” and are incapable of following a “train of thought”. As most DM systems have no notion of the concept of a topic and generate their responses based only on a set of predefined rules that operate on the specific words or phrases retrieved from the last user input.

To address this, we propose a novel data-driven Topic Switch Model (TSM), devise an algorithm for sensible topic switching and instantiate it in a software program that can imbue virtual humans with the capability of staying on topic or making sensible topic switches with the aim of achieving more coherent conversations with humans. Our TSM learns connections between topics, which allows for sensible topic switches, and learns connections between topics and utterances, which allows for the selection of sentences that match the current topic. The system is otherwise naive, in that it does not implement an agent’s goals, or states such as social relations or an agent’s emotion [3, 4]. However, it is entirely data driven and thus does not require crafting of any rule whatsoever. It is thus suggested that for full effectiveness the TSM be integrated into a more complex stateful model, perhaps with one TSM per state.

We evaluate the efficacy of our system by comparing it to a version of our system without the TSM enabled. When tested on over 20 participants, we show that people communicate on average 2.17 times longer with the agent when the TSM is enabled.

* Corresponding Author

2 Related work

A number of different approaches to DM have been proposed and implemented to date. Plan-based DM makes use of a general planner which is responsible for identifying the goal and making a plan. The plan consists of predefined operations and is aimed to achieve the final goal [1]. The most common and simplest approach to DM is to represent the dialogue as a graph where its nodes represents the dialogue states. The nodes usually define the proposed action based on the input of previous node [6]. Another common approach uses the concept of information state. In this approach, the state of conversation is formally represented by some informational components and a set of rules is defined for the DM to update the state and decide on the corresponding action according to the current state, system input and the applicable rule(s) [7].

3 The Topic Switch Model

Topic switches occur constantly in one’s brain, and are influenced by both internal factors (e.g. your own knowledge of topics and their relationships) and external factors, including what you hear and see. There may be several topics in a person’s mind at a time, each taking up a portion of one’s attention. Over time topics in the brain will be replaced by others because of the various influencing factors at any point of time. This is our abstract concept of a TSM.

As we focus here on text-based dialogue systems, topic extraction is comprised of text preprocessing and topic retrieval. Two natural language processing techniques (i.e. stop-word removal and stemming) are chosen to pre-process the text input. With respect to topic retrieval, we maintain a lookup table of words and their corresponding topics. This list was manually created by the authors.

Topic relations, that should be learned by virtual humans consist of three types of topic statistics. The topic frequency $P_f(t)$ is the prior probability of a topic occurring in an utterance. The second relation, concurrency probability $P_{con}(t_1, t_2, \dots, t_n)$, is the probability of two or more topics appearing in the same utterance. The third relation is the adjacency probability, $P_{adj}(t_1, t_2)$ representing the probability of topic t_2 occurring in a utterance if its previous utterance contains t_1 .

$$\text{Topic Frequency: } P_f(t) = \frac{n_t}{N_T} \quad (1)$$

$$\text{Concurrency Possibility: } P_{con}(t_1, t_2, \dots, t_n) = \frac{n_{con}(t_1, t_2, \dots, t_n)}{N_s} \quad (2)$$

$$\text{Adjacency possibility: } P_{adj}(t_1, t_2) = \frac{n_{adj}(t_1, t_2)}{N'_s} \quad (3)$$

where N_s is the number of utterances and N'_s is the potential times of two topics appearing in two adjacent utterances. In this paper, the topic statistics were obtained from the SEMAINE database [5].

Algorithm 1 Topic switch with external factors

Input: all topics L_T , stop words L_{sw} , pairs of words and topics L_p , topic statistics, sentence database, user input

Initialisation:

1. Randomly select 5 topics from L_T to be the topics in the brain, $L_1 = [t_1, t_2, t_3, t_4, t_5]$.
2. Topics of which the virtual human is thinking currently, $L_2 = []$.
3. Topics that are out of the brain, $L_3 = L_T - L_1$.

Procedure:

1. Read the text input and extract the user topic t_u . Find the topic t_i in L_1 which has the largest adjacency possibility with t_u . If no user topic is found, randomly select a topic t_i in L_1 based on the topic frequencies. Add t_i into L_2 , $L_2 = [t_i]$.
2. Generate a random number r between 0 and 1. Continuously try to find topics in $(L_1 - L_2)$ whose concurrency possibility with topics in L_2 is larger than r and add it into L_2 until no such topic is found.

3. Make a response by randomly selecting an utterance which contains all topics in L_2 .

4. Randomly select a topic t_{out} from $(L_1 - L_2)$ to be swapped out based on the adjacency possibilities. Then $L_1 = L_1 - [t_{out}]$ and $L_3 = L_3 + [t_{out}]$. The possibility of each topic

$$\text{being selected is } P_{out}(t) = \frac{\sum_{t_a \in L_2} (1 - P_{adj}(t_a, t))}{\sum_{t_a \in L_2, t_b \in (L_1 - L_2)} (1 - P_{adj}(t_a, t_b))}.$$

5. Randomly select a topic t_{in} from L_3 to be swapped in according to the adjacency possibilities. Then $L_1 = L_1 + [t_{in}]$ and $L_3 = L_3 - [t_{in}]$. The possibility of each topic in

$$L_3 \text{ being chosen is } P_{in}(t) = \frac{\sum_{t_a \in L_2} (P_{adj}(t_a, t))}{\sum_{t_a \in L_2, t_b \in L_3} (P_{adj}(t_a, t_b))}.$$

6. Empty L_2 and go to step 1 until the end of the conversation.

Table 1. Performance comparison between systems.

	System	Weather	Work	Christmas	Other	Total	TS/NTS
Mean	TS	3.7	4.05	5.45	4.15	17.35	2.17
	NTS	1.8	2.15	1.8	2.25	8	
SD	TS	2.03	1.76	3.89	3.73	8.14	3.55
	NTS	0.95	1.04	1.06	1.55	2.29	
Min	TS	1	2	1	1	11	1.10
	NTS	2	2	1	2	10	
Max	TS	6	8	17	12	39	3.90
	NTS	1	1	2	2	10	

4 Evaluation

Twenty participants were invited to interact with our proposed system with TSM implemented and with a baseline version without TSM implemented as a comparison. Participants were asked to start a conversation on a specific topic and stop when they thought the topic switch made by the DM was not sensible. As performance measure we counted the number of user turns before they stopped. Table 1 demonstrates the performance comparisons between both systems, giving mean, standard deviation, min and max number of turns interacted with either system for the different topics. The last column shows the relative

increase in interaction time, measured as the number of turns interacting with Topic Switch (TS) divided by the number of turns with No Topic Switch (NTS).

Table 1 shows that people always communicated longer with the TS model, and on average people interacted 2.17 times longer with TS. Additionally, the performance of TS partially depended on the discussed topic. NTS on the other hand had about the same performance for all specified topics.

5 Discussion

The results indicate that the system with TSM could make a more sensible response which was either staying on the current topic or switched to another related topic. Moreover, the system TS had slightly different performance on different topics. The main reason should be the DM had better knowledge of some topics than that of others because the knowledge it learned from real conversations cannot cover all topics. However, unreasonable topic switches may still occur in the TS system for various reasons, including multiple meanings of an individual word, unfamiliar topics and inaccurate topic extractions. Another drawback of our current implementation is that topics are a flat structure - there is no hierarchy or ontology. This means that choices about the grouping of topics had to be made. For example, ‘sunny’ is part of the topic ‘weather’, but an alternative choice would have been splitting that topic into ‘good weather’ and ‘bad weather’, or constructing a hierarchy.

Acknowledgement

The work by A. Chowanda and M. Valstar is partly funded by European Union’s Horizon 2020 research and innovation programme under grant agreement No 645378, ARIA-VALUSPA.

References

1. James F Allen and C Raymond Perrault. Analyzing intention in utterances. *Artificial intelligence*, 15(3):143–178, 1980.
2. Björn Bringert. *Programming Language Techniques for Natural Language Applications*. Department of Computer Science and Engineering, 2008.
3. Andry Chowanda, Peter Blanchfield, Martin Flinham, and Michel Valstar. Erisa: Building emotionally realistic social game-agents companions. In *Intelligent Virtual Agents*, pages 134–143. Springer International Publishing, 2014.
4. Andry Chowanda, Martin Flinham, Peter Blanchfield, and Michel Valstar. Playing with social and emotional game companions. In *Intelligent Virtual Agents*. 2016.
5. G. McKeown, M.F. Valstar, R. Cowie, M. Pantic, and M. Schröder. The semaine database: Annotated multimodal records of emotionally coloured conversations between a person and a limited agent. *Trans. Affective Computing*, 3(1):5–17, 2012.
6. Bahador Nooraei, Charles Rich, and C Sidner. A real-time architecture for embodied conversational agents: beyond turn-taking. *ACHI*, 14:381–388, 2014.
7. D.R. Traum and S. Larsson. The information state approach to dialogue management. In *Current and new directions in discourse and dialogue*. Springer, 2003.