April 03, 2019

SEQUENCE PREDICTION FOR PRINT/ SCAN ISSUES USING BAYESIAN NETWORKS

HP INC

Sequence Prediction for Print/Scan Issues using Bayesian Networks

Abstract

*Machine Learning algorithms are empowering a lot of software applications today. We address a business need where the Machine Learning model can help to resolve customer issues which arise during printing and scanning on devices or installing a software driver. The model is based on Bayesian Network diagnoses the problem and subsequently resolves it by suggesting a sequence of steps which increase the probability of fix. This is applied in the context of an installer or driver or a diagnostics tool (referred to as *component*) to arrive at a resolution. The model factors in the current state of the print system which comprises of Operating System, Printer model, localization etc. The resolution can be a static sequence which provides a sequence of steps based upon the initial state of the customer or it can be a dynamic sequence wherein user provides feedback against each action suggested by the model and the model takes that feedback into account and suggests the next step. We have a prototype for this model and are applying it as part of Machine Learning service of Print Scan Doctor (PSDr) and are evaluating the possibility of other applications.*

Problem statement

Customers use printers to print/scan documents/photos. Whenever they face issues while printing or scanning they call customer support agents for the resolution which incurs a huge cost. The transcripts of the call records are very rich in information i.e. issue faced by the customer, state of the printer and the environment, which steps solved the problem that customer was facing. If we can extract the solution and perform it automatically, we can save the call cost to customer support and reflect customer satisfaction. But, the transcripts of the conversation are highly unstructured and in bulk quantity. Which makes it very difficult for an engineer to manually go through it and extract the steps which solved the issues and get some related insights.

Our solution

The data i.e. conversation transcripts are freely written unstructured text. We propose a learning model which will learn from the transcripts of the call records and predict the sequence of steps which can solve the issue, given the state information of the customer setup. Fig 1 shows the flow of our approach.

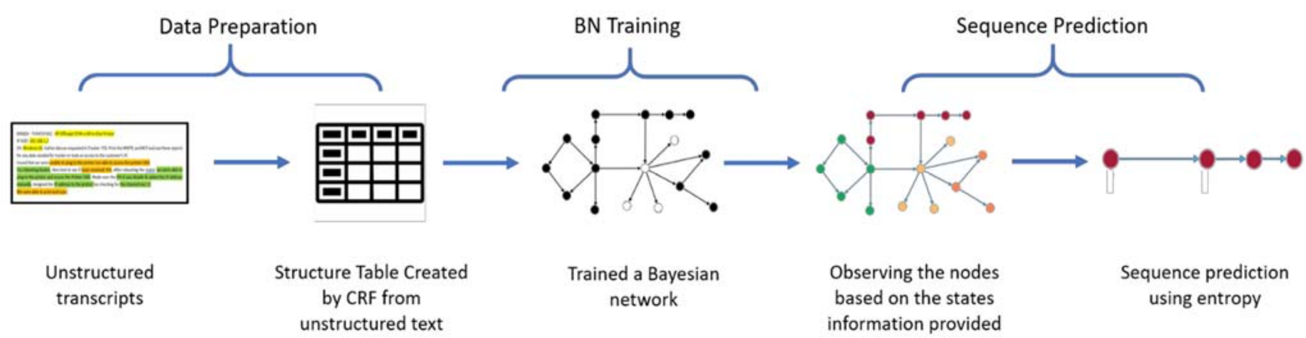
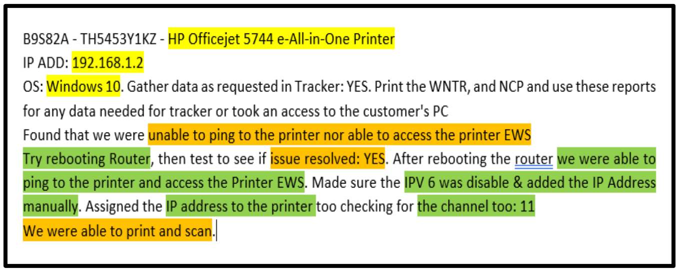


Fig 1

Below are the three modules for the approach:

1. **Data Preparation:**
   To prepare training data we structure the *customer support data* using Conditional Random Fields (CRF). Below (Fig 2.1) is the sample of unstructured text, transcript of a customer support call. The training data for

the CRF is prepared by manually tagging the unstructured text in BIO format. Once CRF is trained it can tag any given unstructured text in three categories i.e. state information, steps performed and junk. For example, in Fig 2.1 state information is marked in yellow and orange and steps performed are marked in green. Fig 2.2 is the output table from CRF, which will have all the information (state and steps performed) in a structured format.



B9S82A - TH5453Y1KZ - HP Officejet 5744 e-All-in-One Printer
IP ADD: 192.168.1.2
OS: Windows 10. Gather data as requested in Tracker: YES. Print the WNTR, and NCP and use these reports for any data needed for tracker or took an access to the customer's PC
Found that we were unable to ping to the printer nor able to access the printer EWS
Try rebooting Router, then test to see if issue resolved: YES. After rebooting the router we were able to ping to the printer and access the Printer EWS. Made sure the IPV 6 was disable & added the IP Address manually. Assigned the IP address to the printer too checking for the channel too: 11
We were able to print and scan.

Fig 2.1

Printer Model	Is this a first ti	Make of the I	OS	How is the	How is the	Is comput	Wireless is	Wireless/A	Are you ab	Able to pin
HP Officejet Pro 862	No	HP Pavillion5	Win 8.1	Wireless	Wireless	Yes	No	No	No	
HP Officejet 5740 e-A	Yes	Asus	Win 7	Ethernet	Wireless	Yes	No	No	No	No
HP ENVY 5640 e-All-i	Yes	HP 15 TS note	Win 10	Wireless	Wireless	Yes	No	No	No	No
HP ENVY 4520 All-in-	Yes	HP Pavilion G	Win 10	Wireless	Wireless	Yes	No	No	Yes	No
HP ENVY 7640 e-All-i	Yes	HP envy	Win 7	Wireless	Wireless	Yes	No	No	No	No
HP ENVY 7640 e-All-i	No	HP	Win 10	Wireless	Wireless	Yes	Yes	No	Yes	Yes
HP OfficeJet 3830 All	Yes	Toshiba	Win 10	Ethernet	Wireless	Yes	Yes	No	Yes	Yes
HP OfficeJet 4650 All	Yes	Asus	Win 8.1	Wireless	Wireless	Yes	unable to	unable to	No	No
HP ENVY 5660 e-All-i	Yes	HP Envy 700	Win 8.1	Ethernet	Wireless	Yes	Yes	No	No	No
HP Officejet Pro 861	No	Dell	Win 10	Wireless	Wireless	Yes	No	No	Yes	No
HP ENVY 4520 All-in-	No	Dell Desktop	Win 10	Wireless	Wireless	Yes	No	No	Yes	Yes
HP OfficeJet 3830 All	No	HP	Win 8.1	Wireless	Wireless	Yes	no	na	no	Yes , 4 pac
HP Officejet Pro 861	No	Acer	Win 7	Wireless	Wireless	Yes	Cant acces	cant check	Yes	Yes
HP OfficeJet 4650 All	Yes	HP 15-g019w	Win 10	Wireless	Wireless	Yes	No	No	No	Yes
HP Officejet Pro 861	No	Lenavo	Win 10	Wireless	Wireless	Yes	Unable to	Unable to	No	No

Fig 2.2

2. **Bayesian Network Training:**

A Bayesian Network is a probabilistic graphical model which represents the joint probability distribution over a set of random variables by assuming certain conditional independence assumptions among variables. The conditional independence assumption allows us an alternative parameterization of the joint distribution such that the number of parameters to be estimated are relatively smaller. Moreover, the inference is tractable using this alternative parameterization. The network consists of nodes which are random variables and an edge between two nodes indicates direct probabilistic dependency such as correlation or causality.

To train a Bayesian Network we use the training data prepared above in Fig 2.2. The columns of the table map to nodes in a Bayesian network. Using the standard structure learning algorithms of Probabilistic Graphical Models, we learn the directed edges between nodes and corresponding conditional probabilities. Fig 3 shows a Bayesian Network learned from the customer support data.

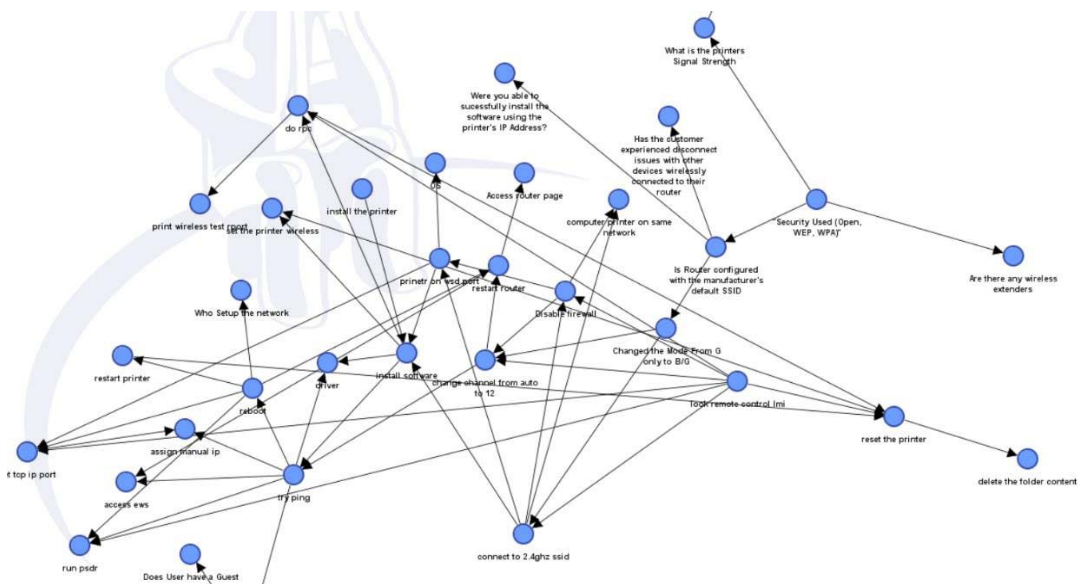


Fig 3

There are two types of nodes in above Bayesian Network:

- User State Nodes
- Agent Action Nodes

User State Nodes are the observable nodes which represent the state of the environmental variables i.e. printer, router, PC etc. For example, printer model, how it is connected to the network, firewall information, router model etc.
Agent Actions are the steps performed by customer support agents to solve the issue, which is predicted after observing all the user state nodes. The basic idea is that once we have learned the Bayesian Network graph which represents a joint probability distribution over set a of random variables mapped to the columns of Table, we can compute the marginal probabilities of all the action nodes present in the graph.

3. **Prediction of the Sequence:**
Fig 3 above represents a joint probability distribution over state and action random variables. To predict a sequence of actions we observe all the state random variables to arrive at the joint probability distribution over action random variables. From this joint probability distribution over action random variables, we can calculate the marginal probability of individual actions. These marginal probabilities when arranged according to minimum entropy will give us the sequence of action which corresponds to a fix for the observed state. We consider the sequence until the probability value is above a certain threshold.
For example, in Fig 4 the suggested sequence is 1-6-2-5-4, ordered acc to minimum entropy.

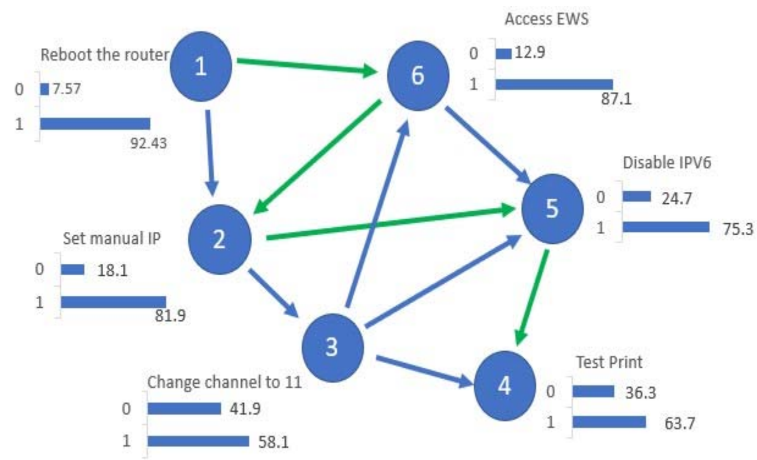


Fig 4

Evidence the solution works

The model was able to predict sequences with more than 80% confidence and was able to solve few run time issues. It is still in the validation phase.

References

- Printer troubleshooting using Bayesian network https://dl.acm.org/citation.cfm?id=352725
- Probabilistic Graphical Models Principles and Techniques By Daphne Koller and Nir Friedman

Disclosed by Gaurav Sood, Pavleen Kaur Brar and Raghu Anantharangachar, HP Inc.