General

<table>
<tr><td>ID[1]</td><td colspan="4"></td></tr>
<tr><td>Use case name</td><td colspan="4">VTrain recommendation engine</td></tr>
<tr><td>Context</td><td colspan="4">Education</td></tr>
<tr><td>Application domain</td><td colspan="4">On-premise systems</td></tr>
<tr><td>Status</td><td colspan="4">In operation</td></tr>
<tr><td rowspan="2">Contributor</td><td>Name</td><td>Affiliation</td><td colspan="2">Contact</td></tr>
<tr><td>Girish Palshikar
C. Anantaram</td><td>Tata Consultancy Services Ltd.</td><td colspan="2">c.anantaram@tcs.com</td></tr>
<tr><td>Scope[2]</td><td colspan="4"></td></tr>
<tr><td>Objective(s)</td><td colspan="4">Recommend a personalised list of “best” training courses to an employee, which will help him/her meet his/her career objectives.</td></tr>
<tr><td rowspan="2">Narrative</td><td>Short description (not more than 150 words)</td><td colspan="3">The vTrain system helps employees improve their skills by recommending appropriate training courses from a given list and historical data.</td></tr>
<tr><td>Complete description</td><td colspan="3">Continuous training is crucial for creating and maintaining the right skill-profile for the industrial organization’s workforce. There is a tremendous variety in the available trainings within an organization: technical, project management, quality, leadership, domain-specific, soft-skills etc. Hence it is important to assist the employee in choosing the best trainings, which perfectly suits him/her background, project needs and career goals. In this work, we focus on algorithms for training recommendation in an industrial setting. We formalize the problem of next training recommendation, taking into account the employee’s training and work history. We have developed several new unsupervised sequence mining algorithms to mine the past trainings data from the organization for arriving at personalized next training recommendation. Using the real-life data about trainings of 118587 employees over 5019 distinct trainings from a large multi-national IT organization, we show that these algorithms outperform several standard recommendation engine algorithms as well as those based on standard sequence mining algorithms.</td></tr>
<tr><td rowspan="2">Key performance indicators (KPIs)</td><td>ID</td><td>Name</td><td>Description</td><td>Reference to mentioned use case objectives</td></tr>
<tr><td>1</td><td>Prediction accuracy</td><td>Number of employees undertaking courses from the recommended list</td><td>Improve accuracy</td></tr>
<tr><td rowspan="4">AI features</td><td>Taks(s)</td><td colspan="3">Recommendation</td></tr>
<tr><td>Method(s)[3]</td><td colspan="3">Deep learning</td></tr>
<tr><td>Hardware[4]</td><td colspan="3">Windows</td></tr>
<tr><td>Terms and concepts used[5]</td><td colspan="3"></td></tr>
<tr><td>Challenges and issues</td><td colspan="4">Challenges: Need large amounts of training data; predicting human behaviour is tricky</td></tr>
</table>

Societal concerns	

Data (optional)

Data characteristics	
Description	
Source[6]	
Type[7]	
Volume (size)	
Velocity (e.g. real time)[8]	
Variety (multiple datasets)[9]	
Variability (rate of change)[10]	
Quality[11]	

Process scenario (optional)

Scenario conditions					
No.	Scenario name	Scenario description	Triggering event	Pre-condition[12]	Post-condition[13]

Training (optional)

Scenario name	Training				
Step No.	Event[14]	Name of process/Activity[15]	Primary actor	Description of process/activity	Requirement

Specification of training data[16]	

Evaluation (optional)

Scenario name	Evaluation				
Step No.	Event[17]	Name of process/Activity[18]	Primary actor	Description of process/activity	Requirement

Input of evaluation[19]	
Output of evaluation[20]	

Execution (optional)

Scenario name	Execution				
Step No.	Event[21]	Name of process/Activity[22]	Primary actor	Description of process/activity	Requirement

Input of Execution[23]	
Output of Execution[24]	

Retraining (optional)

Scenario name	Retraining				
Step No.	Event[25]	Name of process/Activity[26]	Primary actor	Description of process/activity	Requirement

Specification of retraining data[27]	

References

References						
No.	Type	Reference	Status	Impact on use case	Originator/organization	Link
1	Journal				Tata Consultancy Services Limited	R. Srivastava, G.K. Palshikar, S. Chaurasia, A. Dixit, *What's Next? A Recommendation System for Industrial Training*, accepted in **Data Science and Engineering** journal (Springer).
2	Conference				Tata Consultancy Services Limited	R. Srivastava, G.K. Palshikar, S. Chaurasia, *What's Next? A Recommendation System for Industrial Training*, **Proc. of Workshop on Human Capital Management**, held as part of International Conference on Data Management (ICDM 2017), New Orleans, USA, 18–21 November, 2017.
3	Conference				Tata Consultancy Services Limited	R. Srivastava, S. Hingmire, G. K. Palshikar, S. Chaurasia, A. Dixit, *CSRS: A Context and Sequence Aware Recommendation System*, **8th Meeting of the Forum for Information Retrieval Evaluation (FIRE 2016)**, 7 – 10 December 2016, Kolkata, India.

[1] Leave this cell blank.

[2] The scope defines the limits of the use case.

[3] AI method(s)/framework(s) used.

[4] Hardware system used.

[5] Terms and concepts listed here can be used to extend the work of WG 1 (AWI 22989 and AWI 23053) as necessary.

[6] Origin of data, which could be from instruments, IoT, web, surveys, commercial activity, or from simulations.

[7] Structured/unstructured Images, voices, text, gene sequences, and numerical. Composite: time-series, graph-structured

[8] The rate of flow at which the data is created, stored, analysed, or visualized.

[9] Data from a number of domains and a number of data types. The wider range of data formats, logical models, timescales, and semantics complicates the integration of the variety of data.

[10] Changes in data rate, format/structure, semantics, and/or quality.

[11] Completeness and accuracy of the data with respect to semantic content as well as syntactical of the data (such as presence of missing fields or incorrect values)

[12] Describe which condition(s) should have been met before this scenario happens.

[13] Describe which condition(s) should prevail after this scenario happens. The post-condition may also define "success" or "failure" conditions.

[14] The event that triggers the step. This might be completion of the previous event.

[15] Action verbs should be used when naming activity.

[16] Training data can be further specified.

[17] The event that triggers the step. This might be completion of the previous event.

[18] Action verbs should be used when naming activity.

[19] Specify input of evaluation.

[20] Specify output of evaluation.

[21] The event that triggers the step. This might be completion of the previous event.

[22] Action verbs should be used when naming activity.

[23] Specify input of evaluation.

[24] Specify output of evaluation.

[25] The event that triggers the step. This might be completion of the previous event.

[26] Action verbs should be used when naming activity.

[27] Retraining data can be further specified.