Semantic enrichment of spatio-temporal production data to determine lead times for manufacturing simulation – 2019 Winter Simulation Conference
Semantic enrichment of spatio-temporal production data to determine lead times for manufacturing simulation

Carina Mieth | TRUMPF Group, Ditzingen, Germany
1. Motivation
   • introduction
   • objective of this contribution

2. Fundamentals
   • real-time indoor localization systems (RTILS)
   • semantic enrichment
   • related work

3. Contribution
   • proposed approaches
   • comparison of the algorithms
   • sensor fusion concept

4. Conclusion
   • summary
   • future work
   • references
Introduction

Challenges with simulation's input data:

- simulation success relies on high-quality data (Wenzel et al. 2007)
- data gathering and processing is time-consuming
- poor data quality causes problems with credibility (Onggo et al. 2013)

This content had to be removed for online publication.
Introduction

Increasing dissemination of cyber-physical systems:

- valuable proposition of new data sources
- considered essential for digital twins (Yang et al. 2017; Shao and Kibira 2018; Srewil and Scherer 2017)
- the analysis of spatio-temporal data has the potential to discover hidden patterns that result in non-trivial insights (Nikitopoulos et al. 2018)
Flashback to:

RTILS framework

data model for sheet metal production

improvements of data quality

52nd CIRP Conference on Manufacturing Systems
Framework for the usage of data from real-time indoor localization systems to derive inputs for manufacturing simulation
Carina Mieth*, Anne Meyerb, Michael Henkeb

*TRUMPF Werkzeugmaschinen GmbH & Co. KG, Johann-Maus-Straße 2, 71254 Ditzingen, Germany
b TU Dortmund university, Chair of Enterprise Logistics, Leopold-Euler-Straße 5, Dortmund, Germany

Data Quality Dimensions

Motivation
Objective of this contribution

Each production order is tracked with the RTILS, and the process sequence and layout are known. Three algorithms are presented for the semantic enrichment, and semantic trajectories can be used for the calculation of lead time distributions.
Real-time Indoor Localization Systems

- here: based on ultra-wide band technology
  - bandwidth > 500 MHz
  - frequencies 3.1-10.6 GHz
  - TDOA / RTOF

- stationary devices:
  - satellites
  - industrial computer

- mobile devices:
  - markers
Semantic Enrichment

is the process of annotating spatio-temporal trajectories with meaningful context information (Arslan et al. 2018)

context information from...

semantic spatio-temporal trajectories

\((x, y, t, ID_{marker}, ID_{order}, C)\)

Own illustration taken from Mieth et al. 2019
Related Work

Previous approaches consider radio-frequency identification (indoor) or GPS data (outdoor):

- **Zhong et al. (2014)**: mined operating times from RFID-enabled worker data
- **Srewil and Scherer 2017**: enriched RFID-data with location context of construction sites
- **Rashid et al. 2017**: worker tracking at construction sites for training a hidden markov model for trajectory prediction which is used to calculate risk index for workers (location-based safety alerts)

Open Challenges

- **UWB-based RTILS production data ≠ RFID data**
  - **Yan et al. 2013**: techniques for semantic events inferred from raw GPS-like data should be developed
  - **Zheng 2015**: trajectory data mining applications: movements of people, transportation vehicles, animals and natural phenomena → manufacturing not mentioned!
Related Work

approaches to determine stay points in GPS-data

- Li et al. 2008: stay point detection algorithm → no semantics
- Palma et al. 2008: a clustering-based approach with adaptive thresholds, clusters are mapped on polygons
- Rocha et al. 2010: a direction-based spatio-temporal clustering method (frequent changes = POI)
- Alvares et al. 2010: trajectory are split whenever borders of areas are crossed → not robust

Graaff et al. 2016: use accuracy of the location samples, reductions in speed, changes in direction and projection of signals onto parcel polygons

There are no approaches for production environments

Difficulties: outdoor < indoor → proximity, adjacency, overlap signal interferences → inherent inaccuracy of the data (Richly 2018)
Proposed approaches

Online Semantic Annotation (real-time)
- based on the distance to points of interest (POI)
- based on areas of interest (AOI)

Offline Semantic Annotation (when the whole trajectory is known)
- trajectory segmentation as a classification problem (CP)
**Algorithm for Online Semantic Annotation → based on Points of Interest (POI)**

Idea: use $R_j(t, P_i) := \frac{1}{d_{o_j,P_i}(t)} + \sum_{f=1}^{F} \frac{1}{w_fd_{o_j,P_i(t-f)}}$ as discriminator for the allocation of a measurement

- $P_i$ refers to a location $(x_i, y_i)$ on the shopfloor with semantic meaning (e.g. machine or workplace)
- $d_{o_j,P_i}(t)$ euclidean distance between production order $o_j$ and POI $P_i$
- pseudo probability: relating each current rating to the sum of all ratings at the time

Points of interest $P_1$ and $P_2$ with the trajectory of a production order $o_j$
Algorithm for Online Semantic Annotation 
→ based on Areas of Interest (AOI)

Idea: allocate measurements to AOIs, if the majority of the last F measurements are inside

- \( A_i \) refers to an arbitrarily shaped area on the shopfloor that has a semantic meaning
  (e.g. work area around a machine or a storage)

\[
PIA((x,y),A_i) = \begin{cases} 
1 & \text{inside} \\
0 & \text{on border} \\
-1 & \text{outside} 
\end{cases}
\]
the discriminator is the point-in-area (PIA) test
that checks if the measured point \((x,y)\) is in \( A_i \)

- if the sum of all values returned by the PIA-test inside the window size \( F \) is greater or equal zero → location change
Pseudocode can be found in the paper ;)

<table>
<thead>
<tr>
<th>Algorithm 1: Pseudocode for the allocation of measurements to points of interest (POI).</th>
</tr>
</thead>
</table>
| **Input**: spatio-temporal trajectories $S_j$ for each production order $o_j$, window size $F$, weighting factors $w_1, \ldots, w_F$, points of interest $P_i \forall i = 1 \ldots N$  
**Output**: semantic trajectory $S_{o_j}$, event log  
For each $t = 1 \ldots T$ do  
  Check plausibility constraints for measurement $s_j(t)$  
  For each production order $o_j$ do  
    For each point of interest $P_i$ do  
      If $t \leq F$ then  
        $R_j(t, P_i) := \frac{1}{w_{P_i}}$;  
        // initialization of first ratings  
      Else  
        $R_j(t, P_i) := \frac{1}{w_{P_i}} + \sum_{f=1}^{F-1} \frac{1}{w_{P_i,f}(t-f)}$;  
        // rating function  
      Assign $P_i$ with $\max(R_j(t, P_i))$ to $S_{o_j}(t)$;  
      If $P_i(t) \neq P_i(t-1)$ then  
        Save timestamp to eventlog;  
      // location has changed  
    Save $S_{o_j}(t)$, event log.  

Algorithm 2: Pseudocode for the allocation of measurements to areas of interests (AOI). |
| **Input**: trajectories $S_j$ for each production order $o_j$, window size $F$, disjoint areas of interest $A_i \forall i = 1 \ldots N$  
**Output**: semantic trajectory $S_{o_j}$, event log  
For each $t = 1 \ldots T$ do  
  Check plausibility constraints for measurement $s_j(t)$  
  For each production order $o_j$ do  
    For each area of interest $A_i$ do  
      If $(t > F)$ then  
        If $\text{PIA}(S_j(t), A_i) = 1$ then  
          If $A_i(t) \neq A_i(t-1)$ then  
            if $\sum_{j=1}^{F} \text{PIA}(S_j(t-f), A_i) \geq 0$ then  
              Assign event at $t - \left[ \frac{F}{2} \right]$ to eventlog;  
              // location has changed  
              Assign $A_i$ to $S_{o_j}(t)$;  
              // assign measurement  
              Assign $A_i$ to $S_{o_j}(t-1) \ldots S_{o_j}(t - \left[ \frac{F}{2} \right])$;  
              // update previous ones  
          Else  
            Assign previously identified area of interest $A_i$ at $(t-1)$ to $S_{o_j}(t)$;  
        Else  
          Assign previously identified area of interest $A_i$ at $(t-1)$ to $S_{o_j}(t)$;  
      Else  
        If $\text{PIA}(S_j(t), A_i) = 0$ then  
          Assign previously identified area of interest $A_i$ at $(t-1)$ to $S_{o_j}(t)$;  
      Else  
        If $\text{PIA}(S_j(t), A_i) = 1$ then  
          Assign area of interest $A_i$ to $S_{o_j}(t)$;  
          // initialization  
      // location has changed  
    Save $S_{o_j}(t)$, event log.  

Choosing the “right” window size $F$  
Definition of POIs is easy  
AOIs sometimes tricky  
Both definitions influence the results
Classification Problem (CP)

Idea: each measurement is assigned to a process class

- decision variable $k_t$ for each position measurement $s_j(t) \rightarrow$ process class $V_t \in V$
- two possible error functions:
  \[ e_j(t) = \begin{cases} 
  0 & \text{PIA}^*(t) = k_t \\
  1 & \text{PIA}^*(t) \neq k_t 
  \end{cases} \text{ or } e_j(t) = 1 - p(t) \]
- transition function
  \[ u(k_t, k_{t+1}) = \begin{cases} 
  0 & \text{no change detected: } k_t = k_{t+1} \\
  1 & \text{change detected: } k_t \neq k_{t+1} 
  \end{cases} \]

Minimize
\[ \sum_{t=1}^{T} e_j(t) \] assignment error should be minimized w.r.t.

1st constraint
\[ (k_t, k_{t+1}) \in Arc^* \] ensures that the sequence of processes is not violated

2nd constraint
\[ \sum_{t=1}^{T-1} u(k_t, k_{t+1}) = 2 \cdot |V^*| - 1 \] is for the segmentation of trajectories in the number of allowed segments
Comparison of the Algorithms

Cycle-free graph is guaranteed by design of the constraints
- can handle adjacency & overlapping
- rework can be identified → number of classes can be adjusted
Sensor Fusion Concept

combining knowledge from different sensor types

With existing approaches it is only possible to detect time shares between events that correspond to a change in location.

Sometimes it is also of interest to split these times shares into smaller parts.

Handling events can separate the trajectory into suitable segments (e.g. accelerometer or magnetometer measurements).
Summary

- introduction of semantic enrichment
- presentation of three approaches for semantic enrichment (online and offline)
- comparison of these algorithms
- presentation of a sensor fusion concept

Future work

- use real-world data set for validation of proposed algorithms
- derive other simulation inputs from RTILS data
References


References


References


How to calculate lead time distributions from semantic trajectories

\[(x, y, t, ID_{marker}, ID_{order}, C)\]
Figure 2: Manufacturing domain ontology for the semantic enrichment of trajectory from RTILS.
Graph of the production process

For a more reliable assignment of the measurements to the points or areas of interest, technological restrictions will be considered in the form of logical operation sequences. All possible operation sequences within the production system are part of a directed graph $G = (V, Arc)$. Each node $V_i$ in the node set $V$ represents an operation that is performed at the point or area of interest $P_i$ or $A_i$, respectively. The arc set $Arc$ contains all directed arcs $(V_a, V_b)$ that always contain two logically sequentially executable production operations. A path $G_j \in G$ corresponds to a sequence of operations that a production order $o_j$ passes through. Let $G^* = (V^*, Arc^*)$ be an extension of $G_j$ that results by adding edges that join each vertex of the path to itself (loops).