

The Trajectory Collection File Format

The Song Lab
Department of Computer Science
New Mexico State University

October 29, 2012

Contents

1	Introduction	2
1.1	Radix	2
1.2	Missing values	2
2	TCF Version 1 (OBSOLETE)	3
3	TCF Version 2	7
3.1	Version 2.1	9
3.2	Version 2.2	10
4	TCF Version 3 (UNDER DESIGN)	11
5	Possible new features	13
5.1	Steady state data	13
5.2	Unbalanced experimental design	13

Revision History

- September 14, 2010. Added TCF version 2.2
- November 9, 2008.
 - Changed the file format name from Trajectory File Format (TFF) to Trajectory Collection File (TCF) format.
 - Added sections on missing values, steady state data, perturbed time course, and unbalanced experimental design.
 - Proposed new additions to the TCF in the added sections.
 - Re-organized the version definitions into appendices.
- April 6, 2008. The TCF version 2.1 is added, which is a minor change from version 2.
- October 6, 2007. TCF Version 3 design started.

- July 10, 2007. Document created.

1 Introduction

We define a file format, called trajectory collection file (TCF), for storing the trajectories of a dynamical system. This file format can specify multiple trajectories of various lengths in a single file as well as store additional information of the variables or nodes in a dynamical system. TCF format can be used to save trajectories from both discrete and continuous valued dynamical systems.

1.1 Radix

TCF is versatile to use beyond Boolean networks. A radix is specified for each node. The radix should be 2 if a node is binary. When the radix is 0, it means the values of a node is continuous.

1.2 Missing values

Missing values are observations not available or removed in a time course. This issue is addressed in TCF version 2. A value of -1 indicates a missing value. This can pose a problem when the radix of a node is 0 (continuous in value).

2 TCF Version 1 (OBSOLETE)

The version 1 TCF format is defined as follows:

```
TRAJECTORY_VER1
<# OF TRAJECTORIES> <# OF NODES>
<RADIX OF EACH NODE>

<# OF TIME POINTS IN TRAJECTORY 1>
<TRAJECTORY 1>

<# OF TIME POINTS IN TRAJECTORY 2>
<TRAJECTORY 2>

<# OF TIME POINTS IN TRAJECTORY 3>
<TRAJECTORY 3>

...

<# OF TIME POINTS IN LAST TRAJECTORY>
<LAST TRAJECTORY>
```

Notes:

1. The first line is an ID to indicate the file format definition.
2. All other entries are separated by either space or new line.

Example 1:

TRAJECTORY_VER1			
4	4		
2	2	2	2
3			
0	0	1	1
1	0	0	1
1	0	0	0
7			
1	0	1	0
0	0	0	1
1	1	1	0
1	0	1	0
0	0	0	1
1	1	1	0
1	0	1	0
4			
0	1	0	1
1	1	1	1
0	0	0	0
1	1	0	0
4			
0	1	1	1
1	0	0	0
0	1	1	1
1	0	0	0

Example 2 – Trajectories of a 3-node Boolean network.

TRAJECTORY_VER1			
4	3		
2	2	2	
5			
0	1	0	
1	0	1	
1	1	0	
1	0	1	
1	1	0	
5			
0	0	0	
0	0	1	
1	0	0	
0	0	1	
1	0	0	
5			
1	1	1	
1	1	1	
1	1	1	
1	1	1	
1	1	1	
5			
0	1	1	
1	0	1	
1	1	0	
1	0	1	
1	1	0	

Example 3 – Trajectories of a 7-node Boolean network.

TRAJECTORY_VER1							
8	7						
2	2	2	2	2	2	2	2
11							
0	0	0	0	0	1	0	
1	0	0	1	1	0	0	
0	0	0	1	1	0	0	
1	0	0	1	1	0	0	
0	0	0	1	1	0	0	
1	0	0	1	1	0	0	
0	0	0	1	1	0	0	
1	0	0	1	1	0	0	
0	0	0	1	1	0	0	
1	0	0	1	1	0	0	
0	0	0	1	1	0	0	
11							
0	0	0	0	0	1	1	
1	1	0	0	1	0	0	

```

0 0 1 1 1 0 1
1 1 0 0 1 1 0
0 0 1 1 1 0 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
11
0 0 0 0 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
11
0 0 0 0 1 0 1
1 1 0 0 0 0 0
0 0 1 1 1 0 1
1 1 0 0 1 1 0
0 0 1 1 1 0 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
11
0 0 0 0 1 1 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
11
0 0 0 0 1 1 1
1 1 0 0 1 0 0
0 0 1 1 1 0 1
1 1 0 0 1 1 0
0 0 1 1 1 0 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
1 0 1 0 1 1 0

```

```

0 0 1 0 1 1 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
11
0 0 0 1 0 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
1 0 0 1 1 0 0
0 0 0 1 1 0 0
11
0 0 0 1 0 0 1
1 1 0 0 0 0 0
0 0 1 1 1 0 1
1 1 0 0 1 1 0
0 0 1 1 1 0 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
1 0 1 0 1 1 0
0 0 1 0 1 1 0
0 0 1 0 1 1 0

```

3 TCF Version 2

Trajectory file version 2 is defined as follows:

```
TRAJECTORY_VER2
<# OF TRAJECTORIES> <# OF INTERNAL NODES> <# OF EXTERNAL NODES>
<RADIX OF EACH NODE>
<NAMES OF INTERNAL NODES>
<NAMES OF EXTERNAL NODES>

<# OF TIME POINTS IN TRAJECTORY 1>
<TRAJECTORY 1>

<# OF TIME POINTS IN TRAJECTORY 2>
<TRAJECTORY 2>

<# OF TIME POINTS IN TRAJECTORY 3>
<TRAJECTORY 3>

...

<# OF TIME POINTS IN LAST TRAJECTORY>
<LAST TRAJECTORY>
```

Notes:

1. An internal node can have from no to any number of parents.
2. An external node must not have a parent.
3. The names of all internal nodes are listed on line 4.
4. The names of all external nodes are listed on line 5.
5. Each line of a trajectory includes first the values of all internal nodes, followed by the values of all external nodes on the same line.

Example 4.

```
TRAJECTORY_VER2
4 3 1
2 2 2 2
GENE1 GENE2 GENE3
LIGHT

3
0 0 1 1
1 0 0 1
1 0 0 0

7
1 0 1 0
0 0 0 1
1 1 1 0
1 0 1 0
```

0	0	0	1
1	1	1	0
1	0	1	0
4			
0	1	0	1
1	1	1	1
0	0	0	0
1	1	0	0
4			
0	1	1	1
1	0	0	0
0	1	1	1
1	0	0	0

3.1 Version 2.1

Two properties are added for each node

- Can a node be parent or not? coded as BP=Yes and NBP=No
- Can a node have parent or not? coded as HP=Yes and NHP=No

Example 5.

```
TRAJECTORY_VER2.1
```

```
4 3 1
```

```
2 2 2 2
```

```
GENE1 GENE2 GENE3
```

```
LIGHT
```

```
Yes Yes No Yes
```

```
Yes No Yes No
```

```
3
```

```
0 0 1 1
```

```
1 0 0 1
```

```
1 0 0 0
```

```
7
```

```
1 0 1 0
```

```
0 0 0 1
```

```
1 1 1 0
```

```
1 0 1 0
```

```
0 0 0 1
```

```
1 1 1 0
```

```
1 0 1 0
```

```
4
```

```
0 1 0 1
```

```
1 1 1 1
```

```
0 0 0 0
```

```
1 1 0 0
```

```
4
```

```
0 1 1 1
```

```
1 0 0 0
```

```
0 1 1 1
```

```
1 0 0 0
```

3.2 Version 2.2

When a particular gene is perturbed artificially in a time course, the time course should not be used for determining the parents of the gene. This issue came up during the DREAM3 competition in September 2008. The GLN and DDS code was tailored for DREAM3 by specifying a command line option to deal with the situation. Now this is made general in the TCF file without the command line option.

Each trajectory now has its own meta-data information to indicate which nodes are perturbed externally and thus the network inference algorithm should consider excluding them from using this exact trajectory to selecting their own parents.

Example 6.

```
TRAJECTORY_VER2.2
4 3 1
2 2 2 2

GENE1 GENE2 GENE3
LIGHT
Yes Yes No Yes
Yes No Yes No

1 exclude 1
3
0 0 1 1
1 0 0 1
1 0 0 0

0 exclude
7
1 0 1 0
0 0 0 1
1 1 1 0
1 0 1 0
0 0 0 1
1 1 1 0
1 0 1 0

2 exclude 1 2
4
0 1 0 1
1 1 1 1
0 0 0 0
1 1 0 0

0 exclude
4
0 1 1 1
1 0 0 0
0 1 1 1
1 0 0 0
```

In this version, one extra line is added before each trajectory. Before keyword *exclude* is the total number of nodes perturbed and thus should be excluded as being a child node. After the keyword are the actual node ids (1-based) of those excluded nodes.

4 TCF Version 3 (UNDER DESIGN)

New features:

1. There is a node section and a trajectory section
2. It is relatively straightforward to expend node properties in the node section
3. All fields are tab separated
4. Time is included in each trajectory

```
TRAJECTORY_VER3
<# OF NODES>
<NAME OF NODE 1> <NODE RADIX> INTERNAL/EXTERNAL BP/NBP HP/NHP <ADDITIONAL ATTRIBUTES CAN BE ADDED>
<NAME OF NODE 2> <NODE RADIX> INTERNAL/EXTERNAL BP/NBP HP/NHP <ADDITIONAL ATTRIBUTES CAN BE ADDED>
...
<NAME OF THE LAST NODE> <NODE RADIX> INTERNAL/EXTERNAL BP/NBP HP/NHP <ADDITIONAL ATTRIBUTES CAN BE ADDED>

<#TRAJECTORIES>

<#TIME POINTS IN TRAJECTORY 1> <TIME UNIT> <#SAMPLES AT EACH TIME POINT>

<TIME POINT 1>
<SAMPLE 1 OF NODE 1> <SAMPLE 1 OF NODE 2> ...
<SAMPLE 2 OF NODE 1> <SAMPLE 2 OF NODE 2> ...
...

<TIME POINT 2>
<SAMPLE 1 OF NODE 1> <SAMPLE 1 OF NODE 2> ...
<SAMPLE 2 OF NODE 1> <SAMPLE 2 OF NODE 2> ...
...
...

<#TIME POINTS IN TRAJECTORY 2> <TIME UNIT> <#SAMPLES AT EACH TIME POINT>

<TIME POINT 1>
<SAMPLE 1 OF NODE 1> <SAMPLE 1 OF NODE 2> ...
<SAMPLE 2 OF NODE 1> <SAMPLE 2 OF NODE 2> ...
...

<TIME POINT 2>
<SAMPLE 1 OF NODE 1> <SAMPLE 1 OF NODE 2> ...
<SAMPLE 2 OF NODE 1> <SAMPLE 2 OF NODE 2> ...
...
...

.
.
.

<#TIME POINTS IN THE LAST TRAJECTORY> <TIME UNIT> <#SAMPLES AT EACH TIME POINT>

<TIME POINT 1>
<SAMPLE 1 OF NODE 1> <SAMPLE 1 OF NODE 2> ...
<SAMPLE 2 OF NODE 1> <SAMPLE 2 OF NODE 2> ...
...

<TIME POINT 2>
<SAMPLE 1 OF NODE 1> <SAMPLE 1 OF NODE 2> ...
<SAMPLE 2 OF NODE 1> <SAMPLE 2 OF NODE 2> ...
...
...
```

Example 5: Trajectories from a hybrid dynamic system which has three Boolean as well as one continuous variables. Please notice that there are missing values in the 3rd trajectory.

TRAJECTORY_VER3					
4					
GENE1	3	i	BP	HP	
GENE2	2	i	BP	HP	
GENE3	0	i	BP	HP	
LIGHT	2	e	BP	NHP	
3					
3 HOUR 2					
0.0					
0	0	1.3	1		
1	0	1.7	1		
0.5					
1	0	-1.8	0		
0	0	-0.9	1		
1.0					
0	1	-0.2	0		
2	0	0.4	0		
5 MINUTE 3					
0.0					
1	1	3.1	1		
2	1	1.9	0		
1	0	4.0	1		
1.2					
2	0	-0.5	1		
2		-0.5	0		
2	1	-0.5	0		
2.0					
2	1	8.1	0		
2	0	2.1	1		
0	1	9.2			
3.5					
2	0	3.1	0		
1	0	3.1	0		
	1	3.1	0		
4.0					
0	1	-3.6	1		
2	1	-2.8	1		
2	0	-3.2	1		
5.7					
0	1	2.8	1		
2	0	2.1	0		
1	1	2.5	0		
4 DAY 1					
0.0					
0	1	-2.08	1		
1.0					
2	0	11.3	1		
2.0					
1	1	0.05	0		
4.0					
1	0	8.0	0		

5 Possible new features

5.1 Steady state data

It might be necessary to include a field for a time course in the file format, corresponding to whether a time course is from a steady state or not.

<#TIME POINTS> <TIME UNIT> <#SAMPLES PER TIME POINT> <STEADY STATE FLAG> <LIST OF PERTURBED VARIABLES>
--

5.2 Unbalanced experimental design

The number of sample at each time point might be different in a time course. This can be handled by having a field to specify the number of samples at every time point, instead of doing it in a field for time course. E.g.,

<TIME POINT 1> <# SAMPLES> <SAMPLE 1 OF NODE 1> <SAMPLE 1 OF NODE 2> ... <SAMPLE 2 OF NODE 1> <SAMPLE 2 OF NODE 2> <TIME POINT 2> <# SAMPLES> <SAMPLE 1 OF NODE 1> <SAMPLE 1 OF NODE 2> ... <SAMPLE 2 OF NODE 1> <SAMPLE 2 OF NODE 2>
--