人工知能学会 インタラクティブ
情報アクセスと可視化マイニング研究会(第17回)
SIG-AM-17-03

軌跡の直接操作に基づく時系列データの視覚的分析インタフェースの提案

Proposal of Visual Analytics Interface for Time Series Data based on Direct Manipulation of Trajectory

高見 玲 [1]　　高間 康史 [1]
Rei Takami[1]　Yasufumi Takama[1]

[1] 首都大学東京 システムデザイン学部
[1] Faculty of System Design, Tokyo Metropolitan University

Abstract: Recently, various temporal data have been collected in many fields, and visual analytics interface is expected to be useful for utilizing such data. However, when temporal data is visualized using animation, collision would occur between movement of time series data itself and movement caused by interaction with users. This paper focuses on trajectory, which can handle temporal and spatial changes uniformly, and proposes a visual analytics interface based on it. This paper also shows a use case of applying prototype interface to time series data.

1　はじめに

情報技術の普及に伴い, 近年, センサや SNS のストリームデータを代表とする時系列的特徴を持つデータが蓄積されるようになってきている．これらのデータを活用するために, クラスタリングやトレンド・外れ値の発見などのデータマイニング技術の活用が期待される．しかし，それらの技術を適用する前段階として, 分析者自身によるデータの傾向や因果関係の解釈や, 知識形成が重要である．この作業を支援するために, 対象データの可視化に基づく視覚的分析システムの研究が行われている [1].

視覚的分析において, 可視化オブジェクトの直接操作とフィードバックにより, データに対する理解を促進し, 知識形成を支援する手法が研究されてきた [2]．その中でも, 背後に存在するモデルやパラメータを意識せず, 可視化されたオブジェクトの直接操作によりモデルパラメータを修正し, 知識形成を促進する Semantic Interaction が提案されている [3]．しかし, 時系列データに対してこの様な直接操作的アプローチを適用する場合, データ自身の変化とユーザ操作による変化の衝突や, 異なる可視化手法間のトレードオフの存在などが問題点として想定される．衝突の例として, ズーム操作における, 空間もしくは時間のいずれを対象としているかの解釈の困難性が挙げられる．トレードオフの例として, 時系列データの可視化手法としてアニメーションや Small Multiple[4] が一般的だが，前者はデータの時間的変化を発見しやすい半面, 探索時間の増大や見落としが生じる．一方で, 後者では探索の時間と正答率に関しては有効であるが, スケーラビリティの問題がある．

上述の問題に対処するために, 本稿ではデータの時間的変化を空間上に表現することで, 操作との衝突を回避できる軌跡を用いた可視化に着目する．そして, 軌跡の直接操作によりデータの時空間的変化の特徴把握を促進し, 知識形成を支援する視覚的分析インタフェースを提案する．また, 提案インタフェースをプロトタイプとして実装し, 時系列データに適用した事例を示す.

2　関連研究

2.1　視覚的分析

一般に, データの分析者であるドメイン専門家は分析対象については豊富な知識を持つが, 必ずしもデータ分析手法の専門家ではない．そのため, クラスタリングや次元削減アルゴリズム等の背後に存在するパラメータ調整の困難性が指摘されている [5]．この問題に対処するため, データを可視化して提示することで, ドメイン知識による解釈を行い, データに関する洞察を獲得するプロセスを支援する視覚的分析が研究されている [1]．視覚的分析のモデル化に関して, Sacha らはデータ探索, 仮説検証, 知識形成の 3 つのループからなるモデルを提案している [6]．分析者はデータ探索ループで可視化されたデータの特徴的な視覚的傾向に着目し, 仮説検証ループで自身の有する知識に基づき視覚的傾向から仮説を検証, 形成する．そして, 知識形成ループで仮

説の検証と積み上げによりデータに関する知識を形成する.

また, 視覚的分析におけるインタラクションのモデル化の例として, Endert らは可視化オブジェクト自体への直接操作, フィードバックに基づく非明示的な仮説形成ループ (Human is the loop) を提案している [14]. このモデルでは, 直接操作の原則に従いインタラクションは逐次的 (例えば, 検索結果などが即時フィードバックされる), 高速, かつやり直し可能である必要がある [7]. データオブジェクト自体への直接操作を行う手法として, Semantic Interaction のコンセプトが提案されている [3]．このコンセプトでは，システム側が可視化オブジェクトに対するユーザの操作内容を解釈し, その結果に基づき可視化の背後に存在する次元削減やクラスタリングアルゴリズムのモデル構成やパラメータを調整する．調整したモデル・パラメータを用いてデータを処理した操作結果はユーザに即時フィードバックされる. これによって, 分析者がモデルに関する知識を有しない場合でも, データの操作, 可視化から知識形成へのシームレスな移行が可能になるとしている.

2.2 時系列データの分析と可視化

時系列データの主な可視化手法として, 時間軸を含む静的可視化のほかに, 変化を移動として表現するアニメーション, 時系列に従って複数のグラフを表示する Small Multiples, 時間的変化を 1 つのグラフに重ねて表示する軌跡 (Trace) などが存在する [4]. アニメーションを用いて時系列データの分析を行う際に, ユーザはアニメーションで全体の傾向を掴んだ後に, 重要だと判断した時点に対して UI を用いたインタラクティブな探索を行うことが示されている [8]. また, 視覚的分析におけるユーザの認知特性に関して, インタラクティブ性は探索速度や正確さを失うことなくユーザエクスペリエンスを向上させることが示されている [9].

軌跡による時系列データの可視化例として, Kondo, Collins はデータの変化軌跡をなぞる操作に基づく軌跡上の各時点への直感的な時間的ナビゲーションを提案している [10]. この手法によって, スライダなどの間接的なユーザインタフェースを使用せず, オブジェクト中心の直接操作による時系列データの探索が可能になるとしている. また, 散布図に適用した場合には探索速度と正確性について既存手法に対する有効性が示されている [10]. また, Bach らは, 時系列データの変化パターンの把握や, 視覚的概要の提示を行う手段として軌跡に着目し, データの各時点における自己類似性を基準とした軌跡の描画手法を提案している [11]. これらの既存研究では, 時系列データの変化特性の概要把握を支援するために軌跡を用いているが, より詳細なデータ分析やデータマイニングへの応用に関しては考慮されていない.

3 時系列データの可視化における問題点

時系列データに対して, 直接操作による視覚的分析を導入する際に, 以下のような問題点が想定される.

3.1 時間軸上でのインタラクションの衝突

データの可視化手法にアニメーションを採用した場合には, 視覚的分析におけるフィードバックに伴うオブジェクト配置の変化と, アニメーションの再生に伴うオブジェクト配置の変化は共に時間軸上で発生する. この様に, ユーザが行うインタラクションに関する時間軸と, 時系列データが内包する時間軸の衝突が生じる. これに伴い, 操作結果とデータ本来の時系列変化の混同が生じ, ユーザの混乱の原因となりうる.

3.2 時空間軸間でのインタラクションの衝突

時系列データを分析する際にユーザが行う操作は, オブジェクトの配置変更など空間に対し行われる他, 再生速度の変更など, 時間軸に対しても行われる. その結果, 操作適用対象に関して時間および空間軸間の衝突が生じる. 例えば, 時系列データに対して詳細を確認するために拡大（ズームアップ）操作を行った場合, 空間的に拡大表示したいのか, 再生速度を遅くしたいのか, 二通りの解釈がありうる. 採用された解釈により, 望まない変更が行われた場合には分析の妨げとなりうる.

また, 時系列データの変動をアニメーションの様に時間軸の変動として描画するモード, 折れ線グラフの様に空間的配置により描画するモードを用意し, 相互に切り替え可能なシステムも考えられる. このとき, モード間の遷移をユーザの操作を解釈により暗黙的に行う場合には, ユーザは自身が現在どちらのモードに存在するかを判断しづらくなる. 一方で, 明示的に移行する場合, 必要なインタラクション数が増加し, ユーザビリティ低下の原因となりうる.

この他にも, Yi らのユーザ要求に基づくインタラクション分類 [12] の内の以下においても, 時空間軸間でのインタラクションの衝突が発生すると考える.

- Select: 特定対象のマーク
- Abstract/Elaborate: 可視化の詳細度変更
- Filter: 条件に合致するオブジェクトを表示
- Connect: 関連アイテムを表示

3.3 可視化手法間のトレードオフ

時系列データのアニメーションを用いた可視化は, 聴衆にストーリーを理解させるプレゼンテーションのように変化を発見, 理解する場面では有効といえる. しかし, データの時空間的な変化傾向の把握に用いた場合, アニメーションの再生/静止を繰り返しによる認知的負荷の増大や, 探索の正確性や速度の低下が問題点として報告されている [4]. 一方で, 軌跡 (Trace) や Small Multiple は探索の正確性に関してはアニメーションと比べて有効である反面, 前者は多数のオブジェクトが存在する場合に重なりにより探索が困難となる. また, 後者の場合には占有する画面領域が広くなり, 探索時間や認知的負荷が増大する問題点が指摘されている [4].

このように, 時系列データに対して一意に最適な可視化手法は存在しないため, 分析目的や対象に応じて複数の手法を組み合わせることが有効だといえる.

4 提案手法

前述の背景と問題点を踏まえ, 本稿では, 時系列データを二次元散布図および軌跡として可視化し, 直接的操作による視覚的分析を通じた知識形成を支援する可視化インタフェースを提案する. 提案インタフェースが備えるべき要件として, 以下に着目する.

1. 従来の視覚的分析手法との互換性 (時空間的変化特性の把握)
2. 時間軸上でのインタラクションの衝突 (3.1 節) の回避
3. 時空間軸間でのインタラクションの衝突 (3.2 節) の回避
4. 可視化手法間のトレードオフ (3.3 節) への対応

1 点目については, 散布図, 軌跡による可視化や, 詳細ビューといった，可視化手法として一般によく用いられるものを採用する. 2 点目は, 軌跡を用いて, 時空間的な変化を空間上に表現することで対応する. 3 点目について, 軌跡表現により両軸への操作を統一することで対応する. 加えて, 一部の明示的に分離すべき操作に関しては探索モードの導入により対応する. 4 点目に関しては, 探索モードの切り替えにより各可視化手法の利点を生かすことで対応する.

4.1 対象データ

分析対象とする時系列データの種類として, 以下を想定する.

- 数値的データ: ヒストグラムや円グラフなど. 汎用的な可視化手法を適用可能であり, 解釈が容易.
- 集合/関係的データ: テキストストリームやネットワーク. そのまま数値的に表現できないため, 類似度関数などの距離メトリクスを用いて可視化される [2].
- 複合的データ: 複数の数値的データや集合/関係的データから構成されるデータ.

集合/関係的データや複合的データを対象とした場合でも, テキスト解析や次元削減などのデータマイニング技術を適用することで平面などの低次元空間上の座標として描画可能である. そのため, 提案インタフェースでは, 可視化手法として散布図を採用する. また, 座標空間上の近接性が類似性を示すことや, モデルパラメータの調整や値の推定に関するタスクについて, 長さや位置, 距離エンコーディングが面積や濃度と比較してタスクの正確性や達成速度に関して有効であることが報告されている [13]. そのため, 散布図はデータの時空間的特性の把握や, 視覚的傾向に基づく知識形成の点でも適していると考える.

4.2 軌跡を用いた可視化とインタラクション

提案インタフェースでは, 時系列データの散布図による可視化に加えて, データの時空間的な変化を軌跡として表現する. これにより, 時間的変化を空間上のみで表現可能となり, ユーザとのインタラクションに伴う変化との衝突を防げると考える. また, 時系列データを二次元平面上で扱えるため, 時間および空間軸に対する時間的操作を統一できる. そのため, Select, Filter 操作などについてインタラクションの衝突を回避する効果も期待できる. 時系列データの軌跡を用いた可視化の例を図 1 に示す. 軌跡は対象データの各時点における属性値の二次元空間座標間を結ぶ曲線により描画する. このとき, 各点間の補完に cubic Catmull-Rom Spline 曲線 [15] を用いる.

軌跡は, 散布図上のオブジェクトへのマウスオーバー操作に基づき描画される. また, 軌跡を構成する各点のマウスオーバーにより, その時点での散布図が図 2(2) のように薄く重畳表示される. 時間的な文脈を失わずに, 他時点データ分布を確認できる. その状態でクリック操作を行うことで, 図 2(3) のようにその時点への時間的遷移が実行される.

また, 軌跡のパスをクリックすることで, 軌跡の選択 (画面上への固定) が行える. 選択された軌跡は赤色でハイライトされる. これによって, データの時系列的変化による見落としを軽減できると考える. また, 図 3 に示すように, データが線形的に増加した場合には軌跡は

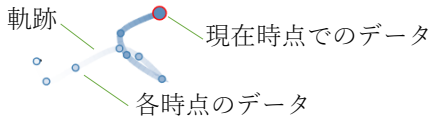


図 1: 軌跡による可視化

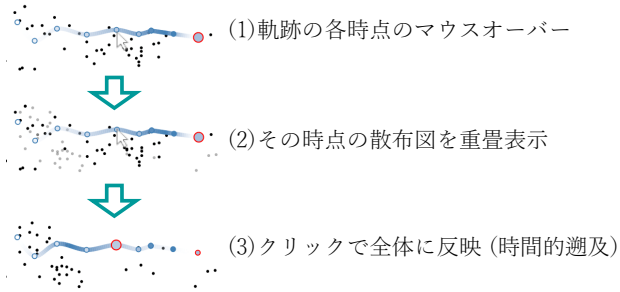


図 2: 軌跡操作に基づく時間的ナビゲーション

直線として現れる．水平あるいは垂直軸のどちらかで周期的に変化した場合には, その軸において往復するような軌跡が, 外れ値を含む場合には広域に渡り急峻な変化をする軌跡が描かれる．このように, 軌跡の形状はデータの時間的変化に対応し, データの視覚的概要の提示手段として有用であるといえる．

4.3 探索モード

アニメーションは時系列データの変化の発見や理解について有効だが, 操作数増加などの原因により探索に関しては不向きである．一方で, 軌跡や Small Multiple による可視化は探索効率と正確性に関しては有効だが, スケーラビリティの問題が存在する．

本稿では両者の利点を活かすために, アニメーションと静止の二つの探索モードの導入によるインタラクションの分岐を行う．以下に, 各モードにおけるユーザの分析行動 (探索目的や着目軸) の例を示す．

- アニメーション: 時系列データの概要としてアニメーションを再生し, 特徴的な変化を検出したり, データの時間軸上の変化特性に基づき, 興味のあるオブジェクトを選択, 追跡する．
- 静止: アニメーションで見出した概要に基づき, 空間軸上にて軌跡の直接操作による詳細な探索や視覚的発見の検証を行い, 知識を形成する．

ここで, 探索はアニメーションモード (視覚的傾向の発見) と静止モード (仮説検証) 間の遷移の繰り返しで行われると仮定する．表 1 に探索モードによる操作とインタラクションの対応関係を示す．モードの変更により, 暗黙のうちに探索目的や着目軸に対応したインタラクションを選択・適用することで，操作を複雑にすることなく両可視化を活用可能である．また, モードの導入に伴う着目軸の分離によって Abstract/Elaborate 操作などにおける時空間的なインタラクションの衝突を回避できると考える．

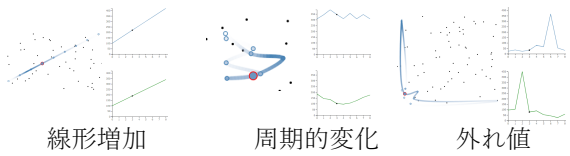


図 3: 軌跡の幾何学的形状の例

表 1: モードによる操作とインタラクションの対応関係

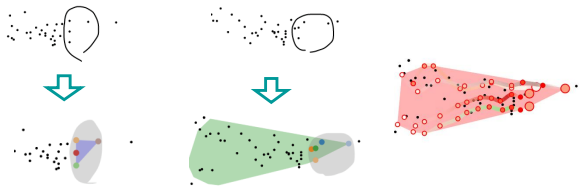
操作	モード	適用インタラクション
ズーム	アニメ	時間的 (再生速度の) ズーム
	静止	空間的ズーム
データ選択	アニメ	データ点のトラッキング
	静止	軌跡の集合, 離散の表示

4.4 オブジェクト/軌跡のグルーピング

提案インタフェースでは, 図 4(a)(b) のようなスケッチベースの入力に基づくオブジェクトの選択とグルーピング, 追跡が可能である．

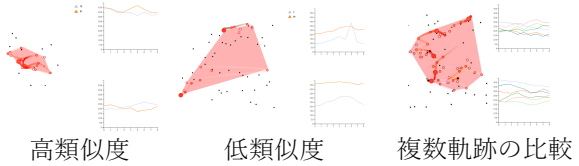
このとき, 複数オブジェクトを選択すると, それらを含む凸包を画面上に描画する．図 5 に示すように, 凸包の面積はデータ間の空間的な近接性 (類似性) に対応する．つまり, データ点同士の凸包は各時点でのデータ群間の類似性を表現し, 軌跡同士の凸包はデータ間の時空間的類似性を表現する．複数オブジェクトに対する選択処理に対して, アニメーションの再生時には, その時点での選択済みデータを含む凸包を計算し, 描画する (図 4(a))．これにより, 着目データ群の時間的な集合/離散を確認できる．また, 選択済データは色と大きさでハイライトされる．そのため, アニメーションによる位置の変化や, 動くデータ点への追随に伴う認知的負荷を軽減できると考える．

対して, 静止モード時には, 選択されたオブジェクト群の全時点におけるデータ座標に対して凸包を計算, 表示する (図 4(b))．また, 複数軌跡に対して選択操作を行った場合にも, 同様の凸包を描画する (図 4(c))．前者では, 空間座標の範囲に基づく対象データの暗黙的な指定により, 時間軸全体にわたるデータ群の変化傾向の把握が促進されると考える．一方で, 後者では明示的に指定した軌跡のグループに対して凸包を重畳することで, 軌跡間の空間的座標の近接性 (類似性) を直感的に把握可能になる．これによって, 分析者が興味を持ったデータの詳細な探索を支援できると考える．



(b) 再生時　(a) 静止時　(c) 複数軌跡選択

図 4: オブジェクト, 軌跡のグルーピングと追跡
(灰色はスケッチされた領域)



高類似度　低類似度　複数軌跡の比較

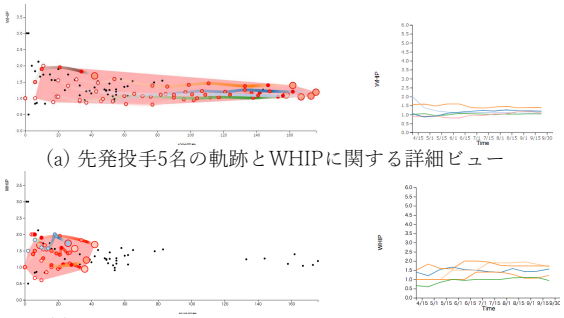
図 5: 凸包による類似性の表現

4.5 相補的なビュー

前述のように, 軌跡による可視化は時系列データの概要の提示手段として有効な一方で, 描画時の曲線補完などによりデータの正確性は喪失する. また 散布図による可視化は, 複数データ間の類似性や関係の判断に関しては有効だが, 視覚的傾向に基づく仮説検証においてはツールチップ等を用いて生データを提示した場合の方が有効であることが示されている [9]. そのため, 提案インタフェースでは散布図ビューに関する詳細情報を提示する詳細ビューを提供する. すなわち, システムは散布図と詳細ビューの協調的マルチビューによって構成される. 詳細ビューでは. 散布図ビューで選択されたオブジェクトに対する操作に基づき, 対応する各属性値の時間的変化を線グラフにより表示する. このとき, 一方のビューへの操作は他方へも適用され, インタフェースの一貫性を保持する. 例えば, 複数軌跡を選択した場合には, それぞれの軌跡における各属性値の比較グラフが詳細ビュー上に表示される. また, 散布図ビュー上で特定の軌跡に対してマウスオーバー処理を行ったときには, 対応するグラフがハイライトされる.

5 ユースケース

提案インタフェースの有効性検証のために, プロトタイプインタフェースを用いたユースケースを示す. インタフェースは Web アプリケーションとして実装し, フレームワークには Ruby on Rails 5, マークアップ言語には HTML と CSS, 可視化とインタラクションの実装に jQuery および D3.js を用いた.



(a) 先発投手5名の軌跡とWHIPに関する詳細ビュー

(b) 抑え投手5名の軌跡とWHIPに関する詳細ビュー

図 6: 先発投手と抑え投手における視覚的傾向の相違

分析対象とする時系列データにはプロ野球ヌルデータ[1]で公開されている 2017 年日本プロ野球パリーグの登板数が 10 試合以上の投手 103 名の投手成績を 15 日間隔でサンプリングしたものを使用した. ここで, 散布図の各軸には WHIP(投球回あたりの与四球と被安打数の合計) および投球回数を使用した.

5.1 投手の評価指標検討のための使用例

提案インタフェースの使用例を, プロ野球における投手成績の評価をタスクとした場合の例を示す.

最初に, 分析者はデータの概要を把握するため散布図ビューを静止モードで各時点について閲覧したり, アニメーションを用いて全体的な傾向の把握を試みた. この時, 投球回数, WHIP をそれぞれ横軸, 縦軸にとった散布図において, 投球回数の多い先発投手と, 少ない中継ぎ投手では空間的位置に明確な違いが見られた. 次に, 分析者は得られた概要や自身のドメイン知識に基づき, 静止モードにて軌跡の探索を用いてシーズンで優良な成績を残した先発/中継ぎ選手の傾向を探索した. その結果, 図 6(a) に示すように, 良い成績を残した先発投手は, 早期に WHIP が特定の値に収束したのに対して, 抑え投手の場合は図 6(b) のように軌跡の変動に個人差が観測された.

その後, 軌跡の選択によるグルーピングなどを行い分析を進めた結果, 同様の最終成績を持つ投手同士でも, 成績の収束時期やパターンに差があることを発見した. 最終的に, 探索で得られた視覚的傾向や洞察に基づき, 分析者は WHIP が早期に収束し, 投球回数が多い先発投手を安定感があるとして評価した. 一方で, 抑え投手に関して, WHIP は投手ごとの傾向の違いは確認できるが, 時系列的に共通する傾向は確認できないため, その他の評価指標を検討すべきであると結論付けた.

このように, 提案インタフェースを利用することで, 時系列データの視覚的分析を支援できると考える.

[1] http://lcom.sakura.ne.jp/NulData/

6 おわりに

本稿では, 時系列データの直接操作による視覚的分析を対象とし, 時間軸上のインタラクションの衝突などの問題点解決を目的として, 軌跡の直接操作に基づく視覚的分析インタフェースを提案した. また, 提案インタフェースをプロトタイプとして実装し, 利用例を示した.

今後は, ユーザビリティテストによる提案インタフェースの定量的・定性的な有効性の検証を行う必要がある. また, 今後の展望として, 知識形成促進のための Semantic Interaction への応用や, データマイニング技術を活用した可視化の品質保証やインタラクション解釈の不確さへの対処, 操作のガイダンス [5] による探索の支援などが挙げられる. 具体的には, 大規模データセットへの拡張を踏まえた DBSCAN などのクラスタリング手法や, 軌跡の形状に基づくハイライトやフィルタリング, 類似性判定などの導入を検討している.

参考文献

[1] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, H. Ziegler, Visual analytics: Scope and challenges, S. J. Simoff, M. H. Bőhlen, A. Mazeika (eds.), Visual Data Mining, Springer, pp. 76-90, 2008.

[2] A. Endert, L. Bradel, C. North, Beyond control panels: Direct manipulation for visual analytics, IEEE computer graphics and applications, Vol. 33, No .4, pp. 6-13, 2013.

[3] A. Endert, P. Fiaux, C. North , Semantic interaction for visual text analytics, Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 473-482, 2012.

[4] G. Robertson, R. Fernandez, D. Fisher, B. Lee, J. Stasko, Effectiveness of animation in trend visualization, IEEE Transactions on Visualization and Computer Graphics, Vol. 14, No. 6, pp. 1325-1332, 2008.

[5] D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, D. A. Keim, What you see is what you can change: Human-centered machine learning by interactive visualization, Neurocomputing, 2017.

[6] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, D. A. Keim, Knowledge generation model for visual analytics, IEEE Transactions on Visualization and Computer Graphics, Vo. 20, No. 12, pp. 1604-1613 , 2014.

[7] B. Shneiderman, Direct manipulation: A step beyond programming languages, IEEE Computer, Vol. 16, No. 8, pp. 57-69, 1983.

[8] S. Kriglstein, M. Pohl, C. Stachl, Animation for time-oriented data: An overview of empirical research, 16th International Conference on Information Visualisation (IV '12), pp. 30-35, 2012.

[9] M. Adnan, M. Just, L. Baillie, Investigating time series visualisations to improve the user experience, Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5444-5455 , 2016.

[10] B. Kondo, C. Collins, DimpVis: Exploring time-varying information visualizations by direct manipulation, IEEE Transactions on Visualization and Computer Graphics, Vol. 20, No. 12, pp. 2003-2012, 2014.

[11] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, P. Dragicevic, Time curves: Folding time to visualize patterns of temporal evolution in data, IEEE Transactions on Visualization and Computer Graphics, Vol. 22, No. 1, pp. 559-568, 2016.

[12] J. S. Yi , Y. A. Kang , J. Stasko, J. Jacko, Toward a deeper uuderstanding of the role of interaction in information visualization, IEEE Transactions on Visualization and Computer Graphics, Vo. 13, No. 6, pp. 1224-1231 , 2007.

[13] B. Saket, A. Srinivasan, E. Ragan, A. Endert, Evaluating interactive graphical encodings for data visualization, IEEE Transactions on Visualization and Computer Graphics, 2017.

[14] A. Endert, M. S.Hossain, N. Ramakrishnan, C. North, P. Fiaux, C. Andrews, The human is the loop: new directions for visual analytics, Journal of intelligent information systems, Vol. 43, No. 3, pp. 411-435, 2014.

[15] C. Yuksel, S. Schaefer, J. Keyser, Parameterization and applications of Catmull-Rom curves, Computer Aided Design, Vol. 43, No. 7, pp. 747-755, 2011.