

## Kizuna: Webble World を用いたデータマイニング視覚化ツール Kizuna: Dataminig Visualizing Tool on Webble World

丸橋 弘明<sup>†</sup>      中元 政一<sup>†</sup>      羽室 行信<sup>†</sup>  
Hiroaki Maruhashi   Masakazu Nakamoto   Yukinobu Hamuro

### 1. はじめに

学術の世界ではデータマイニングに関するさまざまなアルゴリズムが研究・開発されているが、その成果が産官業界において十分に活用されているとは言えない。その理由のひとつとして、アルゴリズムの成果物がコマンドライン実行のみ提供されているケースが多く、一般的なビジネスマンにとって利用する敷居が高いことがあげられる。学術界の成果が産業界へ十分に還元されないことはもったいないことであり、また産業界で実践した結果のフィードバックが少ないことで学術の進歩促進についても不効率をもたらす。

Kizuna はこの課題に問題意識を持って開発されたデータマイニング可視化ツールである。大規模データ前処理支援ツールの NYSOL[1]とデータマイニング支援ツールの Take[2]をクラウドサーバー上で実行し、北海道大学 知識メディアラボラトリーで研究開発されている Webble World[3]を可視化プラットフォームに採用した Web ツールとして提供している\*1。

本稿では、Kizuna のアーキテクチャ構成及びその利用例について述べる。

### 2. Kizuna のアーキテクチャ構成

#### 2.1 概要

Kizuna のアーキテクチャ概要図を図 1 に示す。可視化層、インターフェース層、データ処理層の 3 つの層で構成されている。各層ごとに論理的・物理的にサーバーを分割することが可能であり、サーバーリソースをスケールアウトさせることができる。

インターフェース層・データ処理層は処理対象データごとにそれぞれ同じ構成セットのサーバーが起動する。対象データごとにサーバーを分離することで、企業内の非公開データを対象データとして安全に取り扱うことができるようになっていく。

また、可視化層・データ処理層ともに、可視化部品およびデータ処理アルゴリズムの追加・改良が可能な構成になっている。

#### 2.2 可視化層

可視化層は、処理対象データおよび実行アルゴリズムを選択し入力パラメータを入力する画面機能、処理状況を表示する画面機能、アルゴリズムがデータ処理した結果を表示する画面機能を提供する。

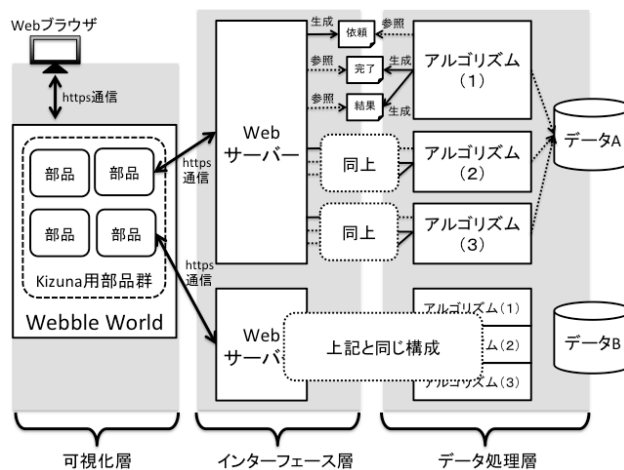


図 1 Kizuna のアーキテクチャ概要図

Webble World[3]を可視化システムのプラットフォームとして採用している。Webble World は北海道大学 知識メディアラボラトリーで研究開発されており、知識が人から人へ共有されて文化が形成されていく meme[4][5]の考え方をコンピュータ上に具現化した可視化プラットフォームである。IntelligentPad の後継バージョンとして、Web 最新技術 (HTML5, Javascript 等) を使用して実装されている。

meme のオープンな文化形成をシステムコンセプトの基調にしているため、GUI 部品の追加・共有および部品の実装コード参照の手段がユーザーに提供されている。

#### 2.3 インターフェース層

インターフェース層は、可視化層から送信されるデータ処理実行依頼内容をデータ処理層に橋渡し役割を担っている。実行するアルゴリズムによっては処理が完了するのに数分かかるものも存在する。よって、このインターフェース層にて、可視化層とデータ処理層が非同期に連携されるように処理プロセスを分離している。

具体的には、可視化層からデータ処理の実行依頼が届くと、対象データ・実行アルゴリズム・インプットパラメータを記録した実行依頼ファイルを作成する。実行依頼ファイルを作成した時点でインターフェース層は実行依頼を受領した旨を可視化層に返して一連の処理プロセスを終了させる。データ処理が完了したかどうかはデータ処理層が生成する実行完了ファイルの有無によって確認し、ファイルが存在すれば実行結果ファイルを参照する。

処理実行依頼、処理完了確認、処理結果参照はすべて可視化層がトリガーとなっており、インターフェース層はそれらをノンブロッキングで処理する。

なお、インターフェース層はアルゴリズムの入出力に依存しないため、可視化部品・アルゴリズムの追加に影響されない。

<sup>†</sup> 関西学院大学 経営戦略研究科, Kwansai Gakuin University

\*1 <https://webble.nysol.jp/#/app?webble=nysol.kizuna>

## 2.4 データ処理層

データ処理層はインターフェース層が配置した実行依頼ファイルの生成を監視している。ファイルが生成されると対象のデータおよびアルゴリズムを実行させる。各アルゴリズムの処理負荷を考慮して、同時実行数を制御する仕掛けも用意している。

処理が完了すると処理完了ファイルを生成するとともに、処理結果ファイルを所定位置に格納する。

データの選択・集計やアルゴリズムに定められたデータ形式への変換は大規模データ前処理支援ツールの NYSOL[1]を用いている。実行可能なアルゴリズムは現時点ではデータマイニング支援ツールの Take[2]を使用したものを用意しているが(表 1), Python 言語で実行できるアルゴリズムであれば追加できるようになっている。

表 1 Kizuna で提供されているアルゴリズム (現時点)

SUMMARY	データの選択・集計
ITEMSET	頻出アイテム集合列挙
POLISH	グラフ研磨
FRIEND	相互類似関係抽出
PAL	2アイテム相関ルールのランクに基づいた類似関係抽出

## 3. Kizuna の利用例

Kizuna ではデモ利用<sup>\*2</sup>を想定して一般公開可能な処理対象データを用意している(表 2)。

表 2 Kizuna で一般公開している処理対象データ (現時点)

焼肉レストラン レシートデータ
Online Retail Data Set <sup>*3</sup>
Wikipedia(戦国時代・安土桃山時代の人物一覧)
Wikipedia(都市一覧)
Wikipedia(プログラミング言語一覧)
Wikipedia(日本の山一覧)
Wikipedia(ポピュラー音楽の音楽家一覧(日本・個人))
Wikipedia(コンピュータゲームのタイトル一覧)
Wikipedia(プロ野球選手)

Kizuna の利用例を 2 つ示す。1 つ目は焼肉レストランのレシートデータに対して polish アルゴリズム[6]を実行したものである(図 2)。polish は複雑に絡み合ったグラフデータを研磨して人間が比較的理解しやすい粒度のクラスタを抽出するアルゴリズムである。焼肉レストランのレシートデータは、メニュー商品名をアイテム、1 レシートを 1 トランザクションとしたデータであり、polish 実行前に同時購買されたアイテムのグラフデータに加工している。店の定番のメニュー組合せやお客の嗜好性などが見て取れる。

2 つ目は Wikipedia のデータを日本の戦国武将のページのみ抽出し、人物名(つまりページ)をアイテム、ページ内に登場する単語をトランザクションとなるように加工したデータに対して、pal アルゴリズム[7]を実行したものである(図 3)。Wikipedia に記載されている単語情報のみの関係性ではあるが、興味深い類似関係が見て取れる。

\*2 <https://webble.nysol.jp/#/app?webble=nysol.kizuna>

(デモ利用時のユーザー/パスワードは guest / guest)

\*3 UCI Machine Learning Repository: Online Retail Data Set (<http://archive.ics.uci.edu/ml/datasets/online+retail>)

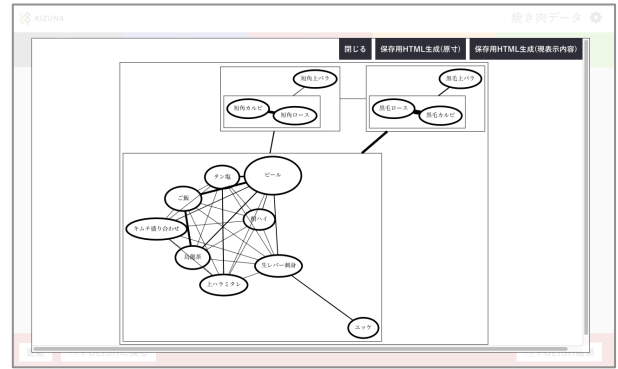


図 2 焼肉レストランレシートデータへの polish 実行結果

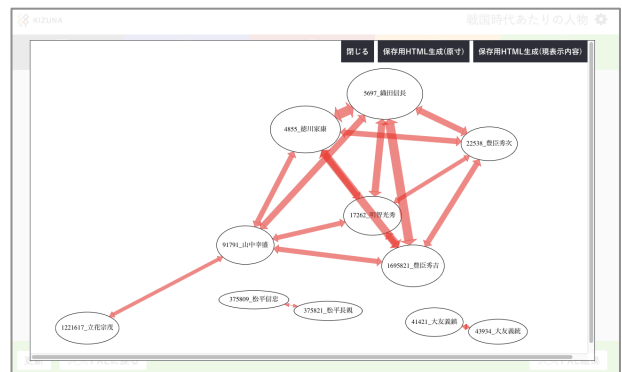


図 3 Wikipedia 戦国武将データへの pal 実行結果

## 4. おわりに

Kizuna は継続して開発・運用している Web システムである。現在は数種類のデータやアルゴリズムに対してのみ提供しているが、今後はアルゴリズム研究者やデータ提供者と協力しながらサービス内容を拡充させていきたいと考えている。

### 謝辞

焼肉レストランのレシートデータは岡山フードサービス株式会社よりご提供いただいた。Webble World 関連のシステム開発においては北海道大学 知識メディアラボラトリーの Webble 開発チームにご協力いただいた。ここに感謝の意を表す。本研究は、JST CREST(グラント番号: JP-MJCR1401)の研究助成を受けている。

### 参考文献

- [1] 中元政一, 羽室行信, NYSOL: Python における大規模データ前処理支援ツール, FIT2018(2018).
- [2] 羽室 行信, 宇野 毅明, 中元 政一, 中原 孝信, 丸橋 弘明, Take: Python におけるデータマイニング支援ツール, FIT2018(2018).
- [3] M. Kuwahara, Micke, Y. Tanaka, Yuzuru "Webble World 3.0: In the Borderland Between Being a User or a Developer", ISIP 2014, CCIS, vol. 497, pp. 85-96, Springer, Heidelberg (2016).
- [4] Dawkins, R: The Selfish Gene. Oxford University Press, Oxford(1976).
- [5] Blackmore, S.: The Meme Machine. Oxford University Press, Oxford(1999).
- [6] Takeaki Uno, Hiroki Maegawa, Takanobu Nakahara, Yukinobu Hamuro, Ryo Yoshinaka, Makoto Tatsuta, "Micro-clustering by data polishing", 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, pp. 1012-1018, Dec., 2017.
- [7] 岩崎幸子, 中元政一, 中原孝信, 宇野毅明, 羽室行信, グラフ構造による相関ルールの視覚化ツール: KIZUNA, 2017 年度人工知能学会(第 31 回), ウィンクあいち, 2017/5/24.