

Use Case Suzuki & Hosoya:

Computational Stylistic Analysis of Popular Songs of Japanese Female Singer- songwriters

Aiming to imprint the research process on the NeDiMAH Methods Ontology, this document presents the scope and description of a case study of humanities research followed by snapshots of its actual modeling in a graph visualization.

Published work containing all modeled information:

Suzuki, T., Hosoya, M.: Computational Stylistic Analysis of Popular Songs of Japanese Female Singer- songwriters. Digital humanities quarterly. 8, 1–13 (2014)

Short summary of modeled scholarly work:

The authors use computational linguistic methods in order to analyze popular songs composed by Japanese female singer-songwriters. In order to accomplish this, they first gather a sample of 116 song lyrics using the on-line database services “Uta-Net” and “Uta-Map”. Second, using a specific tool (“MeCab”), they apply morphological analysis to their data, in order to prepare their sample for the application of the two Machine Learning Methods. In the next step, they employ the “Kernel PCA” method with specific parameters adjusted to their dataset in order to examine the factors affecting the lyrical characteristics of the songs and finally they employ the “Random Forests” method in order to perform the classification experiments, extract important features regarding the classification and find the special distinguishing lyrical characteristic of every singer-songwriter. Each activity uses / produces information resources and results in several statements that constitute their analysis on the subject and are represented through various tables figures or units of text in the paper.

Description of instantiation process:

The two researchers are modeled as instances of the *Person* class [Ac₁: Takafumi Suzuki] and [Ac₂: Mai Hosoya]. Indicative instances of *Activity* class are the general activity [A₁: Analyzed popular songs] which is decomposed into its sub-activities: [A₂: Gathered 116 Songs] followed by [A₃: Applied Random Forests]. These can further be linked through the *hasScope* relationship with the *ActivityType* terms [Analyzing] [Gathering] and [Classifying] respectively.

The method [M₁: Random forests] -as described in [20]- consists of 3 steps, modeled here as [St_{1.1}: Sample from i cases at random from the original text-feature matrix M_[i,j]], followed by [St_{1.2}: Extract random subsets of [root j] variables from a bootstrap sample to make a sample

for constructing an unpruned decision tree], which is followed by [St_{1.3}: Calculate the variable Importance (Vlacu) for the classification experiments]. In addition, the reference in bibliography that is provided in the paper can be modeled here as an instance of the *InformationResource* class connected through the *isReferencedIn* relationship. Finally, the Taxonomy of *ActivityTypes* can be used here also, in order to “tag” instances of *Method* class through the *isEmployedFor* relationship. In our case study the method instance M₁ can be related through the *isEmployedFor* relationship to the *ActivityTypes* term [4.2.23 Principal Component Analysis] whereas M₂ to [4.3.3_Classifying] and [4.2.29_StatisticalAnalysis] respectively.

The different figures (such as figures 1 & 2) that appear are modeled as instances of *InformationRsource* class [IR₃: *Figure1*] & [IR₄: *Figure2*], with Format [Mt: .jpg], Type [*It*: 7.3_Chart] and topics [*Tp*₁: *Three-dimensional scatter plots of the first three principal components by kernel PCA with labels representing the names of the songwriters*] & [*Tp*₂: *Three-dimensional scatter plots of the first three principal components by kernel PCA, with labels representing the years of release*] respectively.

Other elements of the model are: [Software: MeCab, Uta-Map, Uta-Net], [Statement: “Pronouns, final particles, and auxiliary verbs are particularly important for discriminating the songs by 10 Japanese female singer-songwriters”], [ContentItem: Fig.1, Fig.2, dataset of 116 song lyrics], [Goal: find the distinctive lyrical characteristic of Japanese female song-writers], [Topic: Computational Stylistic Analysis of Popular Songs of Japanese Female Singer-songwriters], connected accordingly.

