**Pour une représentation dynamique et analytique des données**

**Objectif** : fournir une interface graphique dynamique pour représenter une classification et permettre une appréciation et une analyse humaine de l’information apportée par la classification.

Le système comporte une fenêtre avec plusieurs onglets qui proposent différents modes de représentation des classifications. On suppose ici que les classifications sont des partitions floues avec des classes comportant un attribut texte informant sur la nature de sa composition.

**1er mode de représentation** Analyse en composantes principales

On commence par rendre stricte la partition en affectant à chaque objet la classe qui maximise son degré d’appartenance. Les cas d’égalité sont brisés de manière arbitraire (aléatoire ?).

On effectue ensuite une Analyse en composantes principales du nuage de points qui permet de représenter les données comme un nuage de points en deux dimensions.

On propose deux vues différentes pour cette représentation avec un bouton qui permet de changer de vue.

Vue 1 : on représente la partition calculée par le classifieur à l’aide d’un code couleur légendé par les attributs texte des classes. A noter, la dernière classe de la classification peut constituer une classe inconnue regroupant les données qui n’ont pas pu être classifiées. On les représente dans une couleur neutre (gris par ex.)

Vue 2 : on représente les données en utilisant une classification a priori des données. Idéalement, le code couleur doit être le même (on suppose que les attributs texte des classes de la partition calculée et de la classification a priori sont identiques). On représente dans une couleur neutre (gris par ex.) les données pour lesquelles on ne dispose pas a priori d’information de classification.

Interactivité : si c’est possible, une infobulle apparaît avec des informations au survol d’un point (avec ses attributs comme l’auteur et le titre) ou d’un élément de la légende (comme la composition exacte de la classe (par ex si un cluster d’attribut texte Auteur A comporte seulement 70% de A et 30% de B, on le précise au survol).

**2ème mode de représentation** Analyse en composantes principales et polygonisation

On commence par rendre stricte la partition en affectant à chaque objet la classe qui maximise son degré d’appartenance. Les cas d’égalité sont brisés de manière arbitraire (aléatoire ?).

On effectue ensuite une Analyse en composantes principales du nuage de points qui permet de représenter les données comme un nuage de points en deux dimensions.

On propose deux vues différentes pour cette représentation avec un bouton qui permet de changer de vue.

Vue 1 : on représente les clusters de la partition calculée à l’aide de polygones (par exemple triangles de Delaunay) et d’un code couleur légendé par les attributs texte des classes. Les points sont représentés dans une couleur neutre (par exemple noire)

Vue 2 : même chose avec la partition a priori. En noir les points pour lesquels ont possède une information a priori et en gris ceux pour lesquels on n’a pas d’information a priori.

Interactivité : si c’est possible, une infobulle apparaît avec des informations au survol d’un point (avec ses attributs comme l’auteur et le titre) ou d’un élément de la légende (comme la composition exacte de la classe (par ex si un cluster d’attribut texte Auteur A comporte seulement 70% de A et 30% de B, on le précise au survol). Des cases à cocher permettent de désactiver la représentation des points ou de chaque cluster individuellement pour démonter les superpositions éventuelles.

Gestion des intersections : si deux clusters s’interpénètrent, on le gère par mélange des couleurs)

Attention à la gestion des points inconnus qui ne doivent pas être représentés par des polygones.

**3ème mode de représentation** Analyse en composantes principales et densité

On ne suppose plus la partition stricte et on conserve une partition floue a priori. On effectue une Analyse en composantes principales du nuage de points qui permet de représenter les données comme un nuage de points en deux dimensions.

On propose deux vues différentes pour cette représentation avec un bouton qui permet de changer de vue.

Vue 1 : On calcule la variance de chaque cluster. A chaque cluster, on associe un point du cercle chromatique extérieur (couleurs intenses). On va calculer k fonctions d’intensité de couleur ou k est le nombre de clusters définis sur le canevas d’affichage graphique.

Pour chaque point p, pour chaque cluster i, on ajoute à la fonction d’intensité du cluster i une Gaussienne radiale centré au point p et de variance lié à celle du cluster i.

On normalise les fonctions d’intensité pour qu’elles soient comprises entre 0 et 1.

On superpose ensuite les différentes colorations en mélangeant les couleurs et en faisant varier l’intensité en se rapprochant de l’intérieur du cercle chromatique.

On représente les points dans les couleurs intenses et on propose un code couleur.

Vue 2 : idem avec la classification a priori. Les points qui n’ont pas d’information a priori sont représentés en gris et ne forment pas de densité de couleur.

Interactivité : si c’est possible, une infobulle apparaît avec des informations au survol d’un point (avec ses attributs comme l’auteur et le titre) ou d’un élément de la légende (comme la composition exacte de la classe (par ex si un cluster d’attribut texte Auteur A comporte seulement 70% de A et 30% de B, on le précise au survol). Des cases à cocher permettent de désactiver la représentation des points ou de chaque cluster individuellement pour démonter les superpositions éventuelles.

**4ème mode de représentation** Schéma d’analyse

En construction