

Применение FBC–кластеризации временных рядов к анализу динамики оценок персонала

Аспирант кафедры “Информационные системы” Сибирев И.В.
Научный руководитель д.т.н. Афанасьева Т.В.

УлГТУ, Ульяновск 2017 г.

Аннотация

В презентации рассматривается применение для анализа динамики оценок персонала (KPI) – средств комбинированной кластеризации временных рядов, аппарата нечётких лингвистических термов, методик извлечения основной тенденции и тренда временного ряда. Предлагается алгоритм Fuzzy Behavior Clustering. FBC – подход к кластеризации, основанный на нечётком поведении временных рядов.

Наиболее распространённые методы оценок персонала.

- **Аттестация** – периодическая проверка трудовых качеств сотрудников с определением уровня их профессиональной подготовки (существуют нормативные документы, регламентирующие проведение аттестации).
- **«Метод 360 градусов»**, в основе которого лежит агрегация мнений о сотруднике со стороны коллег, начальства и подчинённых. Плюсы: оценка объективна, метод учитывает особенности корпоративной организации и культуры, возможны сигналы обратной связи, производится учёт мнений клиентов. Минусы: требование конфиденциальности, нет учёта результативности сотрудников, данные могут быть искажены респондентами.
- **Assessment Centre** (центр оценки) – производится моделирование основных видов деятельности персонала, оценка эффективности с учетом персонально важных качеств, применяются психологические тесты, интервьюирование, специальные упражнения и т.д.

Также используются

метод анкетирования, описательный метод оценки, метод классификации

(ранжирование сотрудников по результатам аттестации), **метод попарного сравнения**

(выдвигается наиболее предпочтительный вариант и сотрудники попарно сортируются по убыванию предпочтения), **рейтинг или**

метод сравнения (оценивается степень соответствия сотрудников с занимаемой должностью) и т.д

Параметры классифицируют на:

- Параметры классифицируют на:
естественно-биологические, социально-экономические, технико-организационные, социально-психологические, рыночные и т.д.
- По отношению к персоналу:
профессиональные, деловые, морально-психологические и др.
- Параметры делят на прямые и косвенные, объективные и субъективные

Используемые нами параметры:

- KPI - ключевые показатели эффективности
- KPI составляется из параметров: срок выполнения, количество и качество работы.

Существующие проблемы оценки персонала:

- Проблема неточностей, связанных с искажением респондентами или устареванием данных.
- Проблема того, что почти все критерии пороговые, то есть подразумевают чёткое разделение на «чёрное и белое».

Данные проблемы можно решить, привнося нечёткость.

- Нечёткие методы подразумевают создание ядер нечётких множеств и описание степени принадлежности к ним.
- Мы предлагаем сделать шаг в сторону нечёткости, считая, что после чёткой кластеризации по параметрам KPI все сотрудники, которые попали в кластер, уже преодолели порог и их параметры относительно однородны. Далее соотносим сотруднику не его вектор параметров, а центр кластера, в который он попал.

Преимущества

- Выдавая сотрудникам задания, подходящие для данного кластера, можно стимулировать профессиональный рост, если параметры работника ниже среднего по кластеру. Можно формировать вектор развития для перехода в другой кластер, если параметры – выше среднего по кластеру.
- Методы кластеризации используют для поиска аномальных объектов, что вполне подходит для «поиска талантов».

Почему именно FBC

- Традиционно используемые в кластеризации методы не учитывают тип поведения временных рядов, что может привести к огрехам при кластеризации.
- Для получения более надёжных результатов кластеризации временных рядов в работе [Afanasieva T., 2017] **предлагается использовать подход иерархической кластеризации FBC.**
- Он основан на моделировании временных рядов с дальнейшей поточечной кластеризацией. Отличительной особенностью предлагаемого подхода является использование F-преобразования [Novák, 2016] и понятия нечеткой тенденции [Afanasieva, 2015] в кластеризации временных рядов на основе модели.

Используемая модель временного ряда

- $X = \{x_t \in R, t = 1, 2, \dots, n\}$ (2) – исходный временной ряд.
- x_t – значение временного ряда X в момент времени $t, t = 1, 2, \dots, n, n$ – длина временного ряда.
- $x_t = \alpha * f_t + \beta * \psi_t$ (1) – упрощенная **аддитивная модель временного ряда.**
- f_t – компонента тренда. ψ_t – компонента колебаний.
- $\psi_t = s_t + e_t, s_t$ и e_t – регулярные и нерегулярные колебания, соответственно.

Идея FBC подхода

- заключается в представлении временных рядов X (2) на трех уровнях иерархии: на уровне общих тенденций gt , на уровне компонент тренда f_t и на уровне компонент колебаний ψ_t .
- $\langle gt, f_t, \psi_t \rangle$, (2)
- При FBC кластеризации временных рядов X на основе модели (1), временные ряды в одном кластере будут сходны по общим тенденциям gt ; компонентам тренда f_t и по компонентам колебаний ψ_t [Afanasieva T., 2017].

Fuzzy Behavior Clustering

Нами используются лингвистические термы, и соответствующие им коэффициенты модели $x_t = \alpha * f_t + \beta * \psi_t (1)$, f_t – компонента тренда. ψ_t – компонента колебаний:

- «стабильность» - $\alpha = 0, \beta = 0$
- «колебания» - $\alpha = 0, \beta = 1$
- «падение» - $\alpha = -1, \beta = 0$
- «рост» - $\alpha = 1, \beta = 0$
- «рост с колебаниями» - $\alpha = 1, \beta = 1$
- «падение с колебаниями» - $\alpha = -1, \beta = 1$

Алгоритм Fuzzy Behavior Clustering

Шаг первый. Преобразование лингвистических термов основных тенденций из множества GT . $GT = \{\text{«fall»}, \text{«stability»}, \text{«growth»}, \text{«fluctuation»}, \text{«growth with fluctuation»}, \text{«fall with fluctuation»}\}$

Используется GTI алгоритм [Afanasieva, 2016], основанный на представлении каждого временного ряда.

$$X_i \rightarrow gt_i \in GT, X_i \in Y, i = 1, 2, \dots, m.$$

Алгоритм Fuzzy Behavior Clustering

- **Шаг второй.** Первый уровень кластеризации временных рядов: кластеризация временных рядов по основным тенденциям с использованием эквивалента в виде лингвистического термина основной тенденции из множества GT .
- $Y = \{X_1, X_2, \dots, X_m\}$ – исходное множество временных рядов различной длины
- $Y = Y_{fall} \cup Y_{growth} \cup Y_{stab} \cup Y_{fluct} \cup Y_{fall_fluct} \cup Y_{growth_fluct}$.

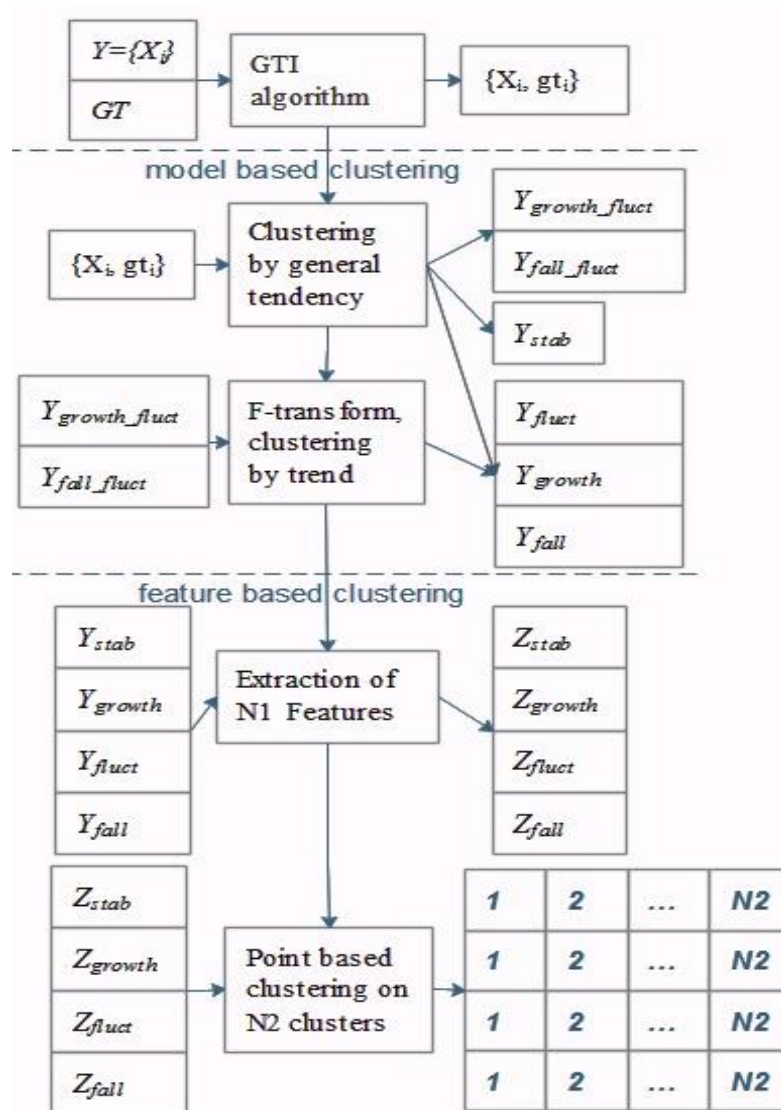
Алгоритм Fuzzy Behavior Clustering

- **Шаг третий.** Второй уровень кластеризации временных рядов: используется модель временного ряда на основе кластеризации по трендам.
- Такая кластеризация базируется на разделении каждого временного ряда согласно его поведению. Y_{growth_fluct} «growth with fluctuation» и Y_{fall_fluct} «fall with fluctuation» оказывают влияние на тренд f_t и флуктуации ψ_t , эти составляющие используются в инверсивном F-преобразовании [Jain, 1988].
- Временные ряды с ярко выраженной трендовой составляющей попадут в кластер с поведением «fall» или «growth». Временные ряды флуктуационной составляющей попадут в кластер «fluctuation».
- $Y_{growt_fluct} = Y_{growth} \cup Y_{fluct}$
- $Y_{fall_fluct} = Y_{fall} \cup Y_{fluct}$
- Двум множествам трендовых компонентов рост и падение Y_{growth} и Y_{fall} можно поставить в соответствие множества: Y_{growt_fluct} и Y_{fall_fluct} . В последние два множества входит множество флуктуационных составляющих Y_{fall} .
- После этого, каждый временной ряд принадлежит одному из множеств: «growth with fluctuation» или «fall with fluctuation».
- Далее, по аналогии, происходит **разделение множества исходных временных рядов на четыре группы** :
- $Y = Y_{fall} \cup Y_{growth} \cup Y_{stab} \cup Y_{fluct}$
- $TR \subset G, TR = \{\text{«fall»}, \text{«stability»}, \text{«growth» and «fluctuation»}\}.$

Алгоритм Fuzzy Behavior Clustering

- **Шаг четвёртый. Комбинированная кластеризация временных рядов.** Третий уровень кластеризации: кластеризация временных рядов на кластеры, описанные в шаге 3. Параметрически ориентированная кластеризация.
- А) Получение вектора характеристик, описывающих временной ряд.
- Группируем величины, характеризующие временной ряд на $N1$ кластеров. Затем рассчитываются барицентры кластеров. Из барицентров, отсортированных в хронологическом порядке, строится вектор параметров z , характеризующих временной ряд. Так получаем множество векторов параметров $Z = Z_{fall} \cup Z_{growth} \cup Z_{stab} \cup Z_{fluct}$, характеризующих временные ряды множества $Y_{fall}, Y_{growth}, Y_{stab}, Y_{fluct}$.
- В) Кластеризация по векторам признаков.
- Кластеризация каждого из множеств $Z_{fall}, Z_{growth}, Z_{stab}, Z_{fluct}$ по векторам признаков на $N2$ кластеров.
- Итак, алгоритм FBC кластеризации делит множество временных рядов Y на $4 * N2$ кластеров. Временные ряды при этом могут обладать разным количеством точек и различными флуктуационными составляющими. При этом каждый временной ряд может принадлежать к одному из кластеров поведенческой составляющей и к одному из кластеров трендовой составляющей. Каждый временной ряд может быть охарактеризован нечётким лингвистическим термом и степенью принадлежности.

Схема FBC подхода.



Заключение.

- Основной идеей предлагаемого FBC подхода является внедрение в кластеризацию временных рядов **группировки поведения временных рядов с использованием лингвистических человеко-ориентированных термов общей тенденции**. Кластеризация поведения временных рядов в лингвистических терминах общей тенденции осуществляется на основе нечеткого представления временных рядов [Afanasieva, 2015], F-преобразования [Novák, 2016] и модели временных рядов.
- **Можно найти сходство временных рядов, учитывая структуру их модели на трех уровнях иерархии**. На уровне общей тенденции временные ряды группируются по лингвистическому представлению их поведения, на уровне тенденции создаются более мелкие кластеры.
- Последний этап кластеризация с помощью численных методов кластеризации. **Алгоритм FBC обладает высоким уровнем модульности** и позволяет использовать любые методы численного кластерного анализа. **Эксперименты показали, что подход повышает качество кластеризации временных рядов по сравнению с численными методами**.
- **Результаты на любом уровне кластеризации временных рядов могут быть использованы в интеллектуальном анализе данных временных рядов для принятия решений, для прогнозирования, классификации и выявления аномалий**.
- FBC подход применим, в частности, для оценки динамики параметров KPI персонала. Он позволяет увеличить гибкость и шумоустойчивость оценочного аппарата за счёт использования нечётких методов. Позволяет изучать динамику параметров. Метод применим для автоматизированной обработки данных. Он основан на человеко-ориентированных лингвистических терминах, результат обработки данных удобен для человеческого восприятия.

Литература

- **[Афанасьева и др., 2015]** Афанасьева Т.В., Заварзин Д.В., Сибирев И.В. Алгоритм поиска и удаления аномалий временных рядов на основе применения кластеризации // Радиотехника. - Москва. - 2015. С. 59-62
- **[Afanasieva, 2015]** T. Afanasieva, N. Yarushkina, M. Toneryan, D. Zavarzin, A. Sapunkov and I. Sibirev. Time series forecasting using fuzzy techniques. In Proc. of IFSA-EUSFLAT, 2015, pp. 1068-1075.
- **[Afanasieva, 2016]** T. Afanasieva, A. Sapunkov. Selection of Time series Forecasting Model Using a Combination of Linguistic and Numerical Criteria. In Proc. of 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), 2016, pp. 341-345.
- **[Afanasieva T., 2017]** Afanasieva T., Yarushkina N., Sibirev I. Time Series Clustering using Numerical and Fuzzy Representations // International Journal of Fuzzy System Applications. – Japan: Otsu 2017.
- **[Jain, 1988]** Jain A. K., and R. C. Dubes. Algorithms for Clustering Data. Prentice-Hall advanced reference series. Prentice-Hall, Inc., Upper Saddle River, NJ, 1988.
- **[Makridakis, 1998]** S. Makridakis, S.C. Wheelwright,, and R.J. Hyndman. Forecasting methods and applications. John Wiley & Sons, Inc., 1998.
- **[Novák, 2016]** Novák V., Perfilieva I., Dvorak A. Insight into Fuzzy Modeling. Wiley, 2016.
- **[Perfilieva, 2006]** I. Perfilieva, Fuzzy transforms: theory and applications. In Fuzzy Sets Systems, 2006, 157, pp. 993–1023.
- **[Sibirev I., 2017]** Sibirev I., Afanasyeva T. Development and Study of a Combined Algorithm for Temporal Series Clustering // Open Semantic Technologies for Intelligent Systems. – Республика Беларусь: Минск. 2017. С. 121-126.
- **[Чемеков В., 2016]** Чемеков В., Оценка по KPI работников бэк-офиса. Разберем по шагам. [Электронный ресурс]// [сайт].URL: <http://hr-media.ru/otsenka-po-kpi-rabotnikov-bek-ofisa-razberem-po-shagam/> (дата обращения: 11.08.2017).
- **[diplomba]** diplomba.ru, Методы оценки персонала [Электронный ресурс]// ": [сайт].URL: <http://diplomba.ru/work/30404> (дата обращения: 11.08.2017).
- **[Grandars]** Grandars.ru, Оценка персонала – система, методы и критерии оценки [Электронный ресурс]// [сайт].URL:<http://www.grandars.ru/college/biznes/ocenka-personala.html> (дата обращения: 11.08.2017).
- **[Studfiles]** Studfiles.net, [Управление персоналом](https://studfiles.net/preview/938148/) [Электронный ресурс]// Санкт-Петербургский государственный электротехнический университет "ЛЭТИ": [сайт].URL: <https://studfiles.net/preview/938148/> (дата обращения: 11.08.2017).