

Negociación aplicada a Recomendación a Grupos: Un Estudio sobre Usuarios en el Dominio de Películas

Christian Villavicencio, Silvia N. Schiaffino, J. Andrés Díaz-Pace, Ariel Monteserin
ISISTAN-CONICET
UNICEN

Tandil, Buenos Aires, Argentina

{christian.villavicencio, silvia.schiaffino, andres.diazpace, ariel.monteserin}@isistan.unicen.edu.ar

Abstract—Providing recommendations to groups of users has become popular in many applications today. Even though there are several group recommendation techniques, the generation of recommendations that satisfy the group members in an even way remains a challenge. Because of this, we have developed a multi-agent approach called MAGReS that relies on negotiation techniques to improve group recommendations. Our approach was tested (on the movies domain) using synthetic data with satisfactory results. Given that the results when using synthetic data may sometimes differ with reality, we decided to assess MAGReS using data from real users. The results obtained showed firstly that, in comparison with the recommendations produced by a traditional approach, the recommendations of MAGReS produce a greater level of satisfaction to the group, and secondly that the proposed approach was able to predict more accurately the satisfaction levels of the group members.

I. INTRODUCCIÓN

En la actualidad la conectividad a Internet permite a los usuarios acceder a una gran cantidad de información en sólo unos instantes. En este contexto, un usuario que está interesado en comprar un producto, contratar un servicio o realizar una determinada actividad (como por ejemplo ver una película) enfrenta el problema de la “sobrecarga de información” [1], [2]. Una simple búsqueda en Internet puede arrojar muchos resultados y el usuario no puede (ni tiene el tiempo necesario para) revisarlos a todos, además de que no todos serán relevantes a sus intereses. Con el fin de aliviar este problema y ayudar al usuario a encontrar lo que desea más rápidamente surgieron los sistemas de recomendación.

Un sistema de recomendación (SR) [3] provee asistencia a los usuarios con el objetivo de resolver el problema de elegir qué ítems, en un dominio dado, se ajustan mejor a sus preferencias, necesidades y gustos. En algunos dominios, como el de películas [4], música [5] y turismo [6], [7], el sujeto al cual se le realizan las recomendaciones puede que no sea un único usuario sino un grupo de usuarios. Esto motiva la necesidad de un sistema de recomendación a grupos (SRG), que es capaz de, dado los miembros del grupo y sus perfiles de usuario, hacer recomendaciones que satisfagan al grupo “como un todo”, complaciendo tanto como sea posible a todos los miembros del grupo [8].

Realizar recomendaciones a grupos implica varios desafíos. Por un lado se debe tener en cuenta que, cuando se realizan recomendaciones a grupo, el pasaje de SR a SRG requiere (según [8]) que se realicen las siguientes subtareas: (i) adquirir información sobre las preferencias grupales, (ii) generar las recomendaciones, (iii) explicar las recomendaciones y (iv) dar soporte en la toma de decisiones final (por parte del grupo). Por otro lado, existen desafíos vinculados al cambio del sujeto de la recomendación (es decir, el cambio de individuo a grupo). En primer lugar, muchos de los sistemas de recomendación existentes fueron creados para recomendar ítems a un único usuario y no contemplan el modelado de grupos de usuarios, especialmente en lo que a preferencias se refiere. En segundo lugar, cada grupo tiene su propia dinámica a la hora de tomar decisiones, por lo que recomendar ítems a un grupo no es lo mismo que recomendarle a un único usuario. Finalmente, hay que tener en cuenta que los intereses de los miembros del grupo suelen entrar en competencia, por lo que generalmente hay que resolver conflictos de intereses.

El problema del enfoque tradicional está dado por las técnicas de agregación que se aplican, dado que si bien existe una gran variedad de técnicas para agregar información, todas comparten una característica: la información agregada muchas veces no refleja fielmente los datos de origen, es decir, aquellos que fueron agregados (por ejemplo, si se utiliza un promedio). A causa de esto, las recomendaciones realizadas para un grupo cuando se utiliza el enfoque tradicional no se suelen corresponder con los intereses reales del grupo, por lo que los ítems recomendados no resultan relevantes para el grupo y, como resultado, la recomendación termina siendo ignorada por los miembros del mismo.

En este contexto, se necesita una forma de realizar recomendaciones grupales que vaya más allá de las técnicas de agregación tradicionales y aplique enfoques que permitan modelar adecuadamente las preferencias grupales y que por ende tienda a generar mejores recomendaciones grupales. Existen diversos enfoques al momento de implementar sistemas de recomendación, los cuales pueden ser puros [9], [10], [4], [11], o híbridos [12], [13], [14], [15], [16]. Los recomendadores híbridos son aquellos que utilizan otro tipo

de sistema para producir las recomendaciones, es decir que no aplican de forma “pura” las estrategias de recomendación. En particular, resulta interesante estudiar un tipo de recomendador híbrido que utiliza Sistemas Multi-Agente (SMA) [17], [18] en el proceso de recomendación. Los SMA son sistemas formados por múltiples agentes inteligentes [19], [20] que interactúan entre sí para lograr un determinado fin. Una modalidad de interacción para estos agentes es a través de un proceso de negociación, el cual varía según el modelo que se siga en el sistema. Hay varios modelos de negociación y la aplicabilidad de cada uno depende del problema que se deba resolver o, en otras palabras, de la razón de existencia del SMA [21], [22], [23], [24].

En algunos casos la combinación entre SR y SMA se realiza explícitamente con el objetivo de mejorar la calidad de las recomendaciones, en otros simplemente porque la arquitectura de un SMA permite modelar más adecuadamente el problema a resolver y la división de responsabilidades entre distintos módulos. En este artículo se describe un enfoque denominado MAGReS (por *Multi-Agent Group Recommender System*, inicialmente presentado como PUMAS-GR [25], [26]), el cual se ubica dentro del primer grupo de sistemas mencionado y tiene como objetivo mejorar la calidad de las recomendaciones buscando: (i) incrementar la satisfacción del grupo como un todo y (ii) que los miembros del grupo sean satisfechos de la forma más igualitaria (o uniforme) posible.

Los resultados de pruebas preliminares con datos sintéticos [25] demostraron que MAGReS es capaz de cumplir ambos objetivos. En este artículo se presenta un estudio acerca del comportamiento de MAGReS cuando se realizan pruebas con usuarios reales, lo cual permitirá evaluar: (i) la calidad “real” de las recomendaciones (en función de cuán satisfechos quedan los miembros del grupo y el grupo como un todo) y (ii) en qué magnitud la misma ha aumentado respecto de las recomendaciones producidas por un enfoque tradicional. La evaluación se llevó a cabo considerando tanto los niveles de satisfacción estimados por los enfoques como los especificados por los usuarios al evaluar las recomendaciones realizadas. Los resultados hallados son prometedores ya que fue posible comprobar que, en muchos casos, MAGReS logró producir recomendaciones que satisficieron en mayor medida a los usuarios. Adicionalmente, un análisis posterior confirmó que el error de estimación respecto a los valores de satisfacción era menor al usar MAGReS, lo que indica que sus recomendaciones representan de forma más fiel los intereses del grupo.

El resto del artículo está estructurado de la siguiente forma. La Sección II se exploran trabajos previos en los que se aplica SMA a SR. En la Sección III se explica el enfoque propuesto. En la Sección IV se presentan los resultados de los experimentos realizados con usuarios reales. Finalmente, la Sección V se mencionan algunas conclusiones y los lineamientos de trabajo futuro.

II. TRABAJOS RELACIONADOS

El problema de la generación de recomendaciones grupales comenzó a ser investigado en la última década [27],

y la mayoría de las soluciones propuestas comparten una característica: buscan reutilizar la tecnología desarrollada para producir recomendaciones a usuarios individuales a través de técnicas de agregación. Dónde y cuándo se utilizan esas técnicas varía de un SR a otro, pero es posible clasificar los sistemas de recomendación en 3 categorías:

- 1) Aquellos que combinan recomendaciones individuales, es decir, generan recomendaciones para cada uno de los miembros del grupo y luego combinan dichas recomendaciones individuales en una única recomendación grupal. [28]
- 2) Aquellos que fusionan la información de los perfiles de usuario de cada miembro del grupo, obteniendo así una aproximación del perfil grupal, el cual es luego asociado a un usuario virtual para el cual se genera la recomendación utilizando un SR. [29]
- 3) Aquellos que agregan las preferencias individuales (ratings). En este caso, se genera un usuario virtual cuyas preferencias se calculan a partir de la combinación de las preferencias de los miembros del grupo, para luego utilizar un SR con el objetivo de generar una recomendación para el mencionado usuario virtual.

Por otro lado, los SMA han sido aplicados en distintas áreas y dominios. En lo que refiere específicamente a los SR, se han propuesto diversas técnicas basadas en SMA para generar recomendaciones tanto individuales como grupales en diferentes dominios, tales como: personalización adaptativa de sitios web [30], e-commerce [31], juegos en teléfonos móviles [32], extracción de conocimiento semántico [15], turismo [33], entre otros. La forma en la que se utilizan los SMA en el desarrollo de los SR varía de un sistema a otro. En algunos de los sistemas propuestos los agentes se utilizan para modelar a los usuarios [34], [15], en otros para generar las recomendaciones (creando un agente recomendador por cada una de las técnicas de recomendación disponibles en el sistema) y/o modelar componentes del sistema que tienen una funcionalidad específica (como buscar información en la Web) [33], [14], [35], [36], y en otros se combinan ambos enfoques [12], [31], [37], [38]. Es importante destacar que en la mayoría de los sistemas analizados el comportamiento de los agentes es cooperativo, y que en muchos de ellos la interacción entre los agentes se da “por necesidad” y sin que exista un proceso de negociación (un agente necesita que otro realice una tarea específica por lo que le pide que la realice y le comunique el resultado).

Si bien en la literatura es posible encontrar una amplia variedad de artículos en los que los SMAs son aplicados a sistemas de recomendación para usuarios individuales, no ocurre lo mismo cuando se trata de recomendación a grupos de usuarios. En [34] los autores exploran la aplicación de SMA a sistemas de recomendación con el objetivo de producir recomendaciones grupales en el dominio de la planificación de viajes para grupos de usuarios. Las recomendaciones grupales surgen como el resultado de la negociación entre los agentes (uno por cada usuario) utilizando dos protocolos

de negociación que utilizan la metodología de negociación cooperativa. A diferencia de nuestro enfoque, el propuesto por Bekkerman tiene dos debilidades: (i) el protocolo que permite la negociación directa sólo resulta aplicable a grupos de 2 usuarios, y (ii) hace uso de técnicas de agregación en el protocolo denominado “combinación de rankings”.

En [14] se propone un enfoque similar al utilizado en [34], que según los autores es aplicable a todo tipo de dominio siempre y cuando éste sea representable con ontologías. En este caso, los agentes también actúan como representantes de los usuarios pero la generación de recomendaciones es distinta. La negociación se realiza por intercambios de propuestas las cuales los agentes pueden aceptar o rechazar (total o parcialmente), y lo que se negocia son las preferencias individuales de cada usuario con el objetivo de decidir cuáles de ellas formarán parte del perfil grupal. Un punto a destacar de este enfoque es que no hace uso en ningún momento de técnicas de agregación, ya que intenta que los agentes decidan como si fueran un grupo de personas. Este trabajo es luego refinado en [39], donde si bien la idea general del funcionamiento del sistema se mantiene, las comunicaciones entre los agentes son ahora centralizadas en un agente negociador que por un lado se encarga de ejecutar el protocolo de negociación, recibiendo, combinando y enviando propuestas, y por el otro actúa de mediador, interviniendo para evitar que la negociación se estanque y facilitar el consenso de los agentes.

Por otro lado, en [38] se propone un sistema llamado *e-Tourism* que utiliza un SMA con el objetivo de generar recomendaciones individuales y grupales también en el dominio del turismo. En este caso, los agentes son utilizados tanto como representantes de los usuarios como también para modelar componentes del sistema. El problema que presenta este enfoque es que si bien los autores definen a un grupo como un conjunto de *UserAgents* que de forma conjunta solicitan una recomendación grupal, para la generación de las recomendaciones grupales el sistema utiliza técnicas de agregación para realizar una agregación de perfiles. Una vez obtenido el perfil grupal (agregado), el sistema trata al grupo como un usuario individual y, por ende, se reutiliza en gran medida los métodos de recomendación para usuarios individuales.

Finalmente, en [40] se propone un enfoque que utiliza un SMA con el objetivo de proveer una recomendación a un grupo buscando que dicha recomendación cumpla con los requerimientos/preferencias de los miembros del grupo. En el sistema propuesto existe, por cada miembro del grupo, un agente que actúa como representante de dicho miembro en la negociación, protege los intereses del mismo y modela su comportamiento en una situación de conflicto (por ejemplo: si el usuario es cooperativo, competitivo, colaborativo, entre otros), lo cual tiene impacto directo en la forma en la que el agente realiza concesiones. Adicionalmente, existe un agente mediador que se encarga del proceso de toma de decisiones, la generación de propuestas (soluciones candidatas) y la negociación en sí (interactuando con todos los agentes representantes de los miembros del grupo).

III. ENFOQUE PROPUESTO

En vista de los problemas que pueden aparecer al utilizar técnicas de agregación para generar recomendaciones a grupos, se propone abordar el problema como una negociación entre agentes inteligentes. El enfoque propuesto, llamado MAGReS, está basado en un SMA en el cual los agentes actúan como asistentes personales de los miembros del grupo. Cada agente, al ser asignado a un miembro del grupo, conoce sus preferencias (las cuales son, en principio, los ratings asignados a los ítems que el usuario ha votado en el pasado) y es capaz tanto de predecir el rating que su usuario asignaría a un ítem que aún no votó, como de generar un ranking de ítems interesantes para el usuario asociado.

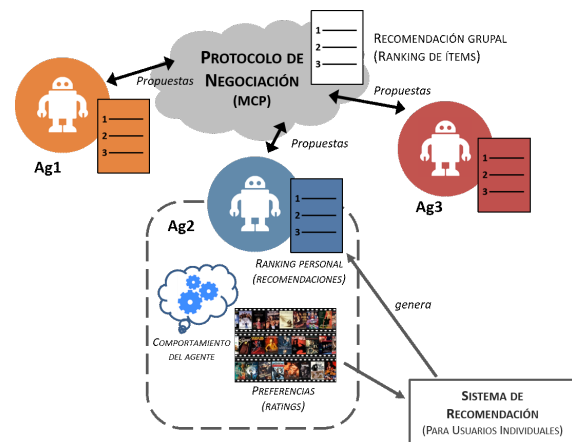


Fig. 1. Enfoque propuesto

El enfoque propuesto difiere de los enfoques tradicionalmente usados para realizar recomendaciones a grupos en que los enfoques tradicionales hacen uso de una entidad centralizada que utiliza algún tipo de técnica de agregación en conjunción con un sistema de recomendación para usuarios individuales. En MAGReS, en cambio, las técnicas de agregación son reemplazadas por un proceso de negociación entre los agentes que representan a los miembros del grupo (los *UserAgent*) (Figura 1). La negociación es guiada por un agente *Moderador*, el cual se encarga de coordinar a los agentes de acuerdo a las reglas del protocolo de negociación utilizado. Si bien existen varios protocolos de negociación, pocos son los que (i) intentan imitar el proceso de negociación utilizado por los humanos (lo cual podría potencialmente aumentar la calidad de las recomendaciones), y (ii) resultan aplicables a negociaciones multilaterales. Es por lo anterior que el protocolo conocido como MCP [41] fue elegido para desarrollar el prototipo del enfoque propuesto.

MCP

Los pasos del funcionamiento del protocolo MCP se resumen en la Figura 2. En el primero, cada agente realiza una propuesta inicial con su ítem “favorito”. Para ello, el agente construye una lista de ítems que son recomendables para su

usuario (es decir, que el usuario no haya votado previamente)¹, la ordena de mayor a menor respecto al rating predicho para las mismos, elige el primer ítem de la lista y, finalmente, construye una propuesta que lo contenga. Luego de esto, las propuestas iniciales de todos los agentes son enviadas al mediador con el objetivo de determinar si un acuerdo puede ser alcanzado.

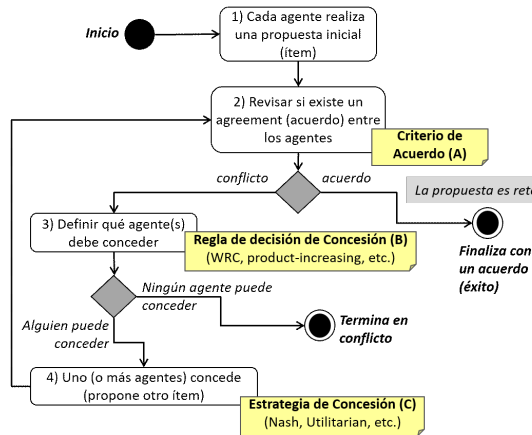


Fig. 2. Pasos del protocolo MCP para MAGReS (adaptado de [41])

La Figura 3 ilustra un caso de ejemplo del funcionamiento de MCP donde el grupo está formado por 3 Usuarios, cada uno de ellos representado por uno de los 3 agentes ($Ag1$, $Ag2$ y $Ag3$). En la primera ronda (Figura 3a) cada uno de los agentes propone una película: $Ag1$ propone $M1$, $Ag2$ propone $M24$ y $Ag3$ propone $M3$. Dado que el *criterio de acuerdo*² del protocolo determina que no existe un acuerdo. Al no haber acuerdo, uno (o más) de los agentes debe conceder, y la *regla de decisión de concesión*³ define que debe ser $Ag2$. Al iniciar la segunda ronda de negociación (Figura 3b), $Ag2$ concede y propone $M52$. Se vuelve a revisar si los agentes están de acuerdo y se halla un conflicto, ya que la película $M52$ es rechazada por $Ag1$ y $Ag3$. El proceso se repite hasta llegar a la ronda k , en la cual $Ag3$ propone la película $M46$, la cual satisface a todos los agentes por lo que la negociación termina exitosamente.

En [25] es posible encontrar más detalles respecto a MCP y su uso en el enfoque propuesto.

IV. EVALUACIÓN

El enfoque propuesto fue evaluado con datos sintéticos con resultados satisfactorios [25]. No obstante, se sabe que las simulaciones con datos sintéticos pueden a veces distanciarse de lo que ocurre en la realidad, por lo que se consideró necesario evaluar MAGReS con usuarios reales.

¹Para esto se utiliza un SR (para usuarios individuales)

²En este caso se utiliza el criterio *Acuerdo Multilateral* (propuesto en [41]): existe un acuerdo sobre una propuesta x sii el valor de satisfacción de cada agente para x es mayor o igual que el valor de satisfacción para su propuesta. Así, por ejemplo, en la *Ronda 1* (del ejemplo) para $Ag1$ ni $M24$ ni $M3$ son aceptables porque 0.7 y 0.56 son menores que 0.80.

³Se utilizó WRC (Willingness to Risk Conflict), una extensión del criterio de Zeuthen propuesta en [41]



Fig. 3. Ejemplo de MCP

El enfoque MAGReS fue implementado en Java como una aplicación Web con arquitectura Cliente-Servidor (Figura 4), utilizando tecnologías como SpringBoot⁴, Thymeleaf⁵, Bootstrap⁶ y Hibernate⁷. Los usuarios interactúan con el cliente Web, el cual les permite: iniciar una sesión en el sistema, votar películas y pedir recomendaciones grupales⁸. Por otro lado, el recomendador para usuarios individuales, utilizado por los agentes para construir las propuestas candidatas (que usarán durante el proceso de negociación) y por TRADGRec (el enfoque tradicional usado para realizar la evaluación) para producir las recomendaciones grupales a través de la técnica de agregación de preferencias, fue implementado utilizando el framework de recomendación de Mahout⁹.

A. Objetivo del experimento

El objetivo principal de este experimento es determinar si las mejoras observadas, tanto en la satisfacción grupal como en la uniformidad de satisfacción de los miembros del grupo, al experimentar con datos sintéticos se mantienen en un contexto real. Esto significa que se espera que, en comparación con un enfoque tradicional, el enfoque MAGReS genere recomendaciones que satisfagan en mayor medida tanto al grupo como un todo como a cada uno de sus miembros, y que también los miembros del grupo sean satisfechos de una forma más uniforme. Dado que mantener a todos los miembros del grupo uniformemente satisfechos no siempre es posible, puesto que depende de los intereses de cada uno

⁴<https://projects.spring.io/spring-boot/>

⁵<http://www.thymeleaf.org/>

⁶<http://getbootstrap.com/>

⁷<http://hibernate.org/>

⁸En [26], [42] es posible encontrar más detalles respecto a la funcionalidad de la aplicación implementada.

⁹<http://mahout.apache.org/>

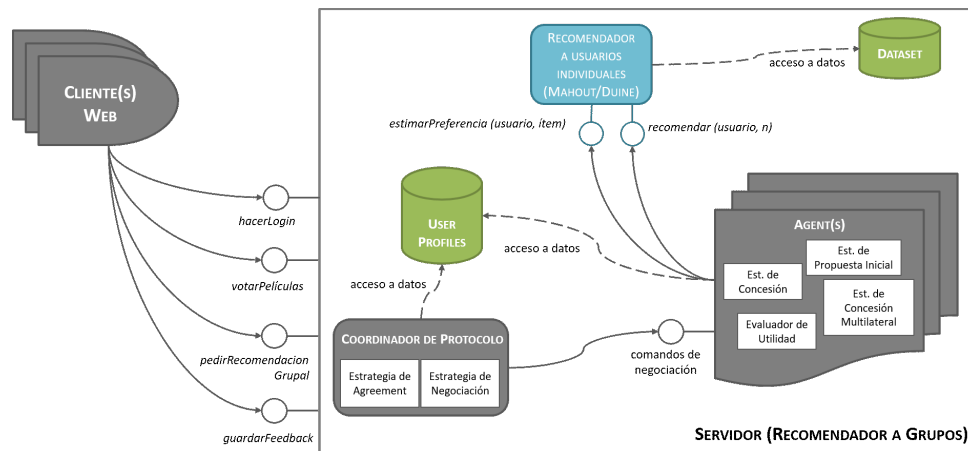


Fig. 4. Arquitectura de MAGReS

de los miembros y de cuán similares sean dichos intereses, se espera que al menos el nivel de satisfacción del miembro menos satisfecho del grupo se incremente.

Este experimento permitirá obtener conocimiento por un lado respecto al aporte de un enfoque basado en SMA en un entorno real, y por el otro respecto a la viabilidad y utilidad de tal enfoque.

B. Dataset

Para la realización del experimento se utilizó el dataset “MovieLens Latest (small)” de MovieLens¹⁰, el cual fue actualizado por última vez en Octubre de 2016 y consta de aproximadamente 100.000 ratings, 9.000 películas y 700 usuarios.

C. Configuración de los algoritmos

Dado que probar todas las posibles configuraciones de MAGReS implica una cantidad elevada de experimentos, y que pedir a los usuarios involucrados en las pruebas que realicen múltiples experimentos es inviable (puesto que puede generarles un alto nivel de estrés, generando desgano y como resultado sesgo en los feedbacks de las recomendaciones), se decidió fijar los parámetros de los algoritmos utilizando la configuración que mejores resultados había generado en las pruebas con grupos sintéticos. El SR individual utilizado para ambos enfoques realiza recomendación basada en ítems y la medida de similitud utilizada fue City Block (también conocida como distancia Manhattan [43]).

- MAGReS: el *criterio de acuerdo* es el llamado *Multilateral* [41], la *regla de decisión de concesión* elegida fue *Willingness to Risk Conflict* [41], la *estrategia de concesión* utilizada fue *Desires Distance* [25]. Adicionalmente, la propuesta inicial de cada agente fue realizada con el *criterio Egocentric* (proponen su película favorita, lo mejor para cada uno de ellos sin importar la opinión de los demás), y los agentes conceden utilizando la estrategia *UtilityThreshold* (la cual determina que podrán conceder

si la siguiente propuesta candidata les produce un nivel de satisfacción por encima de un *threshold* establecido¹¹).

- TRADGRec: realiza agregación de preferencias utilizando el promedio como técnica de agregación.

D. Organización del experimento

El experimento constó de las siguientes etapas:

- Selección de participantes del experimento. Participaron 34 personas, de entre 20 y 40 años de edad.
- Construcción de los perfiles individuales: se le pidió a un conjunto de personas que crearan un usuario en la herramienta y asignaran ratings al menos a 15 películas, con el fin de construir un pequeño perfil de usuario. Los ratings asignables toman valores entre 0 y 5 estrellas. La cantidad total de películas disponibles para asignar ratings por los usuarios es de 300, todas ellas seleccionadas teniendo en cuenta que:
 - fueran de géneros diversos.
 - fueran recientes: tratando que hayan sido lanzadas luego de la fecha de nacimiento de los participantes del experimento (los cuales, en su mayoría, nacieron a fines de la década del 80), o incluyendo películas previas al nacimiento de los mismos pero conocidas por la sociedad en general.
- Formación de grupos: los participantes formaron grupos de 3 personas. Aunque hubo tres excepciones (tres de los grupos sólo tuvieron 2 integrantes).
- Generación de recomendaciones grupales: una vez generados los grupos, cada uno de ellos debió utilizar la herramienta para solicitar una recomendación grupal utilizando cada uno de los enfoques de recomendación disponibles: MAGReS y TRADGRec (nuestra implementación del enfoque de agregación de preferencias tradicional). Se decidió fijar 10 como la cantidad de

¹¹El *threshold* es un número entre 0 y 1, y se fijó en 0.15 para tratar de reducir la cantidad de conflictos. En los experimentos con datos sintéticos se fija el *threshold* en 0.65

¹⁰<https://grouplens.org/datasets/movielens/>

ítems que debían tener todas las recomendaciones durante el experimento.

- (v) Evaluación de recomendaciones: una vez generada la recomendación con una técnica, la herramienta muestra una interfaz que los usuarios deben utilizar para cargar su feedback respecto a la recomendación recibida, tanto individual como grupalmente. El feedback se expresa como un rating (con valor entre 1 y 5 estrellas).

Las pruebas se realizaron con 18 grupos mixtos¹², 15 de 3 integrantes y 3 de 2, pero 3 de los grupos no realizaron el paso 5 del experimento para las recomendaciones producidas utilizando la técnica TRADGRec, por lo cual los datos de dichos grupos no sirven para la comparación entre los enfoques.

E. Datos analizables

1) *Estimaciones:* Al momento de generar cada una de las recomendaciones la herramienta almacena para cada ítem recomendado:

- Los valores de satisfacción individual estimados de cada uno de los miembros del grupo. Estos son calculados utilizando el recomendador a usuarios individuales al que cada agente tiene acceso en el caso del enfoque MAGReS, el mismo SR que se utiliza para generar las recomendaciones usando el enfoque tradicional (TRADGRec).
- El valor de satisfacción grupal estimado, el cual se calcula como una agregación de los valores de satisfacción individual estimados. La agregación se realiza sólo con el fin de tener un valor de satisfacción grupal que resulte comparable entre los dos enfoques¹³.
- La cantidad de información que debió revelar cada usuario para que se pueda producir la recomendación. Se miden dos aspectos: la cantidad de propuestas candidatas reveladas y el porcentaje de información relativa a la función de utilidad del usuario que fue revelado.

Además, para toda la recomendación se calculan los valores promedio de cada uno de los indicadores almacenados, a modo de resumir los datos abarcando toda la recomendación. Toda esta información permite comparar ambos enfoques respecto a la calidad estimada de las recomendaciones que realizan en un contexto real, pudiéndose observar cuál de los enfoques produjo recomendaciones con: el mayor nivel de satisfacción (utilidad) grupal, el mayor nivel de satisfacción individual y la menor dispersión en la satisfacción de los miembros del grupo (y su análogo: la mayor uniformidad de satisfacción).

2) *Feedback:* En la última etapa del experimento los sujetos de prueba deben evaluar las recomendaciones. Esto brinda información acerca de los niveles reales de satisfacción grupal

¹²Se realizaron dos pruebas: en la primera prueba piloto participaron 10 grupos (8 grupos de 3 integrantes y 2 grupos de 2) y en la segunda participaron 8 grupos (7 de 3 integrantes y 1 de 2). Se debe notar que debido a que se disponía de pocos sujetos de prueba, algunos de ellos formaron parte de más de un grupo (la selección de miembros de grupos fue aleatoria).

¹³La técnica de agregación utilizada puede diferir de la utilizada por TRADGRec para realizar la agregación de preferencias (ej. se podría usar "mayoría" para la agregación de preferencias, y "promedio" para mostrar el resultado) y debe ser la misma tanto para TRADGRec como para MAGReS.

e individual, lo cual a su vez permite (a través del cálculo del desvío sobre los niveles de satisfacción individual) determinar cuán uniformemente fueron satisfechos los miembros del grupo. De esta manera, no sólo es posible comparar los enfoques en función de "la calidad real de las recomendaciones", sino que también es posible medir el error de estimación de cada uno de los enfoques utilizando, por ejemplo, el MAE (del inglés, *Mean Absolute Error*). Se debe notar que al disponer de una cantidad reducida de grupos no es posible generalizar los resultados hallados ya que no es posible aplicar el Teorema del Límite Central (T.L.C. [44]). Aún así, los resultados preliminares obtenidos ayudan a tener una idea respecto de la viabilidad del enfoque propuesto al ser aplicado en una situación real.

F. Resultados obtenidos

Análisis de Estimaciones: Al realizar los análisis de los valores de satisfacción individual, grupal y uniformidad de satisfacción estimados los resultados no fueron particularmente alentadores. A pesar de que la satisfacción grupal e individual era superior para las recomendaciones generadas por MAGReS, la uniformidad de satisfacción era inferior en muchos de los casos (en algunos llegó a ser un 8% inferior). Esto principalmente se debió a que algunos parámetros de MAGReS fueron relajados con el objetivo de reducir el tiempo que requería la recomendación, por ejemplo, el threshold que define la utilidad mínima que debe tener una propuesta candidata para que sea elegible por el agente (ver Sección IV-C), por lo que un ajuste en dichos parámetros podría resolver el problema.

Tal como se observa en la Figura 5a, si bien las recomendaciones generadas por MAGReS tienen un mayor valor de satisfacción grupal estimada en la mayoría de los casos, en muchos de ellos la dispersión de satisfacción estimada para el grupo es más alta y por consiguiente la uniformidad de satisfacción es menor. Concretamente, en el 83.3% de los casos la satisfacción grupal fue mayor, pero en el sólo en el 16.7% de los casos se cumple que al mismo tiempo que aumenta la satisfacción grupal se reduce la dispersión de satisfacción. De hecho, en el caso en que se registró el máximo incremento de satisfacción grupal (un 13.58% para el *GRUPO 11*, 0.824 para MAGReS vs. 0.725 para TRADGRec), se produjo al mismo tiempo un aumento en la dispersión de satisfacción (de aproximadamente un 587.4%, 0,090 vs. 0,0132) y por consiguiente una caída en el valor de la uniformidad de satisfacción (de aproximadamente un 7.88%, 0.909 vs. 0.987).

Con el objetivo de medir la relevancia de los datos obtenidos se decidió analizarlos utilizando un test estadístico. Al analizar los datos con el test de Shapiro-Wilk [45] se encontró que una de las cuatro muestras¹⁴ (la uniformidad de satisfacción para TRADGRec) no seguía una distribución normal, por lo que era necesario utilizar un test no paramétrico. Dado que las muestras son pequeñas y estaban relacionadas de a pares

¹⁴Satisfacción grupal (1 muestra por cada enfoque) y Uniformidad de satisfacción (1 muestra por cada enfoque)

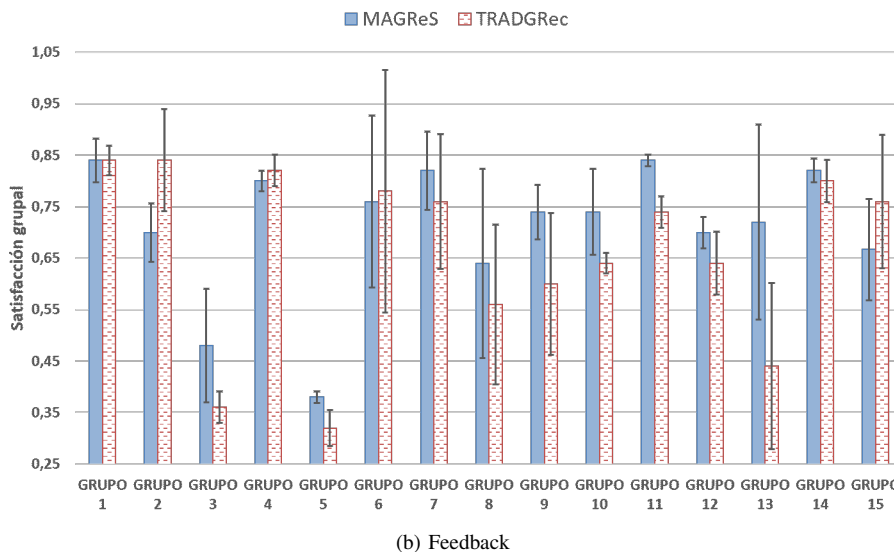
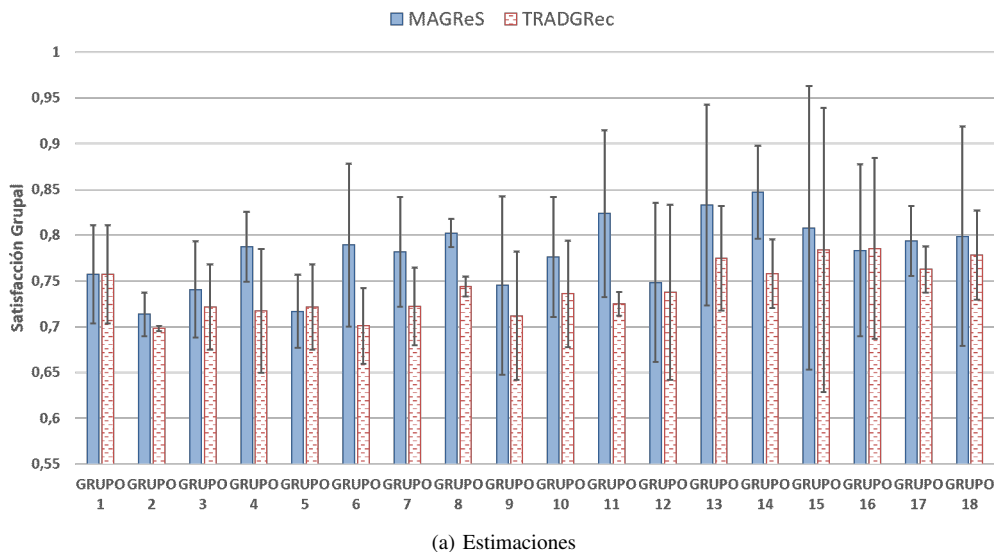


Fig. 5. Estimación de satisfacción grupal con dispersión de satisfacción individual (Satisfacción: mayor es mejor; Dispersión: menor es mejor)

(por cada grupo que participó del experimento se conocían los niveles de satisfacción grupal y dispersión de satisfacción para dos recomendaciones con técnicas distintas) se utilizó el test *Wilcoxon Signed Ranks para muestras relacionadas* [45].

- *Test 1*: Nivel de satisfacción grupal estimado. Dado que se buscaba probar que “la satisfacción grupal estimada para las recomendaciones de MAGReS era superior a la de las recomendaciones de TRADGRec” (esto es la hipótesis alternativa, formalmente $H_A : \mu_g(MAGReS) > \mu_{sg}(TRADGRec)$), se definió la hipótesis nula como la negación de dicha expresión (formalmente, $H_0 : \mu_{sg}(MAGReS) \leq \mu_{sg}(TRADGRec)$). El resultado del test arrojó que, con un nivel de confianza del 95% ($\alpha = 0.05$), había suficiente evidencia estadística para rechazar la hipótesis nula con un p-valor de 0.0006833, lo cual confirma lo observado previamente.

- *Test 2*: Uniformidad de satisfacción estimada. En este caso se realizó el análisis de los desvíos en la satisfacción del grupo (la uniformidad se puede calcular como $1 - \text{desvío}$). Se definió la hipótesis alternativa como “el desvío en la satisfacción grupal es menor para las recomendaciones de MAGReS que para las de TRADGRec” (formalmente, $\mu_{\sigma_g}(MAGReS) < \mu_{\sigma_g}(TRADGRec)$), con lo que la hipótesis nula fue definida como la negación de la hipótesis alternativa (formalmente, $\mu_{\sigma_g}(MAGReS) \geq \mu_{\sigma_g}(TRADGRec)$). El test de Wilcoxon arrojó que con un 95% de confianza no era posible rechazar la hipótesis nula (el p-valor en este caso fue de 0.9742), lo cual confirmó lo observado previamente.

A pesar de todo lo anterior, se debe tener en cuenta que:

- Aún en los casos en los que la satisfacción grupal

es mayor a costa de tener una menor uniformidad, en general, la satisfacción producida para el miembro menos satisfecho del grupo es mayor (observar las barras de error en las Figuras 5a y 5b, específicamente el extremo inferior de la misma) lo cual en sí representa una mejora respecto del enfoque tradicional.

- Si bien los resultados de los test estadísticos confirmaron las observaciones realizadas a priori, se considera necesario aumentar la cantidad de grupos participantes en las pruebas para que resulte factible generalizar las observaciones realizadas.

Análisis de Feedback: Si bien el análisis de las estimaciones arrojó resultados un tanto desesperanzadores, no ocurrió lo mismo con los feedbacks dados por los usuarios. De hecho, tal como se puede observar en la Figura 5b, si bien siguen existiendo casos en los cuales ocurre que se incrementa (considerando las recomendaciones de MAGReS respecto a las de TRADGRec) la satisfacción grupal pero disminuye la uniformidad de satisfacción, o viceversa, para muchos de los grupos (concretamente un 66% de ellos) se registraron incrementos simultáneos en la satisfacción grupal y la uniformidad de satisfacción de los individuos dentro del grupo. Particularmente, el mejor caso registrado fue el del GRUPO 9, para el cual la satisfacción grupal se incrementó en un 23% (de 0.60 para las recomendaciones realizadas por TRADGRec a 0.74 para las de MAGReS) y la uniformidad aumentó en un 9.83% (de 0.862 para TRADGRec a 0.947 para MAGReS). En general, los resultados de la evaluación del feedback dado por los usuarios informaron que al usar MAGReS la satisfacción grupal aumentó en promedio en un 20.2% y la uniformidad de satisfacción individual aumentó en promedio en un 4.4%. Se debe notar que, si bien en algunos casos las recomendaciones de MAGReS tuvieron una menor uniformidad de satisfacción (la “pérdida de uniformidad” se ubicó generalmente alrededor del 3%), esa pérdida de uniformidad se compensaba con grandes ganancias en la satisfacción grupal (siendo los grupos 3, 8 y 13 ejemplos de esta situación).

Adicionalmente, se debe mencionar que en al menos un tercio de los casos, los usuarios “castigaron” a MAGReS en el feedback dado (reduciendo la puntuación) debido a que en la lista de recomendaciones había películas que 2 de los integrantes del grupo habían visto y calificado. Esta situación no se da con TRADGRec por la forma en la que está implementado: al computar la lista de preferencias grupales se agregan a la lista todos aquellos items votados por al menos uno de los miembros del grupo. En MAGReS, en cambio, basta con que uno de los miembros del grupo no haya visto una película para que su agente pueda proponerla, lo que da la posibilidad a que los otros agentes la acepten y así forme parte de la lista de ítems recomendados¹⁵.

Al igual que con el análisis de las estimaciones, los datos obtenidos fueron analizados con un test estadístico. En este

¹⁵Esta situación fue corregida (en versiones posteriores de MAGReS) a través de la agregación de un parámetro adicional a cada agente que modela su flexibilidad respecto a este tipo de situaciones (por ejemplo: permitido, prohibido y flexible hasta cierto punto).

caso, el test de normalidad reveló que ninguna de las muestras analizadas podía ser considerada como normal. Así, se aplicó nuevamente el test de Wilcoxon y se realizaron los mismos test que los mencionados en la sección IV-F. El *Test 1* reveló que, con un nivel de confianza del 95% ($\alpha = 0.05$) existía evidencia estadística suficiente para rechazar la hipótesis nula ($p - valor = 0.04488$), lo que confirmaba las observaciones a priori: al evaluar el feedback dado por los usuarios, MAGReS genera recomendaciones que satisfacen en mayor medida a los grupos. El *Test 2*, por otro lado, no permitió confirmar lo observado a priori ya que al no ser posible rechazar la hipótesis nula de dicho test ($p - valor = 0.1514$), no existía evidencia estadística suficiente para afirmar que las recomendaciones de MAGReS satisfacían “más uniformemente” a los usuarios. Ante esto, se comprobó, utilizando también el test de Wilcoxon (pero a dos colas), que “las poblaciones de las que provenían las muestras fuesen significativamente distintas” (H_0). De acuerdo a los resultados, bajo un $\alpha = 0.05$ no existía evidencia estadística para rechazar la hipótesis nula ($H_0 =$ las poblaciones son significativamente distintas) ya que se obtuvo un $p - valor$ de 0.3028.

Comparativa: Estimaciones vs Feedback: Dado que se disponía tanto de datos de estimaciones como de feedback para las recomendaciones de los dos enfoques, resultó posible analizar la precisión de las predicciones de estos utilizando el Error Medio Absoluto (o MAE, por las siglas del inglés *Mean Absolute Error*). Un error de estimación en la predicción de los niveles de satisfacción individual y grupal implica que el modelado de ambas funciones de satisfacción se condice con la realidad, lo cual repercute directamente en la calidad de las recomendaciones producidas: si las preferencias están adecuadamente modeladas, las recomendaciones se adecuarán de mejor manera a los intereses del grupo.

La información presentada por el gráfico de la Figura 6 demuestra que claramente las predicciones realizadas por MAGReS, tanto en cuanto al nivel de satisfacción grupal como al individual, superan con creces (en precisión) a las generadas por TRADGRec¹⁶. En base a lo anterior, se puede decir que los resultados del análisis de estimaciones (ver Sección IV-F) no eran tan malos como parecían, dado que si bien MAGReS no superaba a TRADGRec en muchas ocasiones esto fue causado, en cierta medida, por las sobre-estimaciones de TRADGRec.

V. CONCLUSIONES

En este artículo se evaluó la aplicación de un enfoque de recomendación a grupos basado en sistemas multi-agente llamado MAGReS en un contexto real, y se compararon las recomendaciones de dicho enfoque con las de un enfoque tradicional. Si bien al evaluar las recomendaciones en función de los valores estimados de satisfacción grupal y uniformidad de satisfacción individual se observó que en muchos de los casos la aplicación de MAGReS no generaba mejoras significativas cuando se tenían ambos aspectos en cuenta, sí

¹⁶En el análisis realizado solo fueron tenidos en cuenta 15 de los 18 grupos que formaron parte del experimento puesto que 3 de los grupos no dieron feedback respecto a la recomendación de TRADGRec.

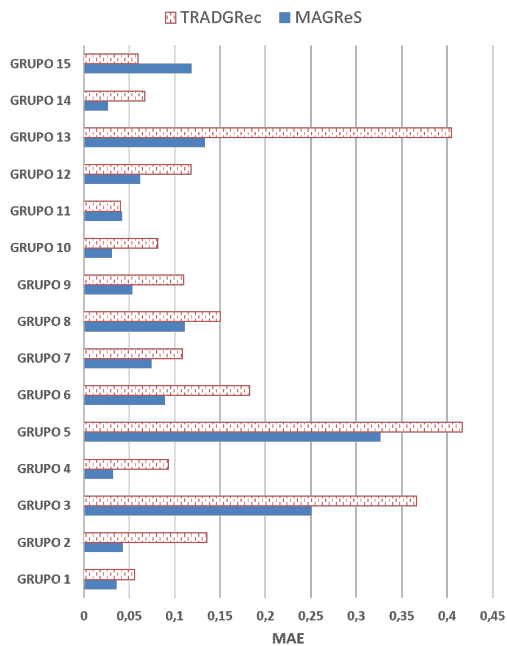


Fig. 6. Error Absoluto Medio (MAE; menor es mejor)

se observó que en general las recomendaciones generaban una mayor satisfacción tanto grupal como individual. Cuando se evaluó el feedback dado por los usuarios que participaron del experimento la situación cambió. Desde ese punto de vista, se observó que las recomendaciones producidas por MAGReS superaron con creces a las producidas por el enfoque tradicional y, además, que la calidad de las predicciones realizadas por el recomendador mejoraba cuando el enfoque utilizado era MAGReS.

Los resultados de los experimentos estuvieron condicionados por una serie de factores que se detallan a continuación. El primer factor es la cantidad de grupos que formaron parte del experimento y la composición de los mismos, ya que a pesar de que los grupos se armaron de forma aleatoria, algunos usuarios formaron parte de dos grupos distintos. El segundo es la parametrización de los algoritmos dado que utilizar, por ejemplo, otra métrica de similitud podría generar cambios radicales en las recomendaciones. Por último, el tercer factor es que actualmente las funciones de utilidad/satisfacción de los agentes son modeladas usando la funcionalidad de predicción del recomendador para usuarios individuales al que los agentes tienen acceso (esto está vinculado a lo explicado para el segundo factor).

A modo de trabajo futuro, se planea realizar el experimento con nuevos grupos, para evaluar la posible generalización de los resultados. Adicionalmente, se considera necesario realizar pruebas utilizando una parametrización más estricta (“menos relajada”) del enfoque MAGReS, dado que la relajación de algunos de los parámetros en pos de agilizar el proceso de recomendación para evitar que el proceso de pruebas sea tedioso para los usuarios puede haber sido la causa de

los resultados observados, particularmente en relación a las estimaciones.

AGRADECIMIENTOS

Este trabajo fue parcialmente financiado por el proyecto PIP 2015 – 2017 número 11220150100030CO.

REFERENCIAS

- [1] P. Rogers, R. Puryear, and J. Root, “Infobesity: The enemy of good decisions,” 2013. [Online]. Available: http://www.bain.com/Images/BAIN_BRIEF_Decision_Insights_Infobesity.pdf
- [2] C. C. Yang, H. Chen, and K. Hong, “Visualization of large category map for internet browsing,” *Decis. Support Syst.*, vol. 35, no. 1, pp. 89–102, Apr. 2003.
- [3] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, 2nd ed. New York, NY, USA: Springer-Verlag New York, Inc., 2015.
- [4] B. Miller, I. Albert, S. Lam, J. Konstan, and J. Riedl, “Movielens unplugged: experiences with an occasionally connected recommender system,” in *Proceedings of the 8th international conference on Intelligent user interfaces (IUI '03)*. ACM, 2003, pp. 263–266.
- [5] A. Piliponyte, F. Ricci, and J. Koschwitz, “Sequential music recommendations for groups by balancing user satisfaction,” in *UMAP 2013 - Workshop Group Recommender Systems: Concepts, Technology, Evaluation*. CEUR, 2013.
- [6] F. Ricci, “Travel recommender systems,” *IEEE Intelligent Systems*, vol. 17, no. 6, pp. 55–57, 2002.
- [7] S. Schiaffino and A. Amandi, “Building an expert travel agent as a software agent,” *Expert Systems with Applications*, vol. 36, pp. 1291–1299, 2009.
- [8] A. Jameson and B. Smyth, *Recommendation to Groups*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 596–627.
- [9] L. Ardissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso, “Intrigue: Personalized recommendation of tourist attractions for desktop and hand held devices,” *Applied Artificial Intelligence*, vol. 17, no. 8-9, pp. 687–714, 2003.
- [10] J. F. McCarthy, “Pocket restaurant finder: A situated recommender systems for groups,” in *Proceeding of Workshop on Mobile Ad-Hoc Communication at the 2002 ACM Conference on Human Factors in Computer Systems*, 2002.
- [11] M. O’Connor, D. Cosley, J. A. Konstan, and J. Riedl, “PolyLens: a recommender system for groups of users,” in *Proc. of ECSCW’01*. Kluwer Academic Publishers, 2001, pp. 199–218.
- [12] Y. Blanco-Fernández, J. J. Pazos-Arias, A. Gil-Solla, M. Ramos-Cabrer, B. Barragáns-Martínez, M. López-Nores, J. García-Duque, A. Fernández-Vilas, and R. P. Díaz-Redondo, *AVATAR: An Advanced Multi-agent Recommender System of Personalized TV Contents by Semantic Reasoning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 415–421.
- [13] R. Burke, “Hybrid recommender systems: Survey and experiments,” *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.
- [14] I. Garcia, L. Sebastia, and E. Onaindia, “A negotiation approach for group recommendation,” in *Proceedings of the 2009 International Conference on Artificial Intelligence*, 2009, pp. 919–925.
- [15] J. S. Lopes, S. Alvarez-Napagao, R. Confalonieri, and J. Vázquez-Salceda, “Use: a multi-agent user-driven recommendation system for semantic knowledge extraction,” 2009.
- [16] S. Ujjin and P. J. Bentley, “Particle swarm optimization recommender system,” in *Proceedings of the IEEE Swarm Intelligence Symposium 2003 (SIS 2003)*, 2003, pp. 124–131.
- [17] M. Niazi and A. Hussain, “Agent-based computing from multi-agent systems to agent-based models: a visual survey,” *Scientometrics*, vol. 89, no. 2, p. 479, 2011.
- [18] M. Wooldridge, *An Introduction to MultiAgent Systems*, second edition ed. John Wiley & Sons, 2009.
- [19] M. Wooldridge and N. R. Jennings, *Agent theories, architectures, and languages: A survey*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995, pp. 1–39.
- [20] M. Wooldridge and N. Jennings, “Intelligent agents: Theory and practice,” *Knowledge Engineering Review*, vol. 10, pp. 115–152, 1995.

- [21] M. N. Huhns and L. M. Stephens, "Multiagent systems," G. Weiss, Ed. MIT Press, 1999, ch. Multiagent Systems and Societies of Agents, pp. 79–120.
- [22] N. Jennings, P. Faratin, A. Lomuscio, S. Parsons, M. Wooldridge, and C. Sierra, "Automated negotiation: Prospects, methods and challenges," *Group Decision and Negotiation*, vol. 10, no. 2, pp. 199–215, 2001.
- [23] J. S. Rosenschein and G. Zlotkin, *Rules of Encounter: Designing Conventions for Automated Negotiation Among Computers*. Cambridge, MA, USA: MIT Press, 1994.
- [24] P. Stone and M. Veloso, "Multiagent systems: A survey from a machine learning perspective," *Autonomous Robots*, vol. 8, no. 3, pp. 345–383, 2000.
- [25] C. Villavicencio, S. Schiaffino, J. A. Diaz-Pace, A. Monteserin, Y. Demazeau, and C. Adam, *A MAS Approach for Group Recommendation Based on Negotiation Techniques*. Cham: Springer International Publishing, 2016, pp. 219–231. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-39324-7_19
- [26] C. Villavicencio, S. Schiaffino, J. A. Diaz-Pace, and A. Monteserin, "A group recommendation system for movies based on mas," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 5, no. 3, 2016.
- [27] I. Cantador and P. Castells, "Group recommender systems: New perspectives in the social web," in *Recommender Systems for the Social Web*, ser. Intelligent Systems Reference Library. Springer Berlin Heidelberg, 2012, vol. 32, pp. 139–157. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-25694-3_7
- [28] L. Baltrunas, T. Makcinskas, and F. Ricci, "Group recommendations with rank aggregation and collaborative filtering," in *Proc. of RecSys '10*. ACM, 2010, pp. 119–126.
- [29] I. Christensen and S. Schiaffino, "A hybrid approach for group profiling in recommender systems," *J. of Universal Computer Science*, vol. 20, no. 4, pp. 507–533, 2014.
- [30] J. Morais, E. Oliveira, and A. Jorge, *Distributed Computing and Artificial Intelligence*, ser. Advances in Intelligent and Soft Computing. Springer, 2012, vol. 151, ch. A Multi-Agent Recommender System, pp. 281–288.
- [31] W.-P. Lee, "Towards agent-based decision making in the electronic marketplace: interactive recommendation and automated negotiation," pp. 665–679, nov 2004. [Online]. Available: <http://dx.doi.org/10.1016/j.eswa.2004.07.001>
- [32] P. Skocir, L. Marusic, M. Marusic, and A. Petric, *Agent and Multi-Agent Systems. Technologies and Applications*, ser. LNCS. Springer, 2012, vol. 7327, ch. The MARS - A Multi-Agent Recommendation System for Games on Mobile Phones, pp. 104–113.
- [33] P. Bedi, S. K. Agarwal, V. Jindal, and Richa, "Marst: Multi-agent recommender system for e-tourism using reputation based collaborative filtering," in *Proceedings of the 9th International Workshop on Databases in Networked Information Systems - Volume 8381*, ser. DNIS 2014. New York, NY, USA: Springer-Verlag New York, Inc., 2014, pp. 189–201.
- [34] P. Bekkerman, K. Sarit, and F. Ricci, "Applying cooperative negotiation methodology to group recommendation problem," in *ECAI Workshop on Recommender systems*, 2006.
- [35] F. Lorenzi, A. Bazzan, and M. Abel, "An Architecture for a Multiagent Recommender System in Travel Recommendation Scenarios," in *ECAI Workshop on Recommender Systems*, vol. 1, 2006, pp. 88–91.
- [36] P. Rodríguez, N. Duque, and D. A. Ovalle, *Multi-agent System for Knowledge-Based Recommendation of Learning Objects Using Metadata Clustering*. Cham: Springer International Publishing, 2015, pp. 356–364.
- [37] D. Rosaci and G. M. L. Sarné, "A multi-agent recommender system for supporting device adaptivity in e-commerce," *J Intell Inf Syst*, vol. 38, no. 2, pp. 393–418, may 2011.
- [38] L. Sebastián, A. Giret, and I. García, "A multi agent architecture for single user and group recommendation in the tourism domain," 2011. [Online]. Available: <http://hdl.handle.net/10251/35364>
- [39] I. Garcia and L. Sebastia, "A negotiation framework for heterogeneous group recommendation," *Expert Systems with Applications*, vol. 41, no. 4.1, pp. 1245 – 1261, 2014.
- [40] S. Rossi, C. D. Napoli, F. Barile, and L. Liguori, "Conflict resolution profiles and agent negotiation for group recommendations," 2016.
- [41] U. Endriss, "Monotonic concession protocols for multilateral negotiation," in *Proc. AAMAS 2006*. New York, NY, USA: ACM, 2006, pp. 392–399. [Online]. Available: <http://doi.acm.org/10.1145/1160633.1160702>
- [42] C. Villavicencio, S. Schiaffino, J. A. Diaz-Pace, and A. Monteserin, *PUMAS-GR: A Negotiation-Based Group Recommendation System for Movies*. Cham: Springer International Publishing, 2016, pp. 294–298. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-39324-7_34
- [43] E. Krause, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*, ser. Dover Books on Mathematics Series. Dover Publications, 1975.
- [44] C. Grinstead and J. Snell, *Introduction to Probability*. American Mathematical Society, 1997.
- [45] G. W. Corder and D. I. Foreman, *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. New Jersey: Wiley, 2009.