Chapter 7

Modeling Virtual Footprints

Rajiv Kadaba
The University of Texas at Austin, USA

Suratna Budalakoti
The University of Texas at Austin, USA

David DeAngelis
The University of Texas at Austin, USA

K. Suzanne Barber
The University of Texas at Austin, USA

ABSTRACT

Entities interacting on the web establish their identity by creating virtual personas. These entities, or agents, can be human users or software-based. This research models identity using the Entity-Persona Model, a semantically annotated social network inferred from the persistent traces of interaction between personas on the web. A Persona Mapping Algorithm is proposed which compares the local views of personas in their social network referred to as their Virtual Signatures, for structural and semantic similarity. The semantics of the Entity-Persona Model are modeled by a vector space model of the text associated with the personas in the network, which allows comparison of their Virtual Signatures. This enables all the publicly accessible personas of an entity to be identified on the scale of the web. This research enables an agent to identify a single entity using multiple personas on different networks, provided that multiple personas exhibit characteristic behavior. The agent is able to increase the trustworthiness of on-line interactions by establishing the identity of entities operating under multiple personas. Consequently, reputation measures based on on-line interactions with multiple personas can be aggregated and resolved to the true singular identity.

1. INTRODUCTION

The way that an individual’s identity is created and experienced is fundamentally different in the virtual world. The basic cues used to uniquely identify individuals in the real world are missing, making the association between an entity and its identity ambiguous (Turkle, 1997). This research creates a model of the virtual world which dispels this ambiguity, allowing the virtual personas cre-
ated by an entity to be linked together. Informally, a virtual persona is a name and its associated attributes, which an entity uses to communicate with other personas.

The virtual world in the context of this research refers collectively to various explicit or inferred social networks. Examples of explicit social networks are websites such as Facebook, Orkut, MySpace, and LinkedIn. Social networks can be inferred from the digital traces of interaction between entities, or individuals, on the Internet, such as in the blogosphere (Kumar et al., 2004), online discussion forums, knowledge sharing sites, IRC logs and the co-occurrence of names in the large amount of textual data on the Internet (Jin et al., 2007). In an explicit social network there exists a framework by which entities can specify to whom they are related and the context of this relationship.

Access to explicit networks is generally controlled (Grimmelmann, 2008) because of the privacy concerns of its participants (Acquisti & Gross, 2006). Inferred social networks lack the privacy mechanisms of explicit networks as its users assume they are as anonymous as they wish to be. The work in this paper counters this assumption since users must engage in consistent and information rich interactions in order to provide value to the framework. The establishment of the reputation of an entity’s virtual persona within the framework is an important motivating factor for its consistent use, as others use reputation to assess the reliability of information associated with the persona (Donath, 1999). Every new persona created will need to establish its reputation within its social network, which requires time and effort. This penalty associated with creating a persona which is capable of meaningful interaction makes a persona valuable. Virtual personas with erratic interactions are not worth detecting as they have little value.

In a more general sense; an entity is an agent interacting within a multi agent framework. An agent could be a software program that acts according to an internally specified set of goals, or objectives. An agent can also be a dedicated piece of hardware designed to accomplish a particular task in the virtual world. An agent can even be a team of humans operating under a common goal. This research explores the canonical example of matching multiple virtual personas to a single human entity.

Search engines treat personal names and pseudonyms as keywords, giving virtual personas the same status as ordinary text. Queries for people’s names only find occurrences with verbatim matches to the query text while they may have interacted extensively, using various personas. The model proposed in this paper can be used to find more accurate results for the information associated with an individual available publicly on the web. Augmenting web search with the ability to link entities and their personas can be perceived as an attack on an individual’s privacy, as information which may have been exchanged with the expectation of anonymity granted by a virtual persona is now linked back to its progenitor. Conversely this research can also contribute to an individual’s ability to safeguard their privacy and protect their identity from theft. As the concept of identity in the virtual world is formalized and an upper bound on the ability of a determined adversarial agent is found, techniques to remain anonymous in spite of sophisticated statistical tools can be developed. Further, software agents who are capable of warning users of the unintended inferences which can be made with the data they publish may also be possible.

The proposed technique for resolving multiple personas requires that an entity exhibit characteristic behavior across multiple personas. Measurable entity behaviors can be both content-based and link-based. Content-based behaviors include the